Nucleotide Recognition and Metal Ion Requirements
of an RNA Ligase Ribozyme

by

Margaret E. Glasner

B.S., Molecular Biology
B.M., Music Performance
University of Wyoming, 1995

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL
FULLFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2003



Signature of Author: _____
Department of Biology
(February 3, 2003)

Certified by: _____
David P. Bartel
Associate Professor of Biology
Thesis Supervisor

Accepted by: _____
Alan D. Grossman
Professor of Biology
Chair, Biology Graduate Committee

Nucleotide Recognition and Metal Ion Requirements
of an RNA Ligase Ribozyme

by

Margaret E. Glasner

Submitted to the Department of Biology
on August 16, 2002 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Biology

## ABSTRACT

The class I ligase, a ribozyme previously isolated from random sequence, catalyzes a reaction similar to RNA polymerization, positioning its 5′-nucleotide via a Watson-Crick base pair, forming a 3′,5′-phosphodiester bond between its 5′-nucleotide and the substrate, and releasing pyrophosphate. The class I ligase forms the catalytic domain of an RNA polymerase ribozyme which is capable of extending a primer by 14 nucleotides. The ability of the class I ligase to support RNA-catalyzed RNA polymerization provides support for the RNA world hypothesis, which states that RNA played a more prominent role early in the evolution of life, functioning as both genome and enzymes. In order to determine ways in which catalysis by this ribozyme might be improved and to provide data for an in-depth comparison between the ligase and other ribozymes and protein-enzymes, we characterized the nucleotide triphosphate recognition and metal ion requirements of the class I ligase.

Nucleotide recognition by the ligase ribozyme was mediated almost exclusively through Watson-Crick base pairing, endowing the ligase and its derived polymerase with high fidelity. However, primer extension was 100-fold slower than the self-ligation reaction of the parent ribozyme. This rate reduction was traced to the deletion of the second and third nucleotides of the self-ligating ribozyme. When these nucleotides were restored by supplying a trinucleotide substrate, pppGGA, the rate of the reaction was nearly as fast as self-ligation, even though primer extension is a trimolecular reaction.

An analysis of the metal ion requirements of the ligase revealed that it has a unique specificity for $Mg^{2+}$ among ribozymes. In addition, these experiments demonstrated that the ligase is folded in the presence of a variety of metal ions, although its activity is abolished. Finally, these experiments suggested that the ligase may use a mechanism requiring two metal ions, like that of polymerases made of protein.

Thesis Supervisor: David P. Bartel
Title: Associate Professor of Biology

## Table of Contents

# Introduction

## Introduction to the RNA World

Early in the evolution of life, RNA may have played an even more prominent role than it does today, functioning as both genome and enzymes [1-6]. Support for this RNA world hypothesis comes from its myriad vital functions in modern life and the existence of natural and artificially selected ribozymes [5-12]. Among the universally conserved roles of RNA are its function in priming DNA synthesis, its many roles in all aspects of protein synthesis, and its role in protein localization. In addition, ribonucleotides serve many other functions in metabolism. Deoxyribonucleotides are synthesized by the reduction of ribonucleotides, and ATP is an intermediate in the synthesis of histidine. Ribonucleotides also serve as the principal currency of energy metabolism, as second messengers in signal transduction, and as cofactors in many metabolic reactions [13]. This last role has been cited frequently as a vestige of the ancient RNA world [7-10]. In coenzymes such as nicotinamide adenine dinucleotide (NADH), adenine is chemically inert and merely serves as a handle for binding enzymes. It is easy to imagine that nicotinamide could bind its enzymes just as well by itself, had the course of evolution run differently. Thus, these coenzymes might be molecular fossils of a largely extinct RNA-mediated metabolism [7-10].

By far the strongest support for the RNA world hypothesis are the universally conserved ribozymes, the peptidyl transferase activity of the ribosome and ribonuclease P (RNase P). Given that protein enzymes can catalyze both peptide synthesis and RNA hydrolysis, it becomes difficult to specify characteristics of RNA that would make it an inherently superior catalyst for these reactions without invoking the RNA world hypothesis.

At the heart of the RNA world hypothesis lies the idea that ribozymes synthesized the first proteins. It has long been thought that the peptidyl transferase activity of the ribosome is catalyzed by RNA, but it proved impossible to completely remove all proteins and observe activity with only the RNA [14]. The crystal structure of the 50S subunit removed all doubts that the ribosome is a ribozyme: there are no proteins within 18 Å of the active site [15, 16]. In fact, the ribosomal proteins decorate the outer surface of the ribosome and only insert thin tendrils which probably serve to stabilize the ribosome's structure. With the proteins in such a peripheral position, it is easy to envision the ribosome beginning as naked RNA and slowly accreting peptides as it evolved.

The other universally conserved ribozyme, RNase P, plays a less glamorous role in metabolism, but it also offers an evolutionary conundrum if the RNA world hypothesis is disregarded. One of the first ribozymes to be discovered, RNase P removes the 5′ leader sequence of pre-tRNA [17]. Even in the absence of the protein subunit, which is required for optimal activity, the RNA component from prokaryotes is one of the fastest ribozymes [18, 19]. There are many ribonucleases made out of protein, and it would be surprising if one evolved from RNA after the invention of protein enzymes—it would be much simpler for a protein enzyme to alter its specificity than for a new activity to arise from random sequence. It is also unclear why RNase P should still exist as an RNA enzyme, when almost all other enzymatic functions have been taken over by proteins. Perhaps the catalytic efficiency of RNase P was more comparable to that of proteins, compared to the efficiency of other ribozymes, so it was able to survive in the new protein world.

### *Exploring the RNA World*

The RNA world would have had to support a complex metabolism in the time leading up to the evolution of protein synthesis [12]. Protein synthesis itself requires a number of activities, including tRNA aminoacylation and peptidyl transferase. One of the central requirements of the RNA world hypothesis is that the RNA genome should be able to replicate itself, which would require an RNA polymerase and possibly several ribozymes for synthesizing and phosphorylating nucleotides. Enzyme cofactors containing nucleotides also provide clues about the kind of metabolism that could have existed in the last ribo-organism. Cofactors, such as NADH and coenzyme A, and metabolic intermediates, such as UDP-glucose, suggest that RNA world metabolism could have included oxidation/reduction reactions, lipid synthesis and carbohydrate metabolism [10].

The main experimental approach for investigating whether RNA is capable of catalyzing these reactions is to discover new ribozymes by using in vitro selection. Selections have been based on mutagenized pools of known ribozymes [20] or randomly synthesized sequences [21, 22]. In addition to discovering ribozymes with new activities, this technique has been used to isolate RNA molecules which bind small molecules or proteins [21], and to determine the sequence requirements for RNA-protein interactions [22]. In this technique, active molecules from a pool of ~$10^{15}$ different sequences are selected based on their ability to covalently modify themselves or to bind a ligand; the selected molecules are then amplified. The selection and amplification steps are repeated iteratively until a large fraction of the pool is active. Many ribozymes have been isolated from random sequence in this manner, including ribozymes that catalyze phosphoryl

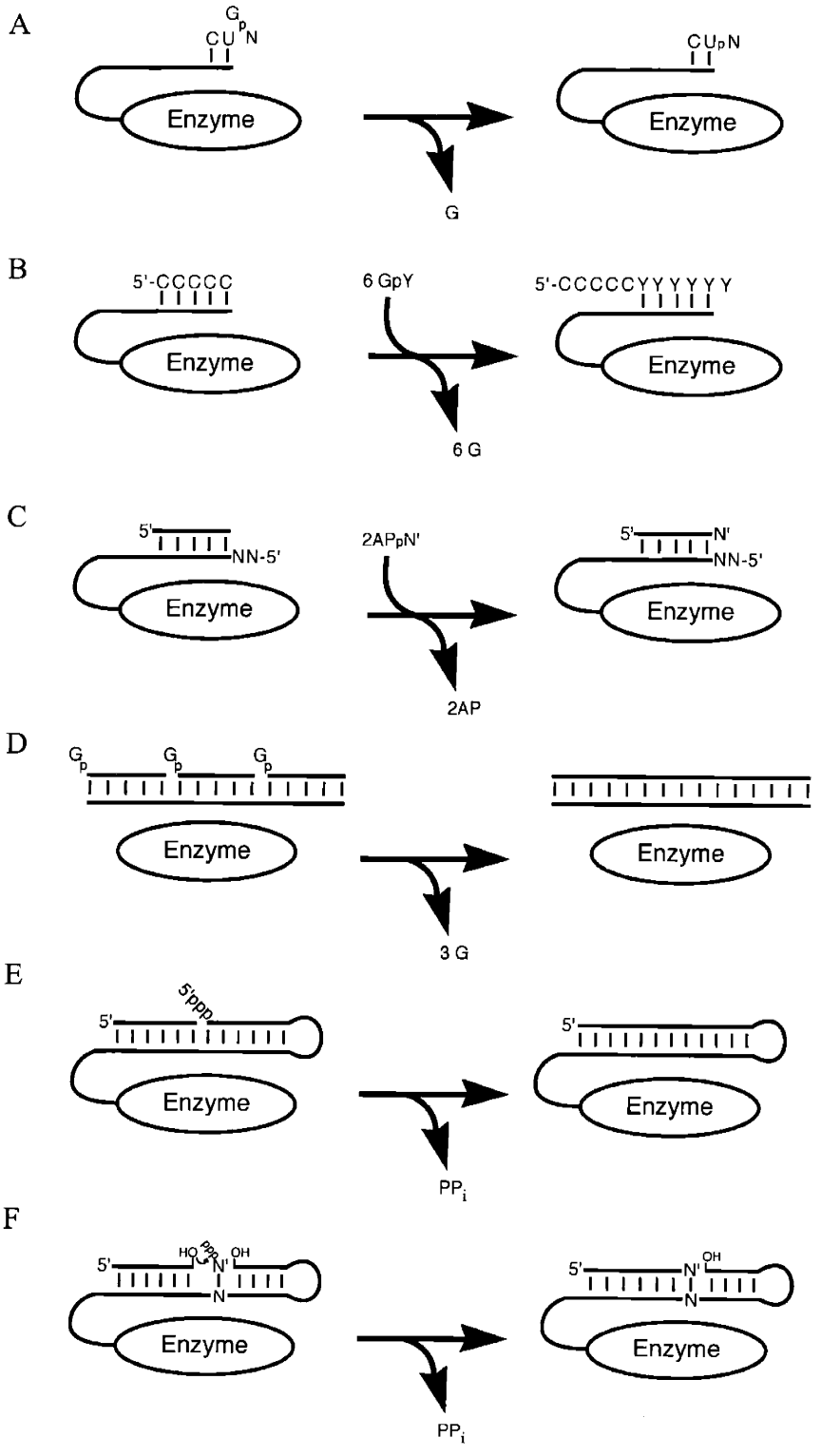
transfer, alkylation, aminoacylation, peptidyl transfer, glycosidic bond formation, and carbon-carbon bond formation [23-36].

## The Quest for an RNA Polymerase Ribozyme

### *Redesigning the Group I Intron*

Foremost among ribozyme activities required by the RNA world hypothesis is an RNA polymerase ribozyme capable of self-replication. Initial attempts to discover whether RNA could catalyze its own replication involved modification of the Group I intron. The first experiments demonstrated that the Group I intron could perform a primer extension reaction (Figure 1A), in which the 3′-hydroxyl of an oligonucleotide (pCpU) that is base-paired to the ribozyme attacks an internal phosphate of another oligonucleotide (GpN) [37, 38]. Only A, C, or U could be added to the primer because using the GpG dinucleotide resulted in cleavage, rather than extension. The Group I intron was also modified to extend a primer by several nucleotides, forming homopolymers as single nucleotides were added and the primer shifted register (Figure 1B) [39]. These reactions are more analogous to the second step in Group I intron splicing than to RNA polymerization catalyzed by proteins, however. A nucleoside is the leaving group, rather than pyrophosphate, and the nucleotide added to the primer is untemplated. In order to introduce template specificity, the Group I intron was modified so that the primer bound the primer-binding site in a single register and a template region was added 5′ to the binding site (Figure 1C) [40]. The guanosine binding site was also

**Figure 1.** RNA synthesis and ligation reactions catalzed by variants of the Group I intron. Adapted from ref. 41 and 51.



modified to recognize 2-aminopurine, so that the primer could be extended by a G. Templated primer extension was observed with any nucleotide, but with low fidelity (only 65% of the extensions were the Watson-Crick match).

In addition to low fidelity, this reaction was very inefficient unless the template was attached to the ribozyme [40]. This problem was overcome by requiring extension of the primer by oligonucleotides complementary to the template (Figure 1D) [42]. Oligonucleotides have a much higher affinity for the template, which facilitates the formation of the enzyme-substrate complex. This ribozyme construct appeared to demonstrate little sequence specificity for the template, an important quality for an RNA polymerase ribozyme [43]. After removing nonessential sequences from *SunY*, a small Group I intron, this shorter construct synthesized its entire complementary strand from 18 oligonucleotides of about 10 nucleotides each [44]. However, only a small fraction of the possible 10-nucleotide substrates could be included in the reaction, limiting the number of possible templates. Shorter oligonucleotides provided for more general polymerization; only 64 3-nucleotide substrates were required [45]. Unfortunately, fidelity varied greatly among the trinucleotides and was particularly low if the trinucleotide-template complex included a G·U wobble.

These studies demonstrated that ribozymes are capable of RNA synthesis, but two limitations prevent these results from firmly establishing the feasibility of general ribozyme-catalyzed RNA replication. First, the substrates for polymerization are oligonucleotides. This is a problem because it presupposes that there is another RNA polymerase ribozyme which synthesizes the oligonucleotides. Second, an RNA polymerase of a length comparable to the *SunY* intron must have a fidelity exceeding

99% (allowing 1 mistake per 100 nucleotides), or it will not produce active copies of itself [46]. The 65% fidelity of the Group I primer extension reaction would require a vast improvement in order to reach the fidelity of relatively inaccurate viral RNA-dependent RNA polymerases, which have fidelities of >99.6% [40, 47, 48].

To address these limitations, an in vitro selection was performed on a pool based on the Group I intron [49]. As a scaffold for stable folding and template binding, the independently folding P4-P6 domain was retained [50]. Two other helices were retained (P3 and P8), and these three pieces were linked together by a total of 85 random nucleotides. The resulting ligase ribozyme (hc ligase) shares 55% sequence identity with the Group I intron and also contains unique structural elements. The hc ligase catalyzes the formation of a 3′,5′-phosphodiester bond between adjacent template-bound oligonucleotides, using pyrophosphate as the leaving group (Figure 1E) [49]. After further selection, a variant of the hc ligase was isolated which has high affinity for a noncovalently bound hairpin substrate and performs ligation in a largely sequence independent manner [51]. In addition to ligation, this ribozyme performs a polymerase-like reaction, in which a mononucleotide triphosphate is ligated to the 3′-terminus of a primer (Figure 1F). The fidelity for GTP addition on a C template is >96%, which is a great improvement over previous engineered Group I introns; however, there is no detectable templated addition of ATP or UTP.

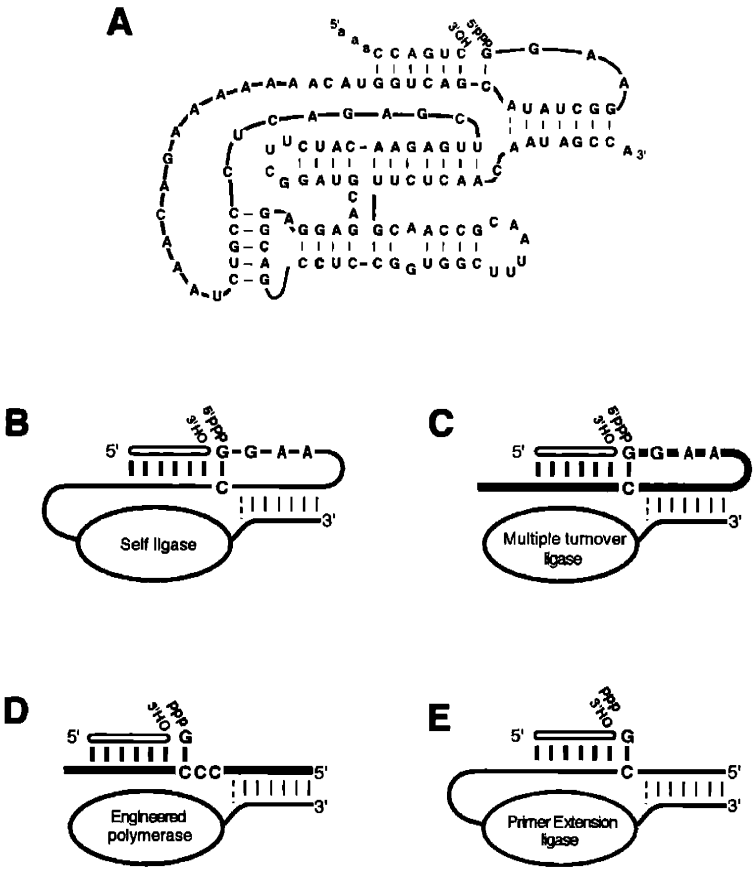
*A Ligase Ribozyme from Random Sequence*

In order to go beyond the limits imposed on redesigning a natural ribozyme, an in vitro selection experiment was performed to discover new ligase ribozymes, which use pyrophosphate as a leaving group instead of a nucleotide [24]. Isolation and optimization of active ribozymes resulted in the discovery of over 65 ribozymes. Seven were characterized and fell into three classes of ligase ribozymes [52, 53]. The class I ligase, represented by a single isolate, catalyzes the formation of a 3′,5′-phosphodiester bond between a primer and the ribozyme's 5′-terminus (Figure 2A,B).

Because the self-ligating ribozyme is extremely fast, a multiple-turnover derivative was generated, in which the template strand is detached from the ribozyme (Figure 2C) [53]. Kinetic analysis of this construct revealed a rate enhancement of $10^9$ over the uncatalyzed ligation rate [54]. In addition, its $k_{cat}$ of 100 $min^{-1}$ compares favorably with the $k_{cat}$ for *Escherichia coli* DNA ligase (28 $min^{-1}$), although its $K_m$ is almost 200 times higher [55]. The pH dependence of the multiple-turnover ligase is log-linear with a slope of 1.0 below pH 6.5, indicating that chemistry is rate-limiting up to this pH, and that substrate dissociation becomes rate-limiting as the pH increases [54]. Increasing the rate of product dissociation by changing the sequence of the stem to which it binds resulted in a much faster observed rate at high pH. The $k_{cat}$ for this construct is 360 $min^{-1}$ at pH 9.0, a rate that is comparable to that of RNase P RNA [19].

The extremely high observed rates of the ligase ribozyme make it an excellent model system for molecular evolution. Using a variant of the ligase, Wright and Joyce developed a continuous evolution system, in which active molecules were selected in a reaction mix containing reverse transcriptase, T7 RNA polymerase, NTPs and dNTPs, in

**Figure 2.** The class I ligase and its derivatives.



addition to the substrate [56]. In this selection scheme, the ribozyme was required to ligate itself to the T7 promoter before it was reverse transcribed; only molecules that reacted would have a promoter for transcription. Using this method, the ligase ribozyme could be kept in continuous culture by serial dilution to replenish the protein enzymes and mononucleotides. Because the RNA is not purified during the selection, continuous evolution provides a more realistic model of biological evolution, allowing analysis of population dynamics under different selective constraints [56]. For instance, continuous evolution was used to study heritability of phenotypes in a population of ligase molecules undergoing selection in an environment in which the magnesium concentration was progressively reduced [57]. Rather than selecting molecules that were more catalytically active in lower magnesium concentrations, molecules which were more likely to fold into an active conformation were selected. This highlights the fact that single RNA molecules have multiple phenotypes (catalytic activity, conformational stability, etc.) which could respond to selection.

The difficulty of designing a selection to enhance a particular characteristic was also demonstrated in a selection to improve the ligase ribozyme's rate of chemistry (N.H. Bergman, C.Y. Yen, D.P. Bartel, in preparation). In this selection, a pool of ligase variants were selected at low pH, where chemistry is rate limiting. The optimized ligase ribozyme that resulted was only 1.2- to 1.5-fold faster than the parent in 10 mM $Mg^{2+}$. However, the ribozyme was 50-fold faster than the parent in 0.5 mM $Mg^{2+}$, corresponding to a reduction in the number of magnesium ions required for activity (N.H.B., C.Y.Y., D.P.B., in preparation) [58]. This was surprising, because the selection was carried out in 10 mM $Mg^{2+}$, and the stringency was increased during the selection by

reducing the pH and the incubation time. It would be difficult to increase the stringency beyond that used in the last round; the pool was allowed to react for only 200 ms at pH 6.0. Although the improved ligase might be as optimal as possible, further improvements might result from optimizing catalysis in low magnesium concentrations (< 1 mM $Mg^{2+}$) or optimizing its rate of folding by performing a selection at high pH, where folding is rate limiting [58, 59].

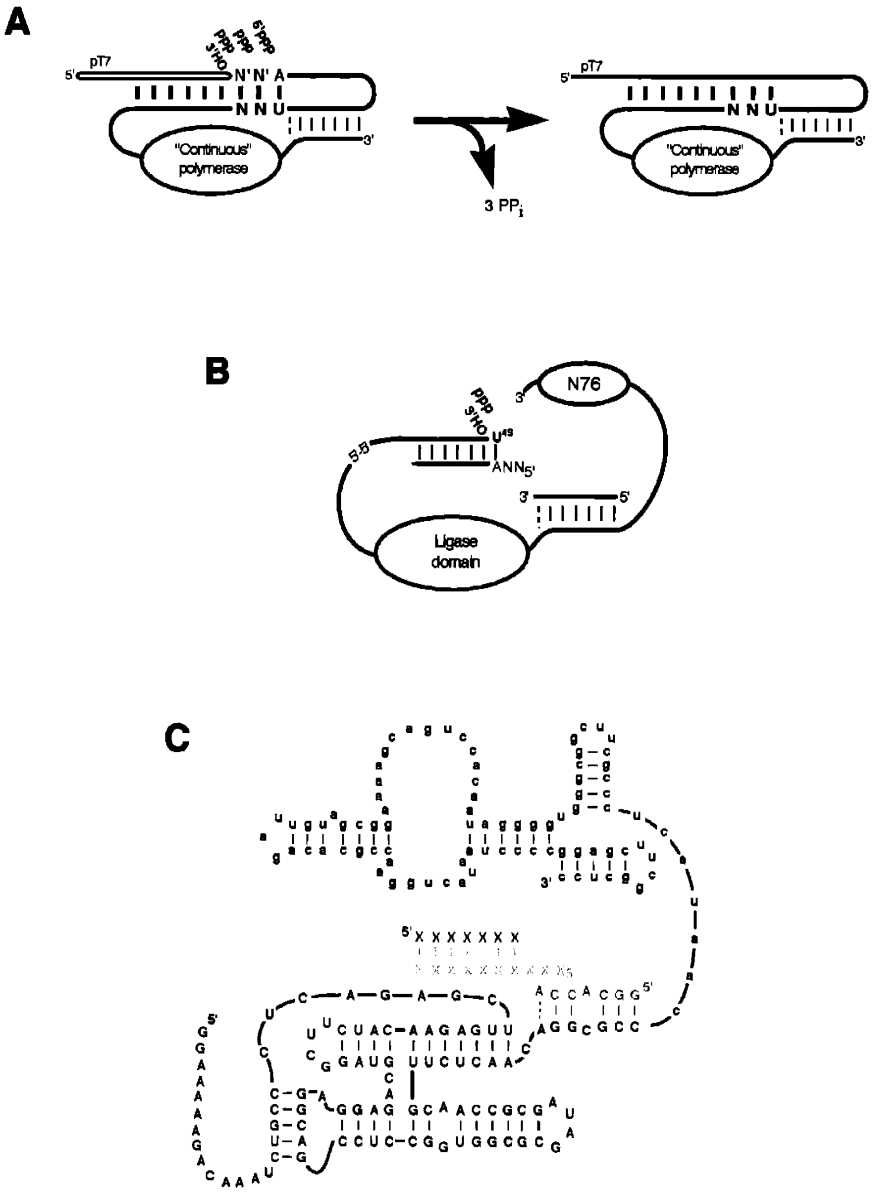
### *A Polymerase Derived from the Class I Ligase*

The class I ligase was converted into a rudimentary polymerase by deleting four nucleotides from the 5′-terminus and separating the template from the core of the ribozyme (Figure 2D) [60]. The engineered polymerase ribozyme could extend a primer by a single, templated nucleotide. The fidelity and efficiency of extension varied according to the template, with a C template giving the highest fidelity (0.996) and a U template giving the lowest (0.71). In the presence of equimolar concentrations of all four nucleotides the fidelity was 85%, and it could be raised to 92% by decreasing the GTP concentration 10-fold. This was a vast improvement over Group I intron primer extension, but still short of the fidelity required for self-replication. In order to observe true polymerization, the template region of this ribozyme was lengthened to three nucleotides. Although the reaction proceeds slowly, this was the first demonstration that RNA can catalyze a polymerization reaction that is identical to that catalyzed by protein-enzyme polymerases. Unfortunately, the template could not be lengthened beyond three nucleotides, because a segment of the template pairs with the ribozyme, forming a required stem.

Because of its limited capability to perform polymerization, the ligase ribozyme has been used as the basis for selecting polymerase ribozymes. In one strategy, a combination of stepwise and continuous evolution was used to isolate ribozymes that could perform three successive reactions, the templated addition of two mononucleotides followed by RNA ligation (Figure 3A) [61]. The resulting ribozymes catalyzed nucleotide addition in both 5′→3′ and 3′→5′ directions. While it is remarkable that these ribozymes perform a three-step reaction, the lack of directionality in nucleotide addition makes the reaction difficult to generalize. In addition, polymerization could not be extended beyond two nucleotides. Variants that were selected for their ability to add three nucleotides followed by RNA ligation evaded the intended selection pressure by evolving a mutant T7 promoter.

A different strategy was used to discover a polymerase ribozyme that has higher fidelity and does not base-pair to the template [62]. The pool was constructed by adding 76 nucleotides of random sequence to the 3′-terminus, randomizing two functionally unimportant loops in the ligase domain, and mixing four subpools in which the ligase domain had been mutagenized to varying degrees or not at all (Figure 3B). The primer was also attached to the 5′-terminus of the pool via a 5′,5′ linkage so that active molecules which had added 4-thio-uridine to the primer could be isolated by binding to mercury. The polymerase ribozyme that was isolated after 18 rounds of selection is an accurate RNA polymerase that can use any template (Figure 3C). The 3′-domain, which was derived from random sequence, is required for binding the primer-template complex, so that covalently attaching the primer to the ribozyme is not required. Remarkably, the ligase domain emerged after 18 rounds of selection with no major changes, suggesting

**Figure 3.** Polymerases derived from the class I ligase.



that it was already well-optimized for single-nucleotide addition. Although the ribozyme was selected for adding only two nucleotides to the primer, it is adept with longer templates and can extend a primer up to 14 nucleotides. Unexpectedly, polymerization is even somewhat processive, suggesting that the polymerase ribozyme will be a good candidate for further selection and optimization with long templates (M.S. Lawrence and D.P. Bartel, in preparation).

During the last few rounds of the selection, all four nucleotides were included in the reaction mix in order to select variants with higher fidelity. In fact, the fidelity of the polymerase ribozyme is greatly improved over its parent's. By measuring the error rates for single-nucleotide addition, the fidelity was determined to be 96.7%, compared to 85% for the parent. When primers that were fully extended by 11 nucleotides were sequenced, the fidelity appeared even higher (99%). This discrepancy was probably due to the fact that incorporating the wrong nucleotide slows further polymerization, so that sequences without mismatches were enriched among the fully extended primers.

## Of Metal Ions and Mechanisms

Selecting ribozymes with new activities has provided extensive information about the types of reactions that RNA can catalyze and the feasibility of the RNA world. RNA catalysts also present a unique opportunity to examine the evolution of biocatalysis in general. Many ribozymes have protein counterparts that perform the same reaction but have entirely different evolutionary histories. A comparison of catalytic strategies requires detailed understanding of the biochemistry of enzymes, but it promises to reveal

insights into the density of catalytic units in sequence space and the likelihood of mechanistic convergence.

***Metal Ions in Ribozyme Catalysis***

Until a few years ago it was thought that all ribozymes require metal ions, which promote catalysis by coordinating to the reacting atoms [63, 64]. The first indication that not all ribozymes are metalloenzymes came from the discovery that the self-cleaving hairpin ribozyme is active in $Co(NH_3)_6^{3+}$ or the polyamine spermine [65-68]. Because $Co(NH_3)_6^{3+}$ is similar in size and shape to a fully hydrated $Mg^{2+}$ ion but is exchange inert, this result meant that some ribozymes do not require direct coordination to a metal ion for catalysis and might use a different mechanism [65, 67, 69]. At about the same time, divalent metal-independent deoxyribozymes and ribozymes were discovered [70-72]. For instance, the deoxyribozyme is equally active in $Na^+$, $K^+$, and $NH_4^+$, and divalent metals have no effect on its activity [70]. In addition, the natural self-cleaving ribozymes, hammerhead, hairpin, Varkud satellite (VS), and hepatitis delta virus (HDV), are also catalytically active in molar concentrations of monovalent cations [73, 74]. These experiments and many others with both natural ribozymes and ribozymes selected in vitro have demonstrated that the metal ion requirements of ribozymes are very diverse.

Despite these examples, most ribozymes do require metal ions for electrostatic shielding, structure, and optimal catalysis [75]. Because RNA is a polyanion, counterions are required for charge neutralization, allowing the RNA to form the compact tertiary structure required for catalysis [76]. As a result, there are 2-4 nonspecifically bound

metal ions per nucleotide [77-82]. These low affinity cations rapidly exchange with ions in the solvent, making them difficult to localize by biophysical techniques [76].

A few metal ions bind to specific sites to participate in catalysis, stabilize the tertiary structure, or both. These ions form discrete contacts to the RNA, so the identity of the metal ion is important. Ribozymes vary in their specificity for metal ions in these sites, depending on the importance of the ionic radius, the metal's coordination geometry (i.e., the preferred number of bound ligands), the identity of the coordinating ligands, and other factors. In contrast to the metal ions found in the active sites of most protein metalloenzymes, RNA-binding metal ions are not fully dehydrated and may bind the RNA via direct, inner-sphere coordination or by outer-sphere coordination mediated through bound water molecules [76].

In general, the larger RNAs have more specific metal ion requirements than smaller ribozymes [76]. For instance, the large ribozymes (Group I intron, Group II intron, and RNase P) have an absolute requirement for divalent metal ions [18, 83-85]. Whereas the Group I and Group II introns are only active in $Mg^{2+}$ and $Mn^{2+}$, RNase P also exhibits some activity in $Ca^{2+}$. In contrast, the small nucleolytic ribozymes (hammerhead, hairpin, VS, and HDV) are active in a much wider set of ionic conditions, including both divalent and monovalent cations [65-68, 86-90].

Ribozymes selected in vitro exhibit an even wider range of metal ion requirements. Some, like the ribonucleolytic leadzyme, are active only in the cations in which they were selected [91-93]. Others exhibit activity in the presence of metal ions which were not present during their selection. For example, an acyl transferase ribozyme, which was selected in the presence of $Mg^{2+}$ and $K^{+}$ only, is active in a wide

variety of cations, including $Co(NH_3)_6^{3+}$ [31, 94]. Several selections have attempted to find ribozymes with unusual metal ion specificities, such as the metal-independent ribozymes discussed above [70-72]. These selections resulted in the isolation of nucleic acid enzymes which are active in many monovalent and divalent metal ions, including $Ca^{2+}$, $Mn^{2+}$, $Cu^{2+}$, and $Zn^{2+}$ [28, 35, 70-72, 91, 94-99].

Since these experiments established that ribozymes can be selected to function under virtually any ionic condition, the question becomes what are the functions of those metal ions in structure and catalysis. Of the many methods used to identify roles of specific metal ions, crystallography has provided unparalled insight, especially into the structural roles of metal ions. For example, the structure of the P4-P6 domain of the Group I intron demonstrated that metal ions cluster between elements of tertiary structure, allowing surprisingly dense packing of the polyanionic backbone [100].

One of the most powerful biochemical methods for discerning the functions of metal ions involves replacing oxygen atoms with sulfur and observing the effect on catalysis in the presence of various metal ions. Replacing an oxygen with a sulfur atom disrupts $Mg^{2+}$ binding because $Mg^{2+}$ has a very low affinity for sulfur relative to its affinity for oxygen; transition metals such as $Mn^{2+}$, $Cd^{2+}$ and $Zn^{2+}$ have much higher affinities for sulfur [101]. The activity of ribozymes with sulfur substitutions at $Mg^{2+}$ binding sites is drastically reduced, and restoring activity with a thiophilic metal ion provides strong evidence for metal ion binding to that oxygen via inner-sphere coordination [76, 102]. These metal "rescue" experiments have been used to test metal binding to single sites, particularly in the active sites of the hammerhead, Group I, RNase P, HDV, and hairpin ribozymes [65-67, 86, 102-104]. It has also been used to

simultaneously screen all possible sulfur substitutions at nonbridging pro-$R_P$ oxygens by cotranscriptionally incorporating phosphorothioate-containing nucleotides, separating active from inactive molecules, and chemically cleaving the RNA at phosphorothioate linkages [105-107]. Comparing the cleavage of the active versus inactive molecules reveals a relative reduction of cleavage of active molecules at sites where the sulfur substitution interferes with activity. In the presence of thiophilic metal ions, an increase in cleavage at that site in the active molecules provides evidence for metal binding to that site. Metal ions involved in both structure and catalysis can be defined using this method, as long as they bind through inner-sphere coordination.

***Mechansims of Ribozyme Catalysis***

The diversity in metal ion requirements among ribozymes suggests that they may use a variety of catalytic strategies. It was initially thought that RNA catalysis would be limited by the fact there are only four chemically-similar nucleotides, compared to the 20 functionally-diverse amino acids [108]. Many selections have taken this into account and provided various cofactors or modified nucleotides to increase functionality. Ribozymes that require histidine, presumably as a general base, have been isolated [109], and deoxyribozymes that require $Cu^{2+}$ might cleave DNA by an oxidation/reduction mechanism [95]. Two groups have isolated ribozymes that perform Deils-Alder condensations from pools with or without modified nucleotides. Tarasow et al. failed to isolate active ribozymes from a pool of unmodified RNA, but succeeded with a pool that contained pyridyl-modified uridines [35]. However, Seelig and Jaschke isolated a ribozyme that performs the same reaction using a pool containing only natural

nucleotides [36]. The rate enhancement of the unmodified ribozyme was actually 25-fold higher than the pyridyl-uridine ribozyme, demonstrating that catalysis by natural RNA had not been fully explored.

The mechanisms of most ribozymes have not been studied thoroughly, and only detailed kinetic and structural analysis can determine the true diversity of mechanistic strategies used by RNA. Natural ribozymes which catalyze phosphoryl transfer have similar reaction pathways, involving nucleophilic attack by a 2′- or 3′-hydroxyl on phosphorous, which generates a pentacovalent transition state and inversion of configuration [110]; however they appear to use a variety of catalytic strategies.

The Group I intron, Group II intron, and RNase P appear to be true metalloenzymes, exhibiting no activity in the absence of divalent metal ions [18, 83-85]. The Group I intron has been the most extensively characterized, and three active site metal ions have been identified by thiophilic metal ion rescue experiments [102, 111-114]. The precise functions of metal ions for the Group II intron and RNase P have not been determined, so it is unknown how their mechanisms differ from that of the Group I intron. At least one active site metal ion has been discovered for each, and kinetic analysis has suggested that RNase P may require three metal ions [103, 115, 116].

Although the small nucleolytic ribozymes catalyze virtually identical reactions, they appear to use radically different mechanisms. The mechanism of the hammerhead ribozyme has been particularly enigmatic. The locations of several divalent metal ions have been determined by crystallographic and biophysical methods, leading to models in which there are either one or two catalytic divalent metal ions [110]. However, much of this data has been interpreted in contradictory ways, and crystallography has failed to

produce a structure in a likely transition state conformation with the 2′-hydroxyl poised for in-line attack, nor has it provided evidence of a catalytic metal ion [110, 117]. In addition, the hammerhead is active in monovalent cations, including $NH_4^+$, indicating that it does not absolutely require divalent cations or even metal ions for a minimal level of activity [73]. Further study of the role of monovalent cations suggested that specific metal ion coordination is not required for catalysis, but cations are merely required to achieve an active structure [90, 118]. Reconciling all the data concerning metal ions in hammerhead cleavage suggests a nonobligatory but stimulatory role for divalent metal ions, but the exact mechanism of the hammerhead remains a mystery.

In contrast to the controversy over the hammerhead ribozyme's mechanism, understanding the mechanism of the HDV ribozyme has revolutionized ideas about RNA catalysis. Biochemical experiments guided by the ribozyme's crystal structure demonstrated that the HDV ribozyme uses an acid/base mechanism [74, 119, 120]. In this mechanism, a cytosine acts as the acid, and a hydrated metal hydroxide acts as the base [74]. That a nucleotide could serve as an acid catalyst was surprising, because the $pK_a$s of nucleotides are far from neutral [121]. In fact, most of the rate enhancement for the reaction arises from the nucleotide acid catalyst, and the metal hydroxide only provides a modest 25-fold increase in rate [122].

Crystal structures of the hairpin ribozyme and the ribosome suggest that the HDV ribozyme may not be alone in using a nucleobase catalyst. There are no metal ions in the active site of the hairpin ribozyme; instead, it is surrounded by four purines which could contribute to catalysis [123]. Several groups have investigated these residues and determined that two of them have perturbed $pK_a$s [124-126]. In addition to proposing

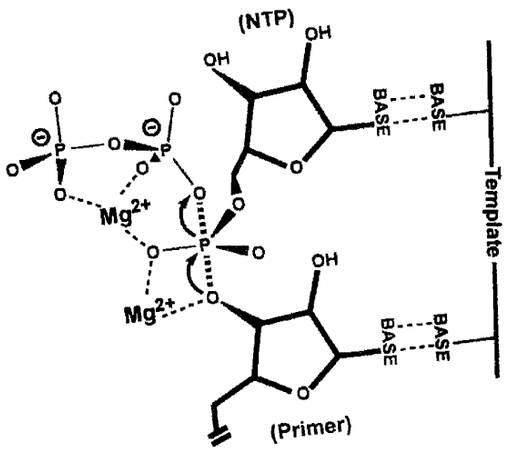
one of these as a candidate for acid/base catalysis [124], they have also been proposed to electrostatically stabilize the transition state or activate the nucleophilic 2′-hydroxyl [124-126]. While more work needs to be done to assess these possibilities, it is apparent that the hairpin ribozyme has a somewhat different mechanism than the HDV ribozyme. In the active site of the ribosome, the extremely well-conserved base A2451 would be perfectly positioned to act as an acid-base catalyst, if it were protonated [16]. Strong biochemical evidence for this has not been forthcoming, and it is possible that the ribosome catalyzes peptidyl transfer solely by proper positioning of the substrates [127-131].

### *A Mechanism for ALL Polymerases?*

It is now apparent that ribozymes are capable of a variety of reaction mechanisms, utilizing metal ions, cofactors, and their own functional groups, just like protein enzymes. Because they have different evolutionary histories from proteins which catalyze similar reactions, the study of ribozyme catalysis offers insights about the variety of mechanisms possible for a given reaction. Of particular interest is the mechanism of the class I ligase and the polymerase derived from it, because these ribozymes catalyze exactly the same reaction as polymerases made of protein.

Even though some protein-enzyme polymerases have no identifiable sequence similarity, all of them use the same mechanism requiring two metal ions (Figure 4) [132-135]. In this mechanism, one metal ion coordinates the 3′-OH of the primer and the pro-$R_P$-oxygen of the α-phosphate belonging to the incoming nucleotide [134], and the other contacts the nonbridging oxygens of the triphosphate, all through inner-sphere

**Figure 4.** The two-metal ion mechanism. Adapted from ref. 132.



coordination [133-135]. The first metal ion promotes the attack of the 3′-OH on the α-phosphate by lowering its affinity for hydrogen, and the second stabilizes the pyrophosphate leaving group [132]. This mechanism is common to several nonhomologous polymerases [132-135], as well as to enzymes that catalyze other phosphoryl transfer reactions (albeit with differing stereochemical preferences), including the 3′–5′ exonuclease of DNA polymerase I and alkaline phosphatase [136-139].

Previously it was suggested that this mechanism might also apply to ribozyme-mediated catalysis, particularly the reaction catalyzed by the Group I intron [111, 112, 136]. It has now been shown that the Group I intron requires three metal ions for catalysis in a mechanism similar but distinct from the two-metal ion mechanism, [113, 114]. Unlike the Group I intron, however, the ligase ribozyme performs the same reaction as polymerases made of protein. Understanding the mechanism of the ligase ribozyme will determine if the two-metal ion mechanism is universal for all polymerases, even if not for all ribozymes.

## Summary: Characterization of the Ligase Ribozyme

The ligase ribozyme is not only the basis for a polymerase ribozyme which addresses the feasibility of the RNA world hypothesis, but it is kinetically equal or superior to natural ribozymes which have had billions of years to evolve [54, 58]. As such, there were several reasons to undertake a detailed characterization of its substrate and metal ion requirements. Our main goals were to discover aspects of the ribozyme's catalysis that could be improved by in vitro selection and to determine if there were any

characteristics that may hinder continued attempts to improve the polymerase ribozyme. Additionally, these studies provide data for an in-depth comparison to other RNA- and protein-enzymes and lay the groundwork for future structural and mechanistic studies.

These studies began with a characterization of nucleoside triphosphate recognition by the ligase ribozyme (Figure 2E). The construct used in these experiments was derived from the self-ligating ribozyme by deleting four nucleotides from the 5′-terminus, and it catalyzes the extension of a primer by a single templated nucleotide. As suggested by the engineered polymerase's high fidelity [60], the ligase recognizes nucleotides almost exclusively through Watson-Crick base-pairing. However, primer extension was 100-fold slower than the self-ligation reaction of the parent ribozyme. This rate reduction was traced to the deletion of the second and third nucleotides of the self-ligating ribozyme. When these nucleotides were restored by supplying a trinucleotide substrate, pppGGA, the rate of the reaction was nearly as fast as self-ligation, even though primer extension is a trimolecular reaction.

As an initial step in the mechanistic characterization of the ligase, the metal ion requirements of the self-ligating ribozyme were also studied. The ligase is unique among ribozymes, in that it has an extreme specificity for $Mg^{2+}$. This is not surprising, considering that it experienced only $Mg^{2+}$ and $K^{+}$ during its selection, and there was no pressure to become more metal-tolerant. In spite of the fact that the ligase is inhibited by all cations other than $Mg^{2+}$, it is well-folded in $Mg^{2+}$, $Ca^{2+}$ and $Co(NH_3)_6^{3+}$. This will be useful in future attempts to crystallize the ligase, both for incorporating electron-dense metal ions and for crystallizing unreacted ribozyme. Kinetic and structural studies suggested that the ligase may require two metal ions for catalysis, providing a tentative

clue that it may use a mechanism similar to that of RNA and DNA polymerases made out of protein.

The detailed overview of catalysis by the ligase ribozyme presented here will be valuable for designing experiments to improve the polymerase ribozyme and to understand the ligase's mechanism. In addition to suggesting ways in which the polymerase might be improved, these results provide a benchmark for measuring the polymerase's progress during future selections. Finally, these experiments provide a basic framework for further mechanistic studies which will greatly add to the rather limited understanding of ribozyme catalytic strategies.

## References


1. Rich, A. (1962). On the Problems of Evolution and Biochemical Information Transfer. In *Horizons in Biochemistry* (Kasha, M. & Pullman, B., eds.). pp. 103-126, Academic Press: New York.
2. Woese, C.R. (1967). The Genetic Code: The Molecular Basis for Genetic Expression, pp. 179-195.
3. Crick, F.H. (1968). The origin of the genetic code. *J Mol Biol* **38**, pp. 367-379.
4. Orgel, L.E. (1968). Evolution of the genetic apparatus. *J Mol Biol* **38**, pp. 381-393.
5. Sharp, P.A. (1985). On the origin of RNA splicing and introns. *Cell* **42**, pp. 397-400.
6. Gilbert, W. (1986). The RNA World. *Nature* **319**, pp. 618.
7. White, H.B. (1976). Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, pp. 101-104.
8. Lamond, A.I. & Gibson, T.J. (1990). Catalytic RNA and the origin of genetic systems. *Trends Genet* **6**, pp. 145-149.
9. Bloch, K. (1996). Some biochemical thoughts on the RNA world. *Chem Biol* **3**, pp. 405-407.
10. Jeffares, D.C., Poole, A.M. & Penny, D. (1998). Relics from the RNA world. *J Mol Evol* **46**, pp. 18-36.
11. Gesteland, R.F. & Atkins, J.F. eds. (1993). *The RNA World.* Cold Spring Harbor Laboratory Press: Cold Spring Harbor.
12. Bartel, D.P. & Unrau, P.J. (1999). Constructing an RNA world. *Trends Cell Biol.* **9**, pp. M9-M13.
13. Stryer, L. (1995). *Biochemistry*. W.H. Freeman and Company: New York.
14. Noller, H.F., Hoffarth, V. & Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**, pp. 1416-1419.
15. Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, pp. 905-920.
16. Nissen, P., Hansen, J., Ban, N., Moore, P.B. & Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, pp. 920-930.
17. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, pp. 849-857.
18. Gardiner, K.J., Marsh, T.L. & Pace, N.R. (1985). Ion dependence of the Bacillus subtilis RNase P reaction. *J Biol Chem* **260**, pp. 5415-5419.
19. Beebe, J.A. & Fierke, C.A. (1994). A kinetic mechanism for cleavage of precursor tRNA(Asp) catalyzed by the RNA component of Bacillus subtilis ribonuclease P. *Biochemistry* **33**, pp. 10294-10304.
20. Robertson, D.L. & Joyce, G.F. (1990). Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, pp. 467-468.
21. Ellington, A.D. & Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, pp. 818-822.

22. Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, pp. 505-510.
23. Williams, K.P., Ciafre, S. & Tocchini-Valentini, G.P. (1995). Selection of novel Mg(2+)-dependent self-cleaving ribozymes. *Embo J* **14**, pp. 4551-4557.
24. Bartel, D.P. & Szostak, J.W. (1993). Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* **261**, pp. 1411-1418.
25. Hager, A.J. & Szostak, J.W. (1997). Isolation of novel ribozymes that ligate AMP-activated RNA substrates. *Chem Biol* **4**, pp. 607-617.
26. Lorsch, J.R. & Szostak, J.W. (1994). In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, pp. 31-36.
27. Chapman, K.B. & Szostak, J.W. (1995). Isolation of a ribozyme with 5'-5' ligase activity. *Chem Biol* **2**, pp. 325-333.
28. Huang, F. & Yarus, M. (1997). 5'-RNA self-capping from guanosine diphosphate. *Biochemistry* **36**, pp. 6557-6563.
29. Wilson, C. & Szostak, J.W. (1995). In vitro evolution of a self-alkylating ribozyme. *Nature* **374**, pp. 777-782.
30. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. (1995). Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, pp. 643-647.
31. Lohse, P.A. & Szostak, J.W. (1996). Ribozyme-catalysed amino-acid transfer reactions. *Nature* **381**, pp. 442-444.
32. Jenne, A. & Famulok, M. (1998). A novel ribozyme with ester transferase activity. *Chem Biol* **5**, pp. 23-34.
33. Zhang, B. & Cech, T.R. (1997). Peptide bond formation by in vitro selected ribozymes. *Nature* **390**, pp. 96-100.
34. Unrau, P.J. & Bartel, D.P. (1998). RNA-catalysed nucleotide synthesis. *Nature* **395**, pp. 260-263.
35. Tarasow, T.M., Tarasow, S.L. & Eaton, B.E. (1997). RNA-catalysed carbon-carbon bond formation. *Nature* **389**, pp. 54-57.
36. Seelig, B. & Jaschke, A. (1999). A small catalytic RNA motif with Diels-Alderase activity. *Chem Biol* **6**, pp. 167-176.
37. Zaug, A.J. & Cech, T.R. (1986). The intervening sequence RNA of Tetrahymena is an enzyme. *Science* **231**, pp. 470-475.
38. Kay, P.S. & Inoue, T. (1987). Catalysis of splicing-related reactions between dinucleotides by a ribozyme. *Nature* **327**, pp. 343-346.
39. Been, M.D. & Cech, T.R. (1988). RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the Tetrahymena ribozyme. *Science* **239**, pp. 1412-1416.
40. Bartel, D.P., Doudna, J.A., Usman, N. & Szostak, J.W. (1991). Template-directed primer extension catalyzed by the Tetrahymena ribozyme. *Mol Cell Biol* **11**, pp. 3390-3394.
41. Bartel, D.P. (1999). Re-creating an RNA replicase. In *The RNA World* (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds.), 2nd ed. pp. 143-162, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
42. Doudna, J.A. & Szostak, J.W. (1989). RNA-catalysed synthesis of complementary-strand RNA. *Nature* **339**, pp. 519-522.

43. Doudna, J.A., Couture, S. & Szostak, J.W. (1991). A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science* **251**, pp. 1605-1608.
44. Green, R. & Szostak, J.W. (1992). Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science* **258**, pp. 1910-1915.
45. Doudna, J.A., Usman, N. & Szostak, J.W. (1993). Ribozyme-catalyzed primer extension by trinucleotides: a model for the RNA-catalyzed replication of RNA. *Biochemistry* **32**, pp. 2111-2115.
46. Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, pp. 465-523.
47. Steinhauer, D.A. & Holland, J.J. (1986). Direct method for quantitation of extreme polymerase error frequencies at selected single base sites in viral RNA. *J Virol* **57**, pp. 219-228.
48. Ward, C.D. & Flanegan, J.B. (1992). Determination of the poliovirus RNA polymerase error frequency at eight sites in the viral genome. *J Virol* **66**, pp. 3784-3793.
49. Jaeger, L., Wright, M.C. & Joyce, G.F. (1999). A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc Natl Acad Sci U S A* **96**, pp. 14712-14717.
50. Zarrinkar, P.P. & Williamson, J.R. (1994). Kinetic intermediates in RNA folding. *Science* **265**, pp. 918-924.
51. McGinness, K.E. & Joyce, G.F. (2002). RNA-Catalyzed RNA Ligation on an External RNA Template. *Chem Biol* **9**, pp. 297-307.
52. Ekland, E.H. & Bartel, D.P. (1995). The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res* **23**, pp. 3231-3238.
53. Ekland, E.H., Szostak, J.W. & Bartel, D.P. (1995). Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, pp. 364-370.
54. Bergman, N.H., Johnston, W.K. & Bartel, D.P. (2000). Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry* **39**, pp. 3115-3123.
55. Modrich, P. & Lehman, I.R. (1973). Deoxyribonucleic acid ligase. A steady state kinetic analysis of the reaction catalyzed by the enzyme from Escherichia coli. *J Biol Chem* **248**, pp. 7502-7511.
56. Wright, M.C. & Joyce, G.F. (1997). Continuous in vitro evolution of catalytic function. *Science* **276**, pp. 614-617.
57. Schmitt, T. & Lehman, N. (1999). Non-unity molecular heritability demonstrated by continuous evolution in vitro. *Chem Biol* **6**, pp. 857-869.
58. Glasner, M.E., Bergman, N.H. & Bartel, D.P. (2002). Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry* **41**, pp. 8103-8112.
59. Bergman, N.H. (2001). *The Reaction Kinetics and Three-Dimensional Architecture of a Catalytic RNA* (Doctoral) Biology, Massachusetts Institute of Technology: Cambridge, MA.
60. Ekland, E.H. & Bartel, D.P. (1996). RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **383**, pp. 192.

61. McGinness, K.E., Wright, M.C. & Joyce, G.F. (2002). Continuous In Vitro Evolution of a Ribozyme that Catalyzes Three Successive Nucleotidyl Addition Reactions. *Chem Biol* **9**, pp. 585-596.
62. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, pp. 1319-1325.
63. Yarus, M. (1993). How many catalytic RNAs? Ions and the Cheshire cat conjecture. *Faseb J* **7**, pp. 31-39.
64. Pyle, A.M. (1993). Ribozymes: a distinct class of metalloenzymes. *Science* **261**, pp. 709-714.
65. Hampel, A. & Cowan, J.A. (1997). A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. *Chem Biol* **4**, pp. 513-517.
66. Nesbitt, S., Hegg, L.A. & Fedor, M.J. (1997). An unusual pH-independent and metal-ion-independent mechanism for hairpin ribozyme catalysis. *Chem Biol* **4**, pp. 619-630.
67. Young, K.J., Gill, F. & Grasby, J.A. (1997). Metal ions play a passive role in the hairpin ribozyme catalysed reaction. *Nucleic Acids Res* **25**, pp. 3760-3766.
68. Earnshaw, D.J. & Gait, M.J. (1998). Hairpin ribozyme cleavage catalyzed by aminoglycoside antibiotics and the polyamine spermine in the absence of metal ions. *Nucleic Acids Res* **26**, pp. 5551-5561.
69. Jou, R. & Cowan, J.A. (1991). Ribonuclease H activation by inert transition-metal complexes. Mechanistic probes for metallocofactors: insights on the metallobiochemistry of divalent magnesium ion. *J Am Chem Soc* **113**, pp. 6685-6686.
70. Geyer, C.R. & Sen, D. (1997). Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme. *Chem Biol* **4**, pp. 579-593.
71. Jayasena, V.K. & Gold, L. (1997). In vitro selection of self-cleaving RNAs with a low pH optimum. *Proc Natl Acad Sci U S A* **94**, pp. 10612-10617.
72. Faulhammer, D. & Famulok, M. (1997). Characterization and divalent metal-ion dependence of in vitro selected deoxyribozymes which cleave DNA/RNA chimeric oligonucleotides. *J Mol Biol* **269**, pp. 188-202.
73. Murray, J.B., Seyhan, A.A., Walter, N.G., Burke, J.M. & Scott, W.G. (1998). The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem Biol* **5**, pp. 587-595.
74. Nakano, S., Chadalavada, D.M. & Bevilacqua, P.C. (2000). General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**, pp. 1493-1497.
75. Pyle, A.M. (1996). Role of metal ions in ribozymes. *Met Ions Biol Syst* **32**, pp. 479-520.
76. Feig, A.L. & Uhlenbeck, O.C. (1999). The Role of Metal Ions in RNA Biochemistry. In *The RNA World* (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds.), 2nd ed. pp. 287-319, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
77. Rialdi, G., Levy, J. & Biltonen, R. (1972). Thermodynamic studies of transfer ribonucleic acids. I. Magnesium binding to yeast phenylalanine transfer ribonucleic acid. *Biochemistry* **11**, pp. 2472-2479.

78. Romer, R. & Hach, R. (1975). tRNA conformation and magnesium binding. A study of a yeast phenylalanine-specific tRNA by a fluorescent indicator and differential melting curves. *Eur J Biochem* **55**, pp. 271-284.
79. Bina-Stein, M. & Stein, A. (1976). Allosteric interpretation of Mg2+ binding to the denaturable Escherichia coli tRNAGlu2+. *Biochemistry* **15**, pp. 3912-3917.
80. Stein, A. & Crothers, D.M. (1976). Equilibrium binding of magnesium(II) by Escherichia coli tRNAfMet. *Biochemistry* **15**, pp. 157-160.
81. Reid, S.S. & Cowan, J.A. (1990). Biostructural chemistry of magnesium ion: characterization of the weak binding sites on tRNA(Phe)(yeast). Implications for conformational change and activity. *Biochemistry* **29**, pp. 6025-6032.
82. Beebe, J.A., Kurz, J.C. & Fierke, C.A. (1996). Magnesium ions are required by Bacillus subtilis ribonuclease P RNA for both binding and cleaving precursor tRNAAsp. *Biochemistry* **35**, pp. 10493-10505.
83. Grosshans, C.A. & Cech, T.R. (1989). Metal ion requirements for sequence-specific endoribonuclease activity of the Tetrahymena ribozyme. *Biochemistry* **28**, pp. 6888-6894.
84. Deme, E., Nolte, A. & Jacquier, A. (1999). Unexpected metal ion requirements specific for catalysis of the branching reaction in a group II intron. *Biochemistry* **38**, pp. 3157-3167.
85. Guerrier-Takada, C., Haydock, K., Allen, L. & Altman, S. (1986). Metal ion requirements and other aspects of the reaction catalyzed by M1 RNA, the RNA subunit of ribonuclease P from Escherichia coli. *Biochemistry* **25**, pp. 1509-1515.
86. Dahm, S.C. & Uhlenbeck, O.C. (1991). Role of divalent metal ions in the hammerhead RNA cleavage reaction. *Biochemistry* **30**, pp. 9464-9469.
87. Chowrira, B.M., Berzal-Herranz, A. & Burke, J.M. (1993). Ionic requirements for RNA binding, cleavage, and ligation by the hairpin ribozyme. *Biochemistry* **32**, pp. 1088-1095.
88. Collins, R.A. & Olive, J.E. (1993). Reaction conditions and kinetics of self-cleavage of a ribozyme derived from Neurospora VS RNA. *Biochemistry* **32**, pp. 2795-2799.
89. Suh, Y.A., Kumar, P.K., Taira, K. & Nishikawa, S. (1993). Self-cleavage activity of the genomic HDV ribozyme in the presence of various divalent metal ions. *Nucleic Acids Res* **21**, pp. 3277-3280.
90. Curtis, E.A. & Bartel, D.P. (2001). The hammerhead cleavage reaction in monovalent cations. *Rna* **7**, pp. 546-552.
91. Pan, T. & Uhlenbeck, O.C. (1992). In vitro selection of RNAs that undergo autolytic cleavage with Pb2+. *Biochemistry* **31**, pp. 3887-3895.
92. Pan, T., Dichtl, B. & Uhlenbeck, O.C. (1994). Properties of an in vitro selected Pb2+ cleavage motif. *Biochemistry* **33**, pp. 9561-9565.
93. Sugimoto, N. & Ohmichi, T. (1996). Site-specific cleavage reaction catalyzed by leadzyme is enhanced by combined effect of lead and rare earth ions. *FEBS Lett* **393**, pp. 97-100.
94. Suga, H., Cowan, J.A. & Szostak, J.W. (1998). Unusual metal ion catalysis in an acyl-transferase ribozyme. *Biochemistry* **37**, pp. 10118-10125.
95. Carmi, N., Shultz, L.A. & Breaker, R.R. (1996). In vitro selection of self-cleaving DNAs. *Chem Biol* **3**, pp. 1039-1046.

96. Santoro, S.W. & Joyce, G.F. (1998). Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry* **37**, pp. 13330-13342.
97. Breaker, R.R. & Joyce, G.F. (1994). A DNA enzyme that cleaves RNA. *Chem Biol* **1**, pp. 223-229.
98. Cuenoud, B. & Szostak, J.W. (1995). A DNA metalloenzyme with DNA ligase activity. *Nature* **375**, pp. 611-614.
99. Landweber, L.F. & Pokrovskaya, I.D. (1999). Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: the spandrels of RNA evolution. *Proc Natl Acad Sci U S A* **96**, pp. 173-178.
100. Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. & Doudna, J.A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, pp. 1678-1685.
101. Pecoraro, V.L., Hermes, J.D. & Cleland, W.W. (1984). Stability constants of $Mg^{2+}$ and $Cd^{2+}$ complexes of adenine nucleotides and thionucleotides and rate constants for formation and dissociation of MgATP and MgADP. *Biochemistry* **23**, pp. 5262-5271.
102. Piccirilli, J.A., Vyle, J.S., Caruthers, M.H. & Cech, T.R. (1993). Metal ion catalysis in the Tetrahymena ribozyme reaction. *Nature* **361**, pp. 85-88.
103. Warnecke, J.M., Furste, J.P., Hardt, W.D., Erdmann, V.A. & Hartmann, R.K. (1996). Ribonuclease P (RNase P) RNA is converted to a Cd(2+)-ribozyme by a single Rp-phosphorothioate modification in the precursor tRNA at the RNase P cleavage site. *Proc Natl Acad Sci U S A* **93**, pp. 8924-8928.
104. Fauzi, H., Kawakami, J., Nishikawa, F. & Nishikawa, S. (1997). Analysis of the cleavage reaction of a trans-acting human hepatitis delta virus ribozyme. *Nucleic Acids Res* **25**, pp. 3124-3130.
105. Gish, G. & Eckstein, F. (1988). DNA and RNA sequence determination based on phosphorothioate chemistry. *Science* **240**, pp. 1520-1522.
106. Christian, E.L. & Yarus, M. (1993). Metal coordination sites that contribute to structure and catalysis in the group I intron from Tetrahymena. *Biochemistry* **32**, pp. 4475-4480.
107. Harris, M.E. & Pace, N.R. (1995). Identification of phosphates involved in catalysis by the ribozyme RNase P RNA. *RNA* **1**, pp. 210-218.
108. Benner, S.A., Burgstaller, P., Battersby, T.R. & Jurczyk, S. (1999). Did the RNA World Exploit an Expanded Genetic Alphabet? In *The RNA World* (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds.), 2nd ed. pp. 163-181, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
109. Roth, A. & Breaker, R.R. (1998). An amino acid as a cofactor for a catalytic polynucleotide. *Proc Natl Acad Sci U S A* **95**, pp. 6027-6031.
110. Takagi, Y., Warashina, M., Stec, W.J., Yoshinari, K. & Taira, K. (2001). SURVEY AND SUMMARY: Recent advances in the elucidation of the mechanisms of action of ribozymes. *Nucleic Acids Res* **29**, pp. 1815-1834.
111. Weinstein, L.B., Jones, B.C., Cosstick, R. & Cech, T.R. (1997). A second catalytic metal ion in group I ribozyme. *Nature* **388**, pp. 805-808.
112. Yoshida, A., Sun, S. & Piccirilli, J.A. (1999). A new metal ion interaction in the Tetrahymena ribozyme reaction revealed by double sulfur substitution. *Nat. Struct. Biol.* **6**, pp. 318-321.

113. Shan, S., Yoshida, A., Sun, S., Piccirilli, J.A. & Herschlag, D. (1999). Three metal ions at the active site of the Tetrahymena group I ribozyme. *Proc. Natl. Acad. Sci. U.S.A.* **96**, pp. 12299-12304.
114. Shan, S., Kravchuk, A.V., Piccirilli, J.A. & Herschlag, D. (2001). Defining the catalytic metal ion interactions in the Tetrahymena ribozyme reaction. *Biochemistry* **40**, pp. 5161-5171.
115. Sontheimer, E.J., Gordon, P.M. & Piccirilli, J.A. (1999). Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev* **13**, pp. 1729-1741.
116. Smith, D. & Pace, N.R. (1993). Multiple magnesium ions in the ribonuclease P reaction mechanism. *Biochemistry* **32**, pp. 5273-5281.
117. McKay, D.B. & Wedekind, J.E. (1999). Small Ribozymes. In *The RNA World* (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds.), 2nd ed. pp. 265-286, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
118. O'Rear, J.L., Wang, S., Feig, A.L., Beigelman, L., Uhlenbeck, O.C. & Herschlag, D. (2001). Comparison of the hammerhead cleavage reactions stimulated by monovalent and divalent cations. *Rna* **7**, pp. 537-545.
119. Ferre-D'Amare, A.R., Zhou, K. & Doudna, J.A. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, pp. 567-574.
120. Perrotta, A.T., Shih, I. & Been, M.D. (1999). Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science* **286**, pp. 123-126.
121. Saenger, W. (1984). *Principles of Nucleic Acid Structure*. Springer-Verlag: New York.
122. Nakano, S., Proctor, D.J. & 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**, pp. 12022-12038.
123. Rupert, P.B. & Ferre-D'Amare, A.R. (2001). Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**, pp. 780-786.
124. Pinard, R., Hampel, K.J., Heckman, J.E., Lambert, D., Chan, P.A., Major, F. & Burke, J.M. (2001). Functional involvement of G8 in the hairpin ribozyme cleavage mechanism. *Embo J* **20**, pp. 6434-6442.
125. Ryder, S.P., Oyelere, A.K., Padilla, J.L., Klostermeier, D., Millar, D.P. & Strobel, S.A. (2001). Investigation of adenosine base ionization in the hairpin ribozyme by nucleotide analog interference mapping. *Rna* **7**, pp. 1454-1463.
126. Lebruska, L.L., Kuzmine, II & Fedor, M.J. (2002). Rescue of an abasic hairpin ribozyme by cationic nucleobases. Evidence for a novel mechanism of RNA catalysis. *Chem Biol* **9**, pp. 465-473.
127. Muth, G.W., Ortoleva-Donnelly, L. & Strobel, S.A. (2000). A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science* **289**, pp. 947-950.
128. Bayfield, M.A., Dahlberg, A.E., Schulmeister, U., Dorner, S. & Barta, A. (2001). A conformational change in the ribosomal peptidyl transferase center upon active/inactive transition. *Proc Natl Acad Sci U S A* **98**, pp. 10096-10101.

129. Muth, G.W., Chen, L., Kosek, A.B. & Strobel, S.A. (2001). pH-dependent conformational flexibility within the ribosomal peptidyl transferase center. *Rna* **7**, pp. 1403-1415.
130. Polacek, N., Gaynor, M., Yassin, A. & Mankin, A.S. (2001). Ribosomal peptidyl transferase can withstand mutations at the putative catalytic nucleotide. *Nature* **411**, pp. 498-501.
131. Thompson, J., Kim, D.F., O'Connor, M., Lieberman, K.R., Bayfield, M.A., Gregory, S.T., Green, R., Noller, H.F. & Dahlberg, A.E. (2001). Analysis of mutations at residues A2451 and G2447 of 23S rRNA in the peptidyltransferase active site of the 50S ribosomal subunit. *Proc Natl Acad Sci U S A* **98**, pp. 9002-9007.
132. Steitz, T.A. (1998). A mechanism for all polymerases. *Nature* **391**, pp. 231-232.
133. Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H. & Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science* **264**, pp. 1891-1903.
134. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature* **391**, pp. 251-258.
135. Kiefer, J.R., Mao, C., Braman, J.C. & Beese, L.S. (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* **391**, pp. 304-307.
136. Steitz, T.A. & Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* **90**, pp. 6498-6502.
137. Freemont, P.S., Friedman, J.M., Beese, L.S., Sanderson, M.R. & Steitz, T.A. (1988). Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc Natl Acad Sci U S A* **85**, pp. 8924-8928.
138. Beese, L.S. & Steitz, T.A. (1991). Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *Embo J* **10**, pp. 25-33.
139. Kim, E.E. & Wyckoff, H.W. (1991). Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *J Mol Biol* **218**, pp. 449-464.

# Chapter I

# Recognition of Nucleoside Triphosphates during RNA-Catalyzed Primer Extension.

The work presented in this chapter was a collaborative effort by myself, Catherine Yen, and Eric Ekland. Catherine synthesized the pppGGA trinucleotide and measured the reverse reaction, Eric assessed the ribozyme's activity with nucleotide anologues, and I performed the other experiments.

[1]Abbreviations: NTP, nucleoside triphosphate; $PP_i$, inorganic pyrophosphate; EDTA, Ethylenediaminetetraacetic acid; 3′-dATP, cordycepin triphosphate; MES, 2-[N-Morpholino]ethanesulfonic acid; BES, N,N-bis[2-Hydroxyethyl]-2-aminoethanesulfonic acid; EPPS, N-[2-Hydroxyethyl]piperazine-N′-[3-propanesulfonic acid].

[2] These $K_d$ values were derived by first fitting for the $K_i$ of the NTP with the mismatched substrate using the standard formula for noncompetitive inhibition (eq 3),

$$k_{obs} = \frac{[S]k_{cat}/(1 + [I]/K_i)}{[S] + K_m} \qquad (3)$$

Assuming that all of the deflection from linearity in the plots for the mismatches is due to inhibition, $K_i^{GTP}$ would be 8 mM (Figure 4B), and $K_i^{CTP}$ would be 21 mM (Figure 4C). These $K_i$ values were used to fit for $K_d$ and $k_{chem}$ (rate constant of the chemical step) with the matched substrates, again using eq 3 but substituting $k_{chem}$ for $k_{cat}$ and $K_d$ for $K_m$ (reasonable substitutions given a simple single-turnover reaction with rate-limiting $k_{chem}$). This generated estimates of $K_d$ and $k_{chem}$ for matched substrates in the case of maximal noncompetitive substrate inhibition. For the GTP ribozyme $K_d^{GTP}$ would be 30 mM, and $k_{chem}^{GTP}$ would be 16 $min^{-1}$; for the CTP ribozyme $K_d^{CTP}$ would be 50 mM, and $k_{chem}^{CTP}$ would be 13 $min^{-1}$. Eq 4

$$K_d^{mismatch} = \frac{K_d^{match}(k_{cat}/K_m^{match})}{k_{cat}/K_m^{mismatch}} \qquad (4)$$

was used to calculate the $K_d$ of the mismatched substrate in the event that all discrimination was at the level of NTP binding.

ABSTRACT: In support of the idea that certain RNA molecules might be able to catalyze RNA replication, a ribozyme was previously generated that synthesizes short segments of RNA in a reaction modeled after that of proteinaceous RNA polymerases. Here, we describe substrate recognition by this polymerase ribozyme. Altering base or sugar moieties of the nucleoside triphosphate only moderately affects its utilization, provided that the alterations do not disrupt Watson-Crick pairing to the template. Correctly paired nucleotides have both a lower $K_m$ and a higher $k_{cat}$, suggesting that differential binding and orientation each play roles in discriminating matched from mismatched nucleotides. Binding of the pyrophosphate leaving group appears weak, as evidenced by a very inefficient pyrophosphate exchange reaction, the reverse of the primer-extension reaction. Indeed, substitutions at the γ-phosphate can be tolerated, although poorly. Thio substitutions of oxygen atoms at the reactive phosphate exert effects similar to those seen with cellular polymerases, leaving open the possibility of an active site analogous to those of protein enzymes. The polymerase ribozyme, derived from an efficient RNA ligase ribozyme, can achieve the very fast $k_{cat}$ of the parent ribozyme when the substrate of the polymerase (GTP) is replaced by an extended substrate (pppGGA), in which the GA dinucleotide extension corresponds to the second and third nucleotides of the ligase. This suggests that the GA dinucleotide, which had been deleted when converting the ligase into a polymerase, plays an important role in orienting the 5′-terminal nucleoside. Polymerase constructs that restore this missing orientation function should achieve much more efficient and perhaps more accurate RNA polymerization.

The RNA world hypothesis proposes that early life relied on catalytic RNAs rather than protein enzymes for catalysis [1]. Much of the appeal of this theory comes from the notion that ribozymes, which could have served as their own genes, would have been much simpler to duplicate than proteins [2-5]. Hence, a vital component of the RNA world scenario is an RNA polymerase ribozyme responsible for replicating the essential ribozymes of early life (including itself). Whether there might be RNA sequences that can promote RNA polymerization with sufficient accuracy and efficiency for RNA self-replication is one of the fundamental unanswered questions related to the RNA world hypothesis [1, 6]. Finding such a sequence would support the idea of the RNA world and provide a key component for the synthesis of simple living cells [7].

The class I ligase (Figure 1A) is a promising starting point for efforts to develop a self-replicating ribozyme. This ribozyme ligates two RNAs, generating a 3′,5′-phosphodiester linkage and releasing pyrophosphate, a reaction similar to that of RNA and DNA polymerases [8]. Engineered forms of the ligase can polymerize short segments of RNA (Figure 1C). In the presence of appropriate RNA templates and nucleoside triphosphates (NTPs[1]), such ribozymes can extend an RNA primer by three or, in special cases, more nucleotides [9]. The overall Watson-Crick error rate is about 10%; i.e., on average, 88-92% of the extension is by the Watson-Crick match to the template, whereas an average of 3-4% of the extension is by each of the three mismatches. Although an error rate of 10% is too large to support RNA self-replication, it is significantly smaller than the 60% error rate calculated solely on the basis of the observed stabilities of Watson-Crick matches relative to mismatches at the ends of helices [6]. The ribozyme is particularly accurate at C and G template residues; at a C template residue, GTP is utilized 1000 times more efficiently than CTP, and at a G, CTP is utilized 600 times more efficiently than GTP [9].

Polymerase-like ribozymes derived from naturally occurring ribozymes, such as self-splicing introns, have not demonstrated comparable Watson-Crick fidelity [6, 10]. Another advantage of the ligase derivatives is that polymerization yields new RNA linkages, whereas the self-splicing derivatives catalyze disproportionation reactions, in which one RNA becomes longer at the expense of another. Furthermore, pyrophosphate activation provides a chemical driving force (2.1 kcal/mole per nucleotide) for the reaction [11]. This disfavors the reverse reaction and might eventually be harnessed to promote translocation and displacement of a non-coding strand or intramolecular template structure [6].

Here we describe the recognition and utilization of NTPs during RNA-catalyzed primer extension. We investigated the basis for the high Watson-Crick discrimination at G and C template residues and the degree to which the primer extension reaction is reversible. We also discovered the importance of a GA dinucleotide segment for substrate utilization, explaining why deleting this dinucleotide from the parent ligase to create the primer-extension construct resulted in a large decrease in activity. Our results reinforce the promise of the polymerase ribozyme as a starting point for developing a self-replicating ribozyme and identify specific elements of substrate binding and utilization to be targeted for improvement during future design and evolution efforts.

## MATERIALS AND METHODS

*Ribozyme and Substrates.* Ribozymes were prepared by run-off transcription of a PCR template as described [12]. The oligonucleotide substrate analogous to a primer was a synthetic DNA-RNA chimera (5′-aaaCCAGUC; lower case indicates DNA). This primer was 5′-radiolabeled using T4 polynucleotide kinase and $\gamma$-$^{32}$P-ATP. G(5′)ppp(5′)G, dGTP, ddGTP, and ultrapure grade ATP, CTP, and GTP were purchased from Pharmacia. GDP, guanosine diphosphomannose, and nicotinamide guanine dinucleotide were from Sigma; $\alpha$-GTP-phosphorothioates were from Dupont/NEN; and 7-deaza-dGTP was from Amersham. Polynucleotide substrates were generated using T7 RNA polymerase in abortive *in vitro* transcription reactions in which only GTP and ATP (for pppGGAA and pppGAAA) or GTP, CTP and UTP (for pppGGCU) were supplied. DNA oligonucleotide templates were GGTGTTCC*TATAGTGAGTCGTATTA*CGC, TCTCATTTC*TATAGTGAGTCGTATTA*CGC, and TGGTTCCAGTCGTATAGCC*TATAGTGAGTCGTATTA*CGC (italics denotes T7 polymerase promotor) for the pppGGAA, pppGAAA, and pppGGCU transcriptions, respectively. To body-label transcription products, $\alpha$-$^{32}$P-GTP was included. The pppGGA from abortive *in vitro* transcription was eluted from a gel and de-salted on a Sep-Pak Classic C18 cartridge (Millipore). Analysis by MALDI-TOF mass spectrometry (PerSeptive Biosystems) was consistent with the major purified product being pppGGA, with a detectable amount of contaminating ppGGA.

*Primer extension assays.* All reactions were performed under single-turnover conditions, with ribozyme in sufficient excess to saturate the labeled substrate. Typical reactions included 1 $\mu$M ribozyme and 0.5 $\mu$M primer. Prior to initiating the reaction, the ribozyme was heated in water (2-3 min at 80 °C) and cooled (2 min at 22 °C). Reactions were initiated with the simultaneous addition of salt, buffer, primer, and the triphosphate-containing substrate. Unless stated otherwise, reactions were in 30 or 50 mM Tris, pH 8.0,

60 mM $MgCl_2$, 200 mM KCl and 0.6 mM EDTA at 22° (or room temperature for incubations less than one minute). Because NTPs chelate $Mg^{+2}$ and the reaction rate is known to be sensitive to $Mg^{+2}$ concentration, the reaction mixture was supplemented with $MgCl_2$ (at a concentration stoichiometric with that of the NTP) to maintain a free $Mg^{+2}$ concentration of 60 mM. Reactions were stopped by addition of at least one volume of 8 M urea/120 mM EDTA and run on denaturing sequencing gels. Radiolabeled reactants and products were quantified with a Fujix BAS 2000 phosphorimager. Nucleotide-analog experiments were performed with a ribozyme that differed from that used in the rest of the experiments: A18-A20 was changed to CGU, A67 to G, and U78 to C. This minor alteration of the ribozyme sequence is not expected to change the relative efficiencies of primer extension.

Rates were determined using eq 1,

$$F = 1 - e^{-kt} \qquad (1)$$

where $t$ equals time, $k$ is the observed rate of catalysis, and F is the fraction of radiolabeled substrate converted to product. Analysis was complicated by a brief lag (<5 seconds) in the course of the reaction. This lag was not observed with the parent ligase [13]. A similar phenomenon was encountered with the Group II intron and was attributed to a $Mg^{2+}$-induced conformation change [14]. Rates were determined using the part of the curve that best fit eq 1.

Apparent second-order rate constants for polynucleotide substrates were determined by first calculating the fold preference for the polynucleotide substrate compared to GTP using eq 2,

$$\text{Preference} = \frac{P\text{–}X}{P\text{–}G} \times \frac{GTP_o}{pppX_o} \qquad (2)$$

where P–X is the fraction of primer extended using the polynucleotide substrate, P–G is the fraction of primer extended using GTP, $GTP_0$ is the initial GTP concentration, and $pppX_0$ is the initial polynucleotide substrate concentration. The preference value was then multiplied by the apparent second-order rate constant for GTP addition (590 $M^{-1}$ $min^{-1}$) to give the apparent second-order rate constant for the polynucleotide substrate.

The substrate for the reverse reaction, aaaCCAGUCGGdA (lowercase, DNA; dA, 3′-deoxyadenosine), represented the product of a polynucleotide extension reaction in which the standard primer had been extended using pppGGdA. This oligonucleotide was generated by extending aaaCCAGUCGG (Dharmacon Research) with a 3′-deoxyadenosine nucleotide, using cordycepin triphosphate (3′-dATP) and yeast poly(A) polymerase (Amersham). To radiolabel the oligonucleotide, $\alpha$-$^{32}$P-3′-dATP was used.

## RESULTS AND DISCUSSION

This study focuses on a ribozyme referred to as the "GTP ribozyme," which extends a primer by a single nucleotide using GTP (Figure 1D). The GTP ribozyme differs from the polymerase construct (Figure 1C) by the covalent attachment of the template. Constructs with this intramolecular configuration are more convenient to produce and characterize, as well as about ten times more efficient than those with the intermolecular configuration. While the ribozyme core can accommodate templates coding for multiple nucleotides, utilization of templates coding for a single nucleotide is most efficient. Thus, our prototype ribozyme codes for primer extension by a single nucleotide. While the GTP ribozyme was chosen for maximal convenience and efficiency, there is every indication that with other templates and configurations the ribozyme core will display the same properties and preferences, albeit with slower overall rates [9]. Using the GTP ribozyme, we investigated the efficiency of RNA-catalyzed primer extension with respect to the three components of the NTP substrate: the reactive phosphate, the nucleoside, and the pyrophosphate leaving group.

*The Reactive Phosphate.* The primer extension reaction centers on the α-phosphate of the NTP. Substitution of either of the non-bridging oxygens of this phosphate decreased the efficiency of primer extension (Figure 2). Substitution of the pro-$R_P$ oxygen decreased the apparent second-order rate constant ($k_{cat}/K_m^{GTP}$) of primer extension by more than 1000 fold, while substitution of the pro-$S_P$ oxygen had a more modest, 37-fold effect. Some decrease in reactivity would be expected from these substitutions due to their destabilization of the associative transition state [15]. For instance, a 60-fold decrease in rate for DNA polymerase upon substitution of the pro-$S_P$ oxygen is attributed to this elemental effect on the transition state [16]. This effect on intrinsic reactivity could also fully account for the 37-fold effect seen with the pro-$S_P$ substitution on the activity of the GTP ribozyme, provided that the chemical step is rate limiting during primer extension.

To investigate whether the chemical step was rate limiting, the pH dependence of primer extension was examined. The logarithm of the rate as a function of pH is a line with a slope of 0.9 (Figure 3), indicating that a single deprotonation is rate limiting. Such behavior for ribozyme reactions involving phosphoryl transfer is taken as evidence of a rate-limiting chemical step, because the attacking hydroxyl must be deprotonated during this step [15, 17-19]. Thus, the thio effect seen upon substitution of the pro-$S_P$ oxygen can be attributed solely to the elemental effect on the reactivity of the phosphate reaction center.

The effect of substituting the pro-$R_P$ oxygen is much more dramatic (1000 fold), suggesting an important ribozyme contact or metal coordination to the pro-$R_P$ oxygen. Intriguingly, the steriospecificity of the thio effects matches that of protein-enzyme polymerases [20]. Both the ribozyme and protein enzymes suffer mere elemental effects upon substitution of the pro-$S_P$ oxygen and much larger effects with the pro-$R_P$ substitution. This leaves open the possibility that both the ribozyme and the protein active sites use the same strategy for promoting the chemistry of primer extension, each employing similar two-metal-ion mechanisms for phosporyl transfer [21-24]. Furthermore, the similar elemental effects upon pro-$S_P$ substitution (37 and 60 fold) suggests that the two transition states lie at similar positions along the continuum between associative and dissociative character [15].

*The Nucleoside: Michaelis-Menten Parameters for Matched and Mismatched Substrates.* Protein-enzyme polymerases discriminate between matched and mismatched NTPs at the level of NTP binding as well as preferential utilization of NTPs bound with Watson-Crick geometry, although the relative importance of these two modes of discrimination varies [16, 25, 26]. To investigate the means of discrimination during RNA-catalyzed polymerization, single-turnover kinetic parameters were determined for matched and mismatched substrates. With increasing concentration of GTP, the primer-extension

rate of the GTP ribozyme displayed saturable behavior that fit well to a Michaelis-Menten curve (Figure 4A, apparent $k_{cat}^{GTP} = 2.4\ min^{-1}$, apparent $K_m^{GTP} = 4.1$ mM). However, the concentration course had to be abbreviated because GTP formed a visible precipitate at concentrations greater than 10 mM. This led to concerns that these parameters might merely reflect aggregation of GTP or noncompetitive inhibition. To control for these nonspecific effects, the concentration course was repeated using the "CTP ribozyme," a ribozyme variant that differed at the template residue such that it coded for extension using CTP (Figure 1E). Although the rates for the CTP ribozyme were nearly a 1000 times slower, they also began to display saturable behavior at high GTP concentrations, fitting well to a Michaelis-Menten curve (Figure 4B, apparent $k_{cat}^{GTP} = 4 \times 10^{-3}\ min^{-1}$ and apparent $K_m^{GTP} = 8.0$ mM). Thus, some degree of GTP aggregation or noncompetitive inhibition cannot be ruled out. Nevertheless, the twofold lower apparent $K_m^{GTP}$ of the GTP ribozyme implies that GTP is at least beginning to saturate the GTP-binding site of this ribozyme.

We performed the analogous set of experiments using CTP, which is more soluble than GTP (Figure 4C and D; for the CTP ribozyme, apparent $k_{cat}^{CTP} = 2.5\ min^{-1}$, apparent $K_m^{CTP} = 7.4$ mM; for the GTP ribozyme, apparent $k_{cat}^{CTP} = 3 \times 10^{-2}\ min^{-1}$, apparent $K_m^{CTP} = 20$ mM). Again, saturable behavior was seen for both ribozymes, with the $K_m$ of the cognate ribozyme lower (threefold) than that of the other ribozyme. This supports the idea that the saturable behavior seen for the cognate substrate of each ribozyme (Figure 4A and D) reflects occupation of the NTP-binding site. The simplest explanation for the Michaelis-Menten behavior of Figure 4B and C is that the NTPs also begin to occupy the NTP-binding sites of the non-cognate ribozymes, though with slightly less affinity than for the cognate ribozymes. The dramatic $k_{cat}$ differences between cognate and non-cognate ribozymes can be explained by much more favorable NTP orientation with the cognate ribozyme.

Although we do not rule out the possibility that NTP aggregation or non-competitive inhibition might generate some of the non-linear behavior seen in Figure 4, we can at least qualitatively surmise that the polymerase uses both binding and orientation to discriminate between matched and mismatched NTPs. The contribution of binding would be least (only 2-3 fold) and that of orientation would be greatest (>100 fold) in the case where $K_m = K_d$ at the NTP-binding site and the mismatched nucleotides are truly saturating the active site. If, for one or both of the ribozymes with mismatched NTPs, the deviation from linearity were instead due to aggregation or non-competitive inhibition, then the contribution of binding would be correspondingly higher. Even in such a scenario, it is doubtful that binding differences could account for all the discrimination, as this would imply $K_d$ (dissociation constant) values for the mismatches of 12 M and 30 M for the GTP and CTP ribozymes, respectively,[2] which approach the concentration of water.

*The Nucleoside: GTP analogs.* Features of the NTP beyond its Watson-Crick face could be important for productive binding by the ribozyme. The contributions of different moieties of the NTP were probed by measuring extension rates with different GTP analogs (Figure 2). Altering base and sugar moieties had a minor effect, provided that Watson-Crick pairing was maintained. Deleting the 2′-oxygen (2′-dGTP) decreased activity threefold, yet the dideoxy analogue (2′,3′-ddGTP) was favored nearly threefold over GTP. Another possible contact to the nucleoside was investigated with 7-deaza-dGTP, which replaces a potential hydrogen bond donor, N7, with carbon. Taking the absence of the 2′-hydroxyl into account, the 7-deaza substitution reduced the rate fourfold.

The slight discrimination between ribo- and 2′-deoxyribonucleotides contrasts with the strict discrimination made by protein enzymes [27]. Biological polymerases are under intense selective pressure to differentiate between ribo- and 2′-deoxyribonucleotides. Both sets of NTPs are present in the cell, and incorporation of ribonucleotides into DNA or

deoxyribonucleotides into RNA is detrimental. Lack of discrimination might have posed a less serious problem in the RNA world before deoxyribonucleotides became common. In contrast to their discrimination at the 2′ position, some bacteriophage RNA and DNA polymerases have only a threefold preference for a 3′-hydroxyl [28, 29]; in their native context, these enzymes are rarely challenged with 3′-deoxynucleoside triphosphates. For the GTP ribozyme, the lack of recognition of the 3′-hydroxyl was expected because the parental ligase ribozyme has a phosphodiester linkage at this position.

*The Nucleoside: 3′-terminal extensions.* Substrate recognition was further examined using different 3′-terminal extensions of the nucleoside. The key modification in the conversion of the ligase (Figure 1B) to a polymerase (Figure 1C) was the replacement of the first four nucleotides of the ligase (pppGGAA) with exogenous GTP (compare Figure 1A with 1D). This engineering exacted a large cost in reaction rate. For example, in conditions where the apparent $k_{cat}^{GTP}$ was 2.4 $min^{-1}$ for the GTP ribozyme (Figure 4A), the $k_{cat}$ for multiple-turnover ligation is 140 $min^{-1}$ and the rate of self-ligation of the parent ligase is 300 $min^{-1}$ [13](M.E.G, N. Bergman, and D.P.B manuscript in preparation). To determine if the activity of the GTP ribozyme could be restored to that of the parent ligase, we examined its activity with polynucleotide triphosphate substrates.

When primer extension using GTP was challenged with a combination of polynucleotide triphosphates, including pppGG, pppGGA, pppGGG, and pppGGAA, the GTP ribozyme preferentially utilized the polynucleotides (Figure 5A). Examining appropriate time points and normalizing to the efficiency of GTP utilization allowed the determination of the $k_{cat}/K_m$ for these and other polynucleotide substrates (Table 1). With polynucleotide substrates corresponding to the first three nucleotides of the parental self-ligating ribozyme the efficiency dramatically increased. The importance of G at the second nucleotide of the substrate can be seen in the increased extension efficiency of pppGG

compared to that of GTP (40 fold) as well as the increased efficiency of pppGGAA relative to that of pppGAAA (80 fold). The importance of A at the third nucleotide can be seen in the increased extension efficiency of pppGGA compared to that of pppGG (40 fold) as well as the increased efficiency of pppGGAA verses that of pppGGCU (14 fold).

The most favored polynucleotide substrate tested, pppGGA, was used 1300-fold more efficiently than GTP. We determined the Michaelis-Menten parameters of pppGGA to investigate whether this high catalytic efficiency was due primarily to tighter binding or faster catalysis. For these reactions the pH was lowered from pH 8.0 to 6.0 in order to slow the reaction sufficiently for time points to be acquired by manual pipetting. The concentration course of pppGGA revealed an apparent $k_{cat}$ of 1.9 $min^{-1}$ and an apparent $K_m$ of 0.6 mM. Extrapolating to pH 8.0 gives a $k_{cat}$ of 190 $min^{-1}$ and a $k_{cat}/K_m$ of $3 \times 10^5$ $M^{-1}min^{-1}$, which is within range of the $k_{cat}/K_m$ value determined earlier by less direct methods ($8 \times 10^5$ $M^{-1}min^{-1}$, Table 1).

A $k_{cat}$ of 190 $min^{-1}$ for pppGGA at pH 8.0 approaches that seen for the rate of self-ligation by the parental ligase, which is 300 $min^{-1}$ [13](M.E.G, N. Bergman, D.P.B manuscript in preparation). Thus, extending the substrate with a GA dinucleotide, which corresponds to the second and third nucleotides deleted when converting the ligase to a polymerase, nearly restores the ligation rate to that of the parent—even though there is no covalent attachment linking pppGGA to the GTP ribozyme. The first G of pppGGA is thought to form a Watson-Crick pair with C8 of the GTP ribozyme. However, there is no obvious Watson-Crick partner for either nucleotide of the dinucleotide extension. To the degree that the apparent $K_m$s of GTP and pppGGA (4.1 mM and 0.6 mM, respectively) reflect substrate binding, the GA extension increases the affinity of the substrate. However, a more important role for the extension appears to be the increased utilization of bound substrate, as reflected in the 80-fold difference in apparent $k_{cat}$ (2.4 $min^{-1}$ and 190 $min^{-1}$, respectively).

We suspect that the GA extension interacts with nucleotides in the core of the ribozyme by means of non-Watson-Crick contacts and positions the pppG portion of the substrate for attack by the 3′-hydroxyl group of the primer. In the absence of the GA extension, bound GTP achieves productive orientation only a small fraction of the time. One target for improving the efficiency (and perhaps fidelity) of primer extension would be to generate ribozyme derivatives that restore this putative orientation function.

*The Leaving Group.* For natural RNA and DNA polymerases much of the NTP affinity is derived from contacts to the triphosphate [22, 30, 31]. When T7 DNA polymerase binds a matched NTP in the closed complex, it contacts all seven nonbridging oxygen atoms of the triphosphate, either directly through interactions with amino acid residues or indirectly via metal ions [22]. In the case of the GTP riboyzme, substitutions of the pyrophosphate leaving group significantly reduced the efficiency of primer extension (Figure 2). Phosphate and substituted phosphates (mannose phosphate and nicotinamide phosphate) were poor leaving groups, reducing the rate two or three orders of magnitude, respectively. GDP, the leaving group for the cap analog G(5′)ppp(5′)G, had a smaller effect, reducing the rate 20 fold. The large rate reduction by leaving-group substitutions may implicate important phosphate contacts, but it may also reflect the altered chemical nature of the leaving group. Phosphate is expected to be a poor leaving group, since its last $pK_a$ is 12.0, compared to 9.1 for pyrophosphate [32, 33]. Substituted phosphates may have additional $pK_a$ perturbations or steric effects.

The affinity of proteinaceous polymerases for the pyrophosphate leaving group is so great that excess pyrophosphate can drive the reaction backward [34, 35]. This pyrophosphate exchange reaction provides an avenue for increased RNA polymerization fidelity; incorporated mismatched nucleotides that block translocation can be excised and replaced [35]. For the GTP ribozyme, attempts to detect the reverse of single-nucleotide

extension have not been successful (N. Bergman and D.P.B., unpublished result). Nevertheless, the reverse of extension using pppGGA can be detected (Figure 6), presumably because of the more productive orientation imparted by the 3′-terminal GA dinucleotide. The GTP ribozyme was incubated with pyrophosphate, an excess of unlabeled pppGGA, and a synthetic oligonucleotide, aaaCCAGUCGG*dA, representing the product of extension using pppGGdA (lowercase, DNA; *, $^{32}$P-radiolabeled phosphate; dA, 3′-deoxyadenosine). The unlabeled pppGGA served as a chase, preventing the newly formed pyrophosphate-exchange product, pppGG*dA, from re-associating with the ribozyme and re-ligating to the primer. The formation of pppGG*dA depended on the pyrophosphate concentration, with an apparent second-order rate constant, $k_{cat}/K_m^{PP_i}$, of 1.3 $\times 10^{-2}$ $M^{-1}$ $min^{-1}$ (Figure 6). The reaction did not begin to display saturable behavior with soluble concentrations of pyrophosphate ($\leq$ 0.3 mM). Thus, the GTP ribozyme has a detectable, though non-quantifiable, affinity for pyrophosphate.

Considering the relative efficiencies of the forward and reverse reactions, the forward reaction would be preferred by a factor of $10^7$ in the presence of moderate concentrations of pppGGA and pyrophosphate (e.g. 0.3 mM each). The preference cannot be determined for single-nucleotide extension, because the reverse reaction is too inefficient to be detected, but the preference is expected to be of a similar magnitude. Furthermore, as concentrations are raised, pyrophosphate will begin precipitating before NTPs, further favoring the forward reaction. Thus, unless improved forms of the polymerase have greatly increased affinity for pyrophosphate, their reactions will be essentially irreversible, obviating any need for a pyrophosphatase activity to drive polymerization to completion. On the other hand, a meager reverse reaction prevents utilization of pyrophosphate exchange for increased fidelity [35].

A second activity was detected during our investigation of the reverse reaction. A product migrating slower on the sequencing gel than pppGG*dA was not pyrophosphate-dependent (Figure 6). This product appears to be pGG*dA, generated by ribozyme-catalyzed hydrolysis at the reactive phosphate, occurring at a very slow rate ($4.5 \times 10^{-6}$ $min^{-1}$). This reaction, in which an RNA linkage is hydrolyzed to generate a 3′-hydroxyl and a 5′-phosphate, is analogous to the reactions of RNAse P and group I-like ribozymes, as well as certain side reactions of group I and group II self-splicing introns [36-39]. Observation of this reaction with the GTP ribozyme demonstrates that, as expected, not all of its transition-state stabilization is mediated through contacts to the pyrophosphate leaving group; at least part occurs at the reactive phosphate or 3′-oxygen.

*Conclusion.* The GTP ribozyme and related constructs have some intriguing properties, particularly when considering that they were derived from a ribozyme that was selected from random sequences based on its ability to perform a simple, sequence-specific ligation reaction—not an ability to perform template-dependent primer extension using NTPs. The steriospecificity of the thio effects at the reactive phosphate matches those of protein enzymes, leaving open the possibility of a common mechanism. The discrimination among NTPs depends predominantly on Watson-Crick interaction with the template. Furthermore, the reaction is essentially irreversible. It will be interesting to examine the extent to which combinatorial methods can generate ribozymes capable of more accurate, general, and extensive RNA polymerization. A promising approach would be to generate a large pool of RNA sequences based on the core of the GTP ribozyme and to select for variants based on their ability to perform primer extension reactions rather than RNA ligation.

**ACKNOWLEDGMENT**

We thank Mike Lawrence and Nick Bergman for helpful comments on this manuscript.

## REFERENCES


1. Joyce, G. F. & Orgel, L. E. Prospects for understanding the origin of the RNA world. in *The RNA World* (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 49-77 (Cold Spring Harbor laboratory press, New York, 1999).
2. Pace, N. R. & Marsh, T. L. *Origins Life* **16**, 97-116 (1985).
3. Sharp, P. A. *Cell* **42**, 397-400 (1985).
4. Cech, T. R. *Proc. Natl. Acad. Sci. USA* **83**, 4360-4363 (1986).
5. Orgel, L. E. *J. Theor. Biol.* **123**, 127-149 (1986).
6. Bartel, D. P. Re-creating an RNA replicase. in *The RNA World* (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 143-162 (Cold Spring Harbor laboratory press, New York, 1999).
7. Bartel, D. P. & Unrau, P. J. *Trends Biochem. Sci.* **24**, M9-M13 (1999).
8. Ekland, E. H., Szostak, J. W. & Bartel, D. P. *Science* **269**, 364-370 (1995).
9. Ekland, E. H. & Bartel, D. P. *Nature* **382**, 373-376 (1996).
10. Bartel, D. P., Doudna, J. A., Usman, N. & Szostak, J. W. *Mol. Cell. Biol.* **11**, 3390-3394 (1991).
11. Erie, D. A., Yager, T. D. & von Hippel, P. H. *Annu. Rev. Biophys. Biomol. Struct.* **21**, 379-415 (1992).
12. Ekland, E. H. & Bartel, D. P. *Nucleic Acids Res.* **23**, 3231-3238 (1995).
13. Bergman, N. H., Johnston, W. K. & Bartel, D. P. *Biochemistry* **39**, 3115-3123 (2000).
14. Costa, M. & Michel, F. *EMBO J.* **14**, 1276-1285 (1995).
15. Herschlag, D., Piccirilli, J. A. & Cech, T. R. *Biochemistry* **30**, 4844-4854 (1991).
16. Wong, I., Patel, S. S. & Johnson, K. A. *Biochemistry* **30**, 526-537 (1991).
17. Beebe, J. A. & Fierke, C. A. *Biochemistry* **33**, 10294-10304 (1994).
18. Mei, R. & Herschlag, D. *Biochemistry* **35**, 5796-5809 (1996).

19. Dahm, S. C., Derrick, W. B. & Uhlenbeck, O. C. *Biochemistry* **32**, 13040-13045 (1993).

20. Eckstein, F. *Ann. Rev. Biochem.* **54**, 367-402 (1985).

21. Steitz, T. A. *Nature* **391**, p231-2 (1998).

22. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. *Nature* **391**, 251-258 (1998).

23. Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. *Nature* **391**, p304-7 (1998).

24. Brautigam, C. A. & Steitz, T. A. *J Mol Biol* **277**, p363-77 (1998).

25. Kuchta, R. D., Benkovic, P. A. & Benkovic, S. J. *Biochemistry* **27**, 6716-6725 (1988).

26. Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. *J. Biol. Chem.* **267**, 25988-25997 (1992).

27. Joyce, C. M. *Proc. Natl. Acad. Sci. USA* **94**, 1619-1622 (1997).

28. Axelrod, V. D. & Kramer, F. R. *Biochemistry* **24**, 5716-5723 (1985).

29. Tabor, S. & Richardson, C. C. *Proc. Natl. Acad. Sci. USA* **92**, 6339-6343 (1995).

30. Beese, L. S., Friedman, J. M. & Steitz, T. A. *Biochemistry* **32**, 14095-14101 (1993).

31. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. *Science* **264**, 1891-1903 (1994).

32. Rohatgi, R., Bartel, D. P. & Szostak, J. W. *J. Am. Chem. Soc.* **118**, 3332-3339 (1996).

33. Dawson, R. M. C., Elliott, D. C., Elliott, W. H. & Jones, K. M. in *Data for Biochemical Research* 400-412 (Clarendon Press, Oxford, 1986).

34. Deutscher, M. P. & Kornberg, A. *J. Biol. Chem.* **244**, 3019-3028 (1969).

35. Kahn, J. D. & Hearst, J. E. *J. Mol. Biol.* **205**, 291-314 (1989).

36. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. *Cell* **35**, 849-857 (1983).

37. Decatur, W. A., Einvik, C., Johansen, S. & Vogt, V. M. *EMBO J.* **14**, 4558-4568 (1995).

38. Inoue, T., Sullivan, F. X. & Cech, T. R. *J Mol Biol* **189**, 143-165 (1986).

39. Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L. & Perlman, P. S. *J. Biol. Chem.* **263**, 3432-3439 (1988).

40. Moroney, S. E. & Piccirilli, J. A. *Biochemistry* **30**, 10343-10349 (1991).

Table 1: Efficiency of Extension Using Polynucleotide Triphosphate Substrates

| Substrate | $k_{cat}/K_m(M^{-1}min^{-1})$ |
|---|---|
| pppG | 590 |
| pppGG | 21,000 |
| pppGGA | 800,000 |
| pppGGAA | 25,000 |
| pppGGCU | 1800 |
| pppGAAA | 310 |

## FIGURES

**Figure 1.** Design of RNA polymerase-like ribozymes. In the schematic representations (panels A-E) the core of the ribozyme is indicated by an oval, and Watson-Crick pairing involving the substrates is indicated by vertical dashes (with a proposed A:A interaction indicated by a double dash). (A) The parent class I ligase (solid line, GenBank #U26413) hybridized to the primer (open line) [8, 12]. The 5′-terminal residues (pppGGAA) are specified, as is the cytidine that pairs with the first nucleotide of the ribozyme. (B) The trans ligase, a multiple turnover ribozyme [8, 13]. In this ribozyme, the former 5′-leader of the ligase (heavy line) is detached from the ribozyme. (C) The RNA polymerase [9]. This ribozyme requires three components: a template (heavy line), primer (open line) and NTPs. (D) The GTP ribozyme (GenBank #AF303466). This ribozyme is similar to the polymerase ribozyme except that it is covalently attached to a template specifying primer extension using a single GTP. (E) The CTP ribozyme. To specify extension using CTP, the template base (residue 8) was changed from C to G. (F) Sequence and secondary structure of the GTP ribozyme.

**Figure 2.** Primer extension using GTP analogs. Primer extension was assayed using 1 mM of the indicated analogue (except GpppG, which was used at 0.5 mM to compensate for its twofold symmetry). Observed rates are reported relative to that using 1 mM GTP.

**Figure 3.** pH dependence of the GTP ribozyme. The reaction was carried out with 9 mM GTP in standard conditions, except Tris buffer was replaced by 50 mM MES (pH 5.7-6.5), BES (pH 6.7-7.4) or EPPS (pH 8.0). The slope of the line is 0.9.

**Figure 4.** Single-nucleotide extension as a function of nucleoside-triphosphate concentration. (A) GTP with the GTP ribozyme. The line corresponds to $k_{cat}^{GTP} = 2.4\ min^{-1}$, $K_m^{GTP} = 4.1$ mM, and $k_{cat}/K_m^{GTP} = 590\ M^{-1}\ min^{-1}$. (B) GTP with the CTP ribozyme. The line corresponds to $k_{cat}^{GTP} = 4 \times 10^{-3}\ min^{-1}$, $K_m^{GTP} = 8$ mM, and $k_{cat}/K_m^{GTP} = 0.5\ M^{-1}\ min^{-1}$. (C) CTP with the GTP

ribozyme. The line corresponds to $k_{cat}^{CTP} = 3 \times 10^{-2}$ min$^{-1}$, $K_m^{CTP} = 20$ mM, and $k_{cat}/K_m^{CTP} = 1.5$ M$^{-1}$ min$^{-1}$. (D) CTP with the CTP ribozyme. The line corresponds to $k_{cat}^{CTP} = 2.5$ min$^{-1}$, $K_m^{CTP} =$ 7.4 mM, and $k_{cat}/K_m^{CTP} = 340$ M$^{-1}$ min$^{-1}$. Reactions were performed with either 50 mM (▲) or 100 mM Tris (●).

**Figure 5.** Primer extension using polynucleotide triphosphates. (A) A mixture of GTP and polynucleotide triphosphates with combined concentration < 5 $\mu$M was incubated with 5 $\mu$M unlabeled primer and 5 $\mu$M GTP ribozyme. Shown is a phosphorimage of a 22.5% sequencing gel used to separate the radiolabeled substrates (GTP, pppGG, pppGGA, pppGGG, and pppGGAA) and extension products (P-G, P-GG, P-GGA, P-GGG, and P-GGAA, where P denotes the primer). The abortive transcription product indicated by an asterisk (*) was surmised to be pppGGG based on its gel mobility and the known propensity of T7 polymerase to add an untemplated G to the 5′-terminus of transcripts [40]. (B) Primer extension as a function of pppGGA concentration at pH 6.0. The line corresponds to $k_{cat}^{pppGGA} = 1.9$ min$^{-1}$, $K_m^{pppGGA} = 0.6$ mM, and $k_{cat}/K_m^{pppGGA} = 3 \times 10^3$ M$^{-1}$ min$^{-1}$. Substrate precipitation prevented extension of the concentration course beyond 0.75 mM pppGGA.

**Figure 6.** Pyrophosphate-exchange and hydrolysis reactions promoted by the GTP ribozyme as a function of time and pyrophosphate concentration. Reactions were performed with 1 $\mu$M ribozyme, <0.7 $\mu$M radiolabeled oligonucleotide (aaaCCAGUCGG*dA; * denotes $^{32}$P), 100 $\mu$M pppGGA and the indicated concentration of sodium pyrophosphate. The radiolabeled pyrophosphate-exchange product, pppGG*dA, was identified by co-migration with an authentic standard generated by T7 transcription using GTP and $\alpha$-$^{32}$P-3′-dATP (data not shown). The radiolabeled hydrolysis product, pGG*dA, was identified by co-migration with a standard generated by partial digestion of aaaCCAGUCGG*dA with S1 nuclease (data not shown). The slower migration of pGG*dA relative to pppGG*dA can be explained by their respective charges of –4 and –6.

Figure1

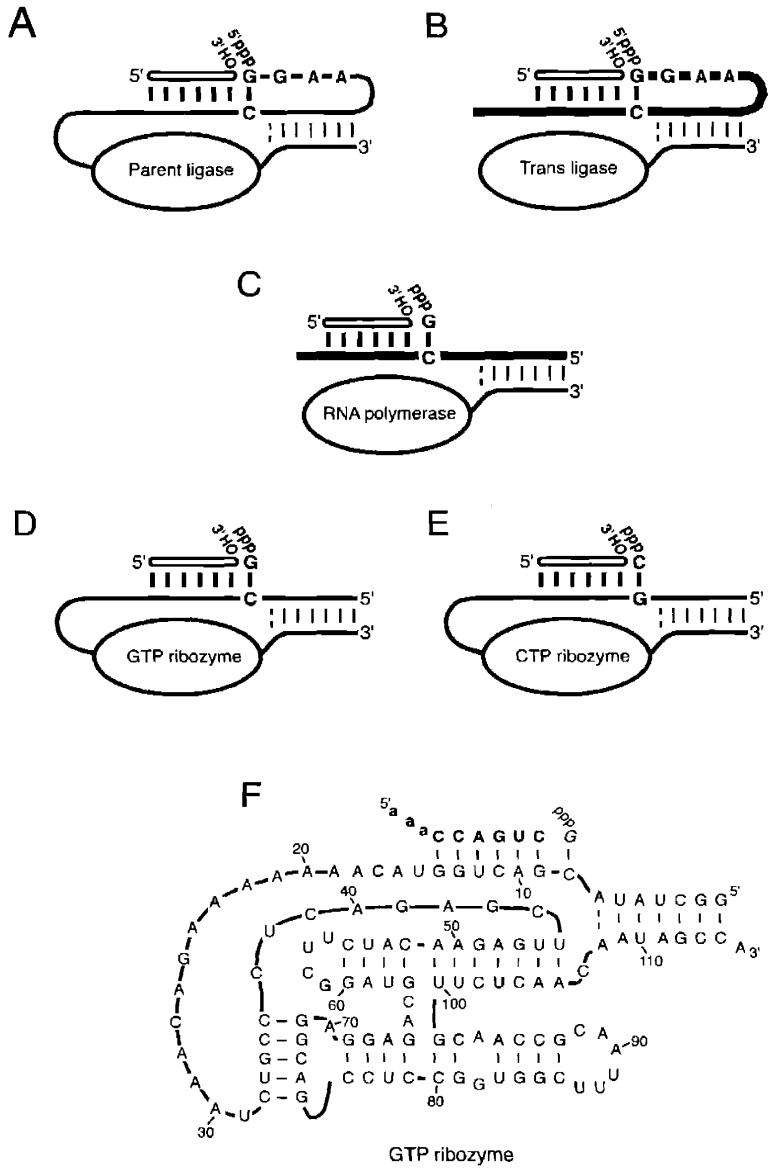

A
5'ppp
3'HO
5'
G-G-A-A
C
3'
Parent ligase
B
5'ppp
3'HO
5'
G-G-A-A
C
3'
Trans ligase
C
ppp
3'HO
5'
G
C
5'
3'
RNA polymerase
D
ppp
3'HO
5'
G
C
5'
3'
GTP ribozyme
E
ppp
3'HO
5'
C
G
5'
3'
CTP ribozyme
F
5'
a a
C C A G U C
ppp
G
20
A-A-A-A A C A U G G U C A G-C
40
C-A-G-A-G-C
10
A U A U C G G 5'
50
U
U U C U A C-A A G A G U U
A A U A G C C A 3'
110
C
C G G A U G U U C U C A A C
60
C
100
C-G A 70
A
C-G G G A G G C A A C C G C A A 90
G-C
U-A C C U C-C G G U G G C U U U
C-G
80
A U
30
GTP ribozyme

Figure 2

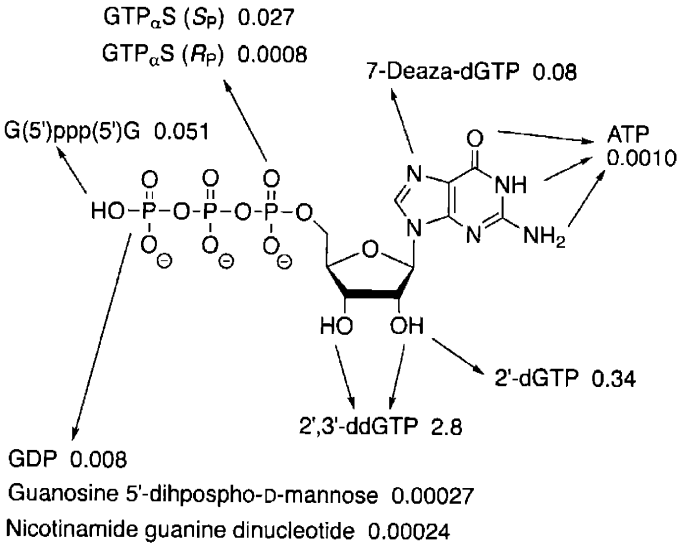

GTPαS (SP) 0.027
GTPαS (RP) 0.0008
7-Deaza-dGTP 0.08
G(5')ppp(5')G 0.051
ATP
0.0010
2'-dGTP 0.34
2',3'-ddGTP 2.8
GDP 0.008
Guanosine 5'-dihpospho-D-mannose 0.00027
Nicotinamide guanine dinucleotide 0.00024

Figure 3

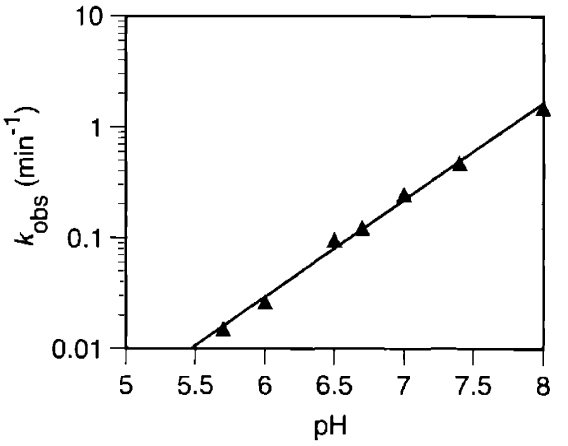


Figure 4

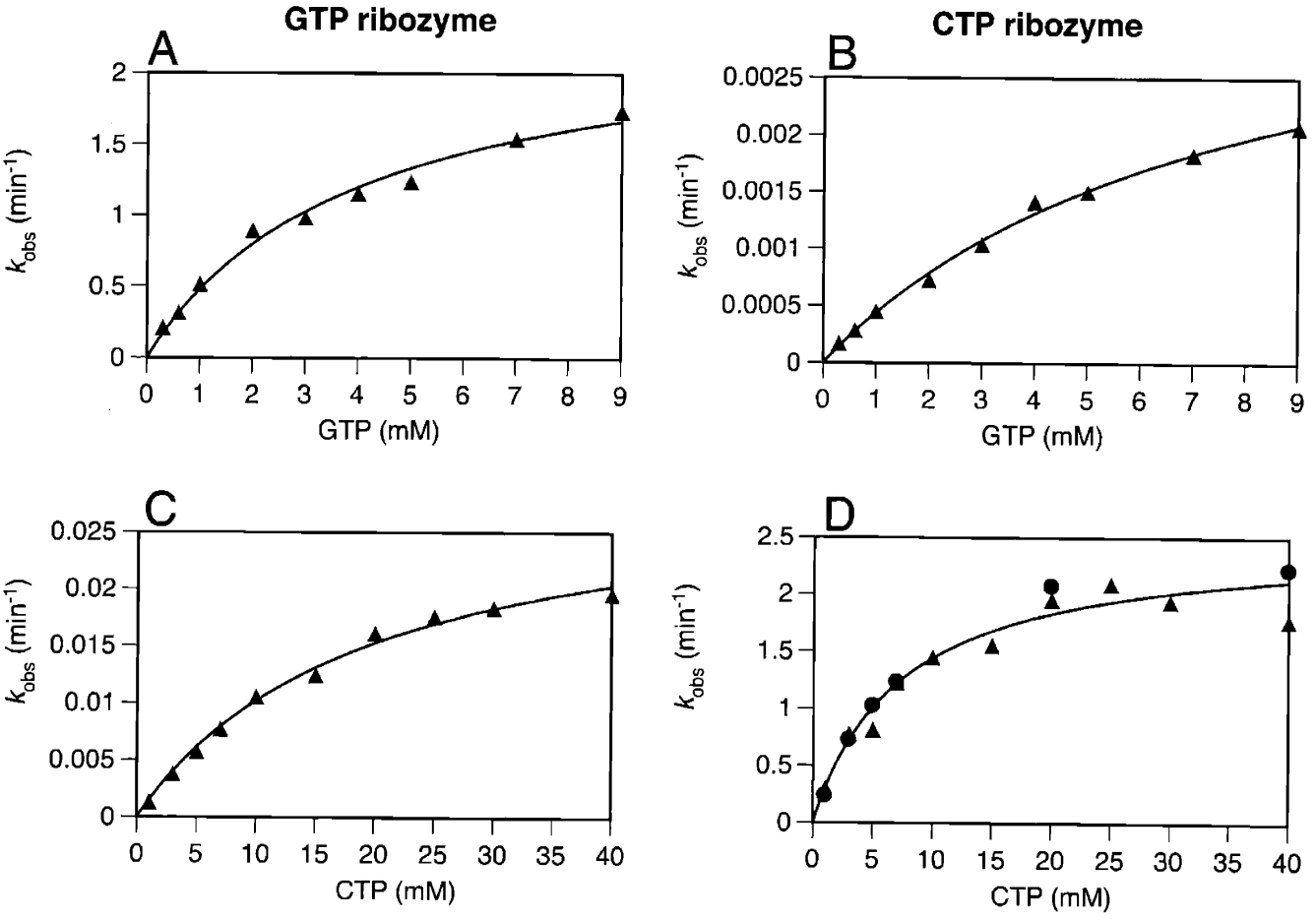

GTP ribozyme
CTP ribozyme
A
B
C
D
$k_{obs}$ (min$^{-1}$)
GTP (mM)
CTP (mM)
0
0.5
1
1.5
2
0.0005
0.001
0.0015
0.002
0.0025
0.005
0.01
0.015
0.02
0.025
2.5
0 1 2 3 4 5 6 7 8 9
0 5 10 15 20 25 30 35 40

Figure 5A

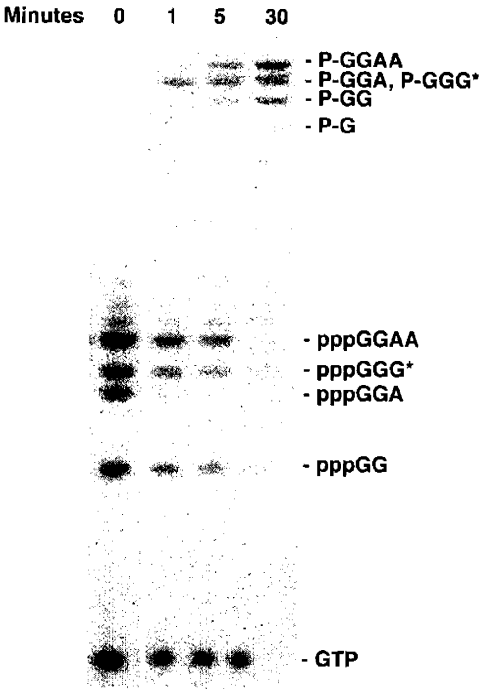

Minutes 0 1 5 30
- P-GGAA
- P-GGA, P-GGG*
- P-GG
- P-G
- pppGGAA
- pppGGG*
- pppGGA
- pppGG
- GTP

Figure 5B

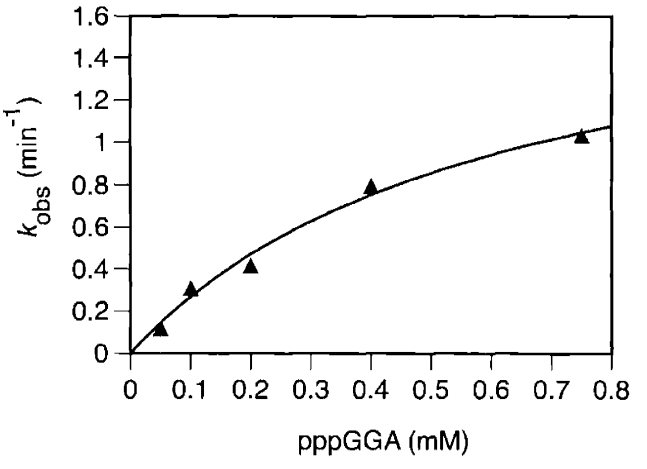
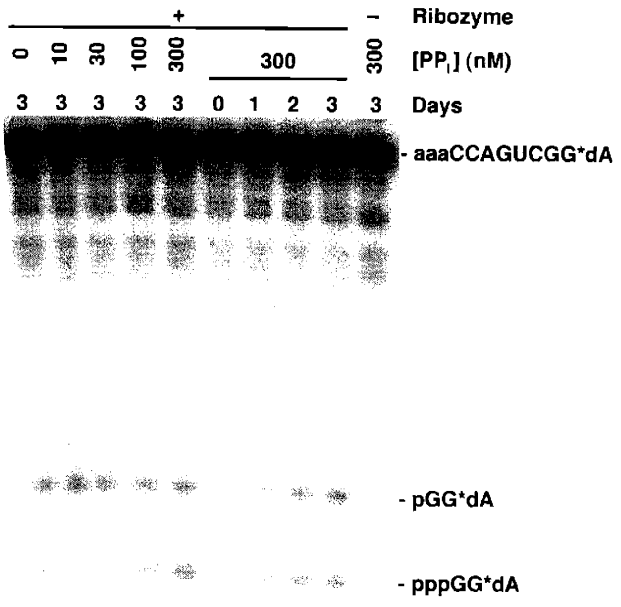


Figure 6


+
–
Ribozyme
0
10
30
100
300
300
300
[PPi] (nM)
3
3
3
3
3
0
1
2
3
3
Days
- aaaCCAGUCGG*dA
- pGG*dA
- pppGG*dA

# Chapter II

## Metal Ion Requirements for Structure and Catalysis of an RNA Ligase Ribozyme.

The work presented in this chapter was a collaborative effort between myself and Nick Bergman. Nick measured the chemistry rate at high pH using $Ca^{2+}$ preincubation, and I performed all other experiments.

[1]Abbreviations: EDTA, Ethylenediaminetetraacetic acid; EGTA, Ethylene glycol-bis[β-aminoethyl ether]-*N*,*N*,*N*′*N*′-tetraacetic acid; MES, 2-[*N*-Morpholino]ethanesulfonic acid; BES, *N*,*N*-bis[2-Hydroxyethyl]-2-aminoethanesulfonic acid; EPPS, *N*-[2-Hydroxyethyl]piperazine-*N*′-[3-propanesulfonic acid]; CHES, 2-[*N*-cyclohexylamino]-ethanesulfonic acid; RNase P, Ribonuclease P.

[2] These Hill coefficients were interpreted assuming that the rate-limiting step does not change at low $Mg^{+2}$ (<10 mM) in the presence of additional cations ($K^+$, $Ca^{2+}$, or $Co(NH_3)_6^{3+}$). This assumption that the rate of folding (or another step preceding chemistry) does not make a significant contribution to the observed rate can be justified as follows. The rate of folding ($k_{fold}$) is 60 $min^{-1}$ at 60 mM $Mg^{2+}$ and 200 mM $K^+$ and is independent of pH from pH 7.0–9.0 (discussed later). At 0.5 mM $Mg^{2+}$ and pH 6.0, the lower limit for the rate of folding is 2 $min^{-1}$, based on the pH dependence of activity at 0.5 mM $Mg^{2+}$ (data not shown), and the upper limit is at least 60 $min^{-1}$, if folding is independent of pH and $Mg^{2+}$ concentration. For folding to have an effect on the observed rate in the presence of $Mg^{2+}$ plus $K^+$, $Ca^{2+}$, or $Co(NH_3)_6^{3+}$, the rate of folding would have to decrease by at least 1000-fold at the lowest $Mg^{2+}$ concentration tested. While such a large decrease is not inconceivable, it would be surprising given that the ribozyme is well-folded at low concentrations ($[M^{2+}]_{1/2}^{fold} \leq 1$ mM) of $Mg^{2+}$ (in the presence or absence of $K^+$), $Ca^{2+}$, or $Co(NH_3)_6^{3+}$, as discussed later.

ABSTRACT: The class I ligase, a ribozyme previously isolated from random sequence, catalyzes a reaction similar to RNA polymerization, positioning its 5′-nucleotide via a Watson-Crick base pair, forming a 3′,5′-phosphodiester bond between its 5′-nucleotide and the substrate, and releasing pyrophosphate. Like most ribozymes, it requires metal ions for structure and catalysis. Here, we report the ionic requirements of this self-ligating ribozyme. The ligase requires at least five $Mg^{2+}$ for activity and has a $[Mg^{2+}]_{1/2}$ of 70–100 mM. It has an unusual specificity for $Mg^{2+}$; there is only marginal activity in $Mn^{2+}$ and no detectable activity in $Ca^{2+}$, $Sr^{2+}$, $Ba^{2+}$, $Zn^{2+}$, $Co^{2+}$, $Cd^{2+}$, $Pb^{2+}$, $Co(NH_3)_6^{3+}$, or spermine. All tested cations other than $Mg^{2+}$, including $Mn^{2+}$, inhibit the ribozyme. Hill analysis in the presence of inhibitory cations suggested that $Ca^{2+}$ and $Co(NH_3)_6^{3+}$ inhibit by binding at least two sites, but they appear to productively fill a subset of the required sites. Inhibition is not the result of a significant structural change, since the ribozyme assumes a nativelike structure when folded in the presence of $Ca^{2+}$ or $Co(NH_3)_6^{3+}$, as observed by hydroxyl-radical mapping. As further support for a nativelike fold in $Ca^{2+}$, ribozyme that has been prefolded in $Ca^{2+}$ can carry out the self-ligation very quickly upon the addition of $Mg^{2+}$. Ligation rates of the prefolded ribozyme were directly measured and proceed at 800 $min^{-1}$ at pH 9.0.

Metal ions are essential for the activity of most ribozymes, functioning in electrostatic shielding, structure, and directly in catalysis (*1*). Natural ribozymes vary widely in their metal ion specificity, from the large ribozymes (Group I intron, Group II intron, and RNase $P^1$), which require a divalent metal, preferably $Mg^{2+}$, for activity, to the promiscuous self-cleaving ribozymes (hammerhead, hairpin, VS, and HDV), which are active in many different cations (*2-13*). Hammerhead, hairpin and VS ribozymes are even active in nonmetal ions such as $NH_4^+$ (*12, 13*). Ribozymes selected in vitro also exhibit a wide range of metal ion requirements. Some, like the ribonucleolytic leadzyme, are active only in the cations in which they were selected (*14-16*). Others exhibit activity in the presence of metal ions which were not present during their selection. For example, an acyl transferase ribozyme, which was selected in the presence of $Mg^{2+}$ and $K^+$ only, is active in a wide variety of cations, including $Co(NH_3)_6^{3+}$ (*17, 18*).

The class I ligase, which was also selected in the presence of only $Mg^{2+}$ and $K^+$, performs a reaction similar to RNA polymerization, positioning its 5′-nucleotide via a Watson–Crick base pair, forming a 3′,5′-phosphodiester bond between its 5′-nucleotide and the substrate, and releasing pyrophosphate (Figure 1) (*19*). In fact, a version of this ribozyme can act as a general RNA polymerase, accurately extending a primer up to 14 nucleotides (*20*). Not only does the ligase perform the same reaction as naturally occurring RNA and DNA polymerases, but the stereospecificity of its preference for sulfur substitutions at the nonbridging oxygens on the α-phosphate matches that of these enzymes, raising the possibility that a catalytic mechanism common to all polymerases extends to ribozyme-mediated

polymerization as well (*21, 22*). The ligase is also one of the fastest ribozymes, among both natural ribozymes and those selected in vitro (*23*). These characteristics raise a number of interesting questions, including how the ligase's metal ion requirements compare to that of other ribozymes, how these requirements compare to those of polymerases made of protein, and to what degree its selection in the presence of only $Mg^{2+}$ and $K^{+}$ has constrained its use of metal ions.

To address these questions, we surveyed the metal ion requirements for activity and folding of the ligase. This ribozyme has an unusual $Mg^{2+}$ specificity for activity and was inhibited by all other cations tested. However, the ligase can achieve its native, global structure in the presence of several different cations, as determined by hydroxyl radical mapping. That the structure observed by this technique is similar to the active conformation was demonstrated by prefolding the ribozyme in $Ca^{2+}$ and initiating the reaction with the addition of $Mg^{2+}$ and EGTA, thus changing the rate-limiting step from folding to catalysis. The rate, 800 $min^{-1}$ at pH 9.0 and 60 mM $Mg^{2+}$, is one of the fastest rates yet observed for ribozyme catalysis.

## MATERIALS AND METHODS

*RNA preparation.* The class I ligase ribozyme (b1–207, Genbank #U26413) was transcribed in vitro from a plasmid template linearized with *Ear*I and gel purified on a 6% denaturing polyacrylamide gel (*23*). The substrate was an RNA–DNA hybrid (5′-aaaCCAGUC; DNA bases lowercase) synthesized by standard phosphoramidite chemistry. It was 5′-radiolabeled using [$^{32}$P]γ-ATP or [$^{33}$P]γ -ATP and T4 polynucleotide kinase.

*Manual kinetic assays.* Metal salts were the highest purity available. RNA, water, buffer, spermine, and non-transition metal salts were extracted with diphenylthiocarbazone in chloroform to remove contaminating transition metal ions (*6, 24*). Residual chloroform did not affect the reaction rates (data not shown). The ribozyme was renatured in water by heating at 85 °C for 2 min and cooling at 22 °C for 2 min. Reactions were carried out with 2 $\mu$M ligase ribozyme and ~25 nM substrate in 50 mM MES, pH 6.0, at 22 °C (or room temperature for incubations <1 min). Reactions were started by adding [$^{32}$P]-labeled substrate, buffer, and metal ions simultaneously and stopped by the addition of at least 1 vol 8 M urea and EDTA in excess of the divalent metal. Product and substrate were separated on 20% denaturing polyacrylamide gels and quantified with a Fujix BAS 2000 phosphorimager.

*Rapid-quench flow assays.* Ribozyme reactions requiring time points earlier than 5 seconds were performed in an RQF-3 rapid-quench flow apparatus (KinTek corporation, Austin, TX). Ribozyme in water was heated to 80 °C for 2 min, and then incubated at 22 °C for 2 min. EDTA was added to a concentration of 1 mM to chelate any trace divalent metal contaminants, and radiolabeled substrate was then added. To start the reaction, the ribozyme-substrate complex was mixed with an equal volume of reaction buffer. Final reaction conditions were 1 $\mu$M ligase ribozyme, ≤16 nM substrate, 60 mM $MgCl_2$, 200 mM KCl, 0.6 mM EDTA, and 50 mM BES (pH 7.0), BES (pH 7.4), EPPS (pH 8.0), EPPS (pH 8.5), or CHES (pH 9.0). Reactions were quenched with 0.5 vol 0.5 M EDTA, and then collected and analyzed by gel electrophoresis and phosphorimaging. Reactions in which $CaCl_2$ was used to pre-fold the ribozyme–substrate complex were performed as above, except that $CaCl_2$ (3 mM) was

added to the ribozyme–substrate complex after addition of substrate, and the reaction buffer was supplemented with 5 mM EGTA. At pH 7.0, rates measured with rapid-quench flow matched previous manual measurements of ligation (*23*).

*Rate determination.* Rates were determined using eq 1,

$$\text{fraction reacted} = F_a(1 - e^{-kt}) \qquad (1)$$

where $t$ equals time, $k$ is the observed rate of catalysis, and $F_a$ equals the fraction of substrate that reacts in an initial burst. In rapid-quench flow experiments in which conformational interconversion was negligible, reaction data were fit to eq 1 using $F_a = 0.7$ to adjust for the fraction of enzyme in an active conformation in 60 mM $MgCl_2$ (*23*) (N. H. B. and D. P. B., unpublished data). In all manual kinetic experiments, the rates were not corrected for the amount of ribozyme in the active conformation ($F_a = 1$), because the inactive fraction may be dependent on the identity and concentration of the metal ions (*2, 25*), and it was important not to discount the misfolded molecules in comparing the relative reaction rates in the presence of various metal ions. To fit the data to a single exponential, only times points in which ≤35% of the substrate had reacted were used.

*Hydroxyl radical mapping.* Structural studies were performed on the ligated product because the ribozyme does not fold properly in the absence of substrate (N. H. B. and D. P. B., unpublished data). 10 μM ribozyme and 3 μM [$^{33}$P]-labeled substrate were reacted to completion in 10 mM $MgCl_2$ and 50 mM MES, pH 6.0. The $Mg^{2+}$ was chelated with EDTA, and the ribozyme was precipitated to remove the EDTA/$Mg^{2+}$ complex. After resuspension, the

reacted ribozyme was extracted with diphenylthiocarbazone. The product was renatured by heating at 85 °C for 2 min and cooling at 22 °C for 2 min. The renatured product (final concentration = 1 $\mu$M) was mixed with 20 mM MES, pH 6.0 and various monovalent and/or divalent metals, then incubated at 22 °C with 2 mM $NH_4Fe(II)SO_4$, 2.2 mM EDTA and 2 mM ascorbic acid (*26, 27*). After 15 min, the cleavage reaction was quenched with 50 mM thiourea. The cleavage reactions were run on 10% denaturing polyacrylamide gels and quantified by phosphorimaging. After normalizing for loading differences and the overall amount of cleavage in each lane, "protection factors" were determined by dividing the number of counts of a nucleotide in the unfolded ribozyme (0 mM $Mg^{2+}$ at 60 °C) by the number of counts of that nucleotide in the "folded" ribozyme (various metal concentrations at 22 °C) (*26*). Except for nucleotide C68, the spectrum of protected residues observed in the renatured product was identical to that seen in previous experiments in which the product was not precipitated and renatured, but was mapped immediately after ligation (N. H. Bergman, N. Lau, V. Lehnert, E. Westhof, and D. P. Bartel, manuscript in preparation).

As a standard for complete folding, each experiment included the ribozyme folded in 10 mM $MgCl_2$ (N. H. Bergman, N. Lau, V. Lehnert, E. Westhof, and D. P. Bartel, manuscript in preparation). The magnitude of the protection factors for ribozymes folded in 10 mM $MgCl_2$ varied in different experiments, due to differences in the overall degree of cleavage and to differences in the amount of signal over the background radiation of the exposed gels. To compare experiments performed on different days and run on different gels, the protection factors of each nucleotide were normalized to the average protection of that nucleotide in 10

mM $MgCl_2$ (the average of four experiments). To assess the global structure of the ligase in the presence of various cations, protection factors of the most protected nucleotides (A46–C48 and G71–A74) were averaged.

## RESULTS

*Metal requirements for activity.* To study the metal ion requirements for folding and catalysis of the class I ligase, the single-turnover, self-ligating version (Figure 1) was chosen because it offered a simpler context relative to the polymerase and multiple-turnover derivatives (*23, 28*). Previous experiments established that chemistry is rate-limiting from pH 5.7–7.0 in 60 mM $MgCl_2$ and 200 mM KCl, but the kinetic analysis was not extended to higher pH because the rates could not be accurately measured by manual pipetting (*23*). In addition, no other ionic conditions have been explored. We therefore began this study by determining the metal ion requirements of the ligase at pH 6.0.

Varying the concentration of $Mg^{2+}$ revealed a high optimum (>200 mM; Figure 2A). The observed rate is dependent on pH even at very low $Mg^{2+}$ concentrations (log-linear dependence, slope ~1), so chemistry appears to be rate-limiting throughout the range of concentrations tested (data not shown). In the absence of $K^+$, the Hill coefficient between 0.5 and 2 mM $Mg^{2+}$ is 3.5; at higher concentrations, the Hill coefficient decreases to ~1 (Figure 2B). While this decrease may be due to inhibition by high $Mg^{2+}$ concentrations, it is probably partially or wholly due to the saturation of some $Mg^{2+}$ binding sites (*29*). At very low $Mg^{2+}$ concentrations, the Hill coefficient approaches the minimal number of ligands required for

activity. The nonintegral value of the Hill coefficient indicates either that some sites are becoming saturated at the lowest $Mg^{2+}$ concentration tested or that not all binding sites must be filled for the ligase to achieve some level of activity (*29*).

Because all previous kinetic analysis of the ligase was performed in the presence of 200 mM KCl, its $Mg^{2+}$ dependence was also assayed in the presence of $K^+$. Above 8 mM $Mg^{2+}$, $K^+$ does not alter the Hill coefficient significantly; however, $K^+$ inhibits the ribozyme at low $Mg^{2+}$ concentrations and increases the Hill coefficient to 5.0.[2] While it is possible that $K^+$ somehow increases the number of absolutely required binding sites, another explanation is that in the absence of $K^+$ there are also 5 (or more) required $Mg^{2+}$ binding sites that would only be revealed by Hill analysis at $Mg^{2+}$ concentrations below 0.5 mM. $K^+$ may compete with one or more $Mg^{2+}$ (by binding overlapping or nonoverlapping sites), increasing their apparent dissociation constant(s) such that these sites are not saturated at the $Mg^{2+}$ concentrations tested (0.5–4 mM), and the Hill coefficient is higher. Thus, the ligase requires at least five $Mg^{2+}$ ions for optimal activity, most of which have high affinities ($K_d$ < 2 mM). At $[Mg^{2+}]_{1/2}$ the Hill coefficient is ~1, suggesting that there is a single low-affinity $Mg^{2+}$ binding site with a $K_d$ near $[Mg^{2+}]_{1/2}$, 70–100 mM (assuming that there is no inhibition by high $Mg^{2+}$ concentrations).

We next surveyed the activity of the ligase in the presence of other cations. No activity was detected in $Ca^{2+}$, $Sr^{2+}$, $Ba^{2+}$, $Zn^{2+}$, $Co^{2+}$, $Cd^{2+}$, $Pb^{2+}$, $Co(NH_3)_6^{3+}$, or spermine (limits of detection, $1 \times 10^{-5}$ $min^{-1}$ for $Ca^{2+}$ and $1 \times 10^{-6}$ $min^{-1}$ for the rest), and only marginal activity was detected in $Mn^{2+}$ (Figure 3). $Mn^{2+}$ and $Mg^{2+}$ have similar sizes and coordination

geometries, and most natural ribozymes, as well as RNA and DNA polymerases, are very active when $Mn^{2+}$ is substituted for $Mg^{2+}$ (*1-6, 8, 9, 30*). In contrast, the activity of the ligase is orders of magnitude lower in the presence of $Mn^{2+}$. At the optimal $Mn^{2+}$ concentration (2 mM), the observed rate was 50-fold less than in 2 mM $Mg^{2+}$ and 1500-fold less than at the optimal $Mg^{2+}$ concentration. The activity decreased at higher $Mn^{2+}$ concentrations, and was undetectable above 20 mM. For the hammerhead and an acyl transferase ribozyme, a decrease in activity at high $Mn^{2+}$ concentrations was attributed to the formation of insoluble metal hydroxides at pH ≥8.0 (*17, 31*). However, this is unlikely at pH 6.0 (*31, 32*). If the affinity of $Mn^{2+}$ were merely lower than the affinity of $Mg^{2+}$ for required metal ion binding sites, the activity of the ligase would plateau at a higher metal ion concentration. Because the activity decreases with increasing $Mn^{2+}$ concentration, it is more probable that $Mn^{2+}$ binds to inhibitory sites, possibly perturbing the active structure by coordinating to the nitrogens of the bases (*33*).

*Inhibition by divalent metals.* Although the ligase achieves full activity only in the presence of $Mg^{2+}$, it is possible that other cations can successfully fill certain sites. In the reactions of RNase P and the Group I intron, cations that do not promote activity on their own can reduce the amount of $Mg^{2+}$ required for catalysis, providing evidence for different classes of metal ion binding sites (*2, 5*). To explore whether this is true for the ligase, it was assayed in the presence of 4 mM $MgCl_2$ and 0, 0.1, 1 or 10 mM additional cation (Table 1). Every divalent metal ion tested was found to inhibit the ligase. The transition metals $Zn^{2+}$, $Co^{2+}$, $Cd^{2+}$, and $Pb^{2+}$ were particularly potent, substantially inhibiting the ligase at micromolar concentrations. This effect was not due to degradation of the ribozyme—only $Pb^{2+}$ at high

concentrations (10 mM) cleaved the ribozyme noticeably (data not shown). The alkaline earth metals were less effective inhibitors, with $Ca^{2+}$ being more potent than $Sr^{2+}$ and $Ba^{2+}$, which have larger radii. Despite having low levels of activity in the presence of $Mn^{2+}$, the ligase was inhibited quite effectively by this cation. Experiments at other $Mg^{2+}$ concentrations failed to reveal that any other cation, including those in Table 1 as well as $K^+$, $Co(NH_3)_6^{3+}$, and spermine, could stimulate the activity of the ligase (data not shown). Thus, only $Mg^{2+}$ can fulfill all the catalytic and structural requirements of the ligase, but determining if a subset of the required metal ion binding sites are more tolerant required another approach.

If cations other than $Mg^{2+}$ can successfully fill some of the metal ion binding sites, it may be reflected in the Hill coefficient when both $Mg^{2+}$ and another cation are present[2]. Competitive inhibition should increase the apparent dissociation constants, shifting the curve to higher $Mg^{2+}$ concentrations, while appropriate binding by the other cation should decrease the Hill coefficient at low $Mg^{2+}$ concentrations where these sites are visible. A decrease in the Hill coefficient, however, cannot establish the number of sites productively bound by the competing cation, because the other cation may not bind the same sites as $Mg^{2+}$ (*34*). $Co(NH_3)_6^{3+}$, $Ca^{2+}$, and $Mn^{2+}$ disrupt $Mg^{2+}$ binding to at least the lowest affinity site, raising $[Mg^{2+}]_{1/2}$ to ~200 mM (Figure 4A,B, data not shown). Below 4 mM $Mg^{2+}$, the Hill coefficient is 3.1 in 60 $\mu$M $Co(NH_3)_6^{3+}$, compared to 3.5 in $Mg^{2+}$ alone or 5.0 in $Mg^{2+}$ plus 200 mM $K^+$. If activity is completely cooperative (i.e., all metal ion binding sites must be filled to observe any activity), at least 2 of the ≥5 required sites are saturated below 4 mM $Mg^{2+}$ in the presence or absence of $Co(NH_3)_6^{3+}$. Unless $Co(NH_3)_6^{3+}$ has no affinity for these saturated sites, it must be able to fill

them (or other functionally equivalent sites) without inhibiting the ribozyme. However, if the ribozyme is active even when not all sites are filled (i.e., activity is partially cooperative or noncooperative), this result provides no information about whether $Co(NH_3)_6^{3+}$ can productively fill any of the metal ion binding sites. It may merely inhibit the ribozyme.

In contrast, the Hill coefficient between 1 and 10 mM $Mg^{2+}$ is 2.0 in the presence of 5 mM $Ca^{2+}$ (Figure 4B). If activity is completely cooperative, fewer $Mg^{2+}$ are required in the presence of $Ca^{2+}$ than in its absence, indicating that $Ca^{2+}$ can replace $Mg^{2+}$ at some sites (or at other functionally equivalent sites). If activity is not completely cooperative, the lower Hill coefficient in the presence of $Ca^{2+}$ indicates that there are a higher proportion of ribozymes incompletely bound by $Mg^{2+}$ contributing to activity. Thus, in either case, binding by $Ca^{2+}$ reduces the number of $Mg^{2+}$ required for activity.

The role of $Co(NH_3)_6^{3+}$ and $Ca^{2+}$ was also addressed by varying the inhibitor at constant $Mg^{2+}$ concentration (Figure 4C,D). In this alternative Hill plot, the reaction rate in the presence of inhibitor is compared to the rate in its absence, rather than to the maximum rate. The data fit a curve with a limiting slope at high inhibitor concentrations equal to the number of inhibitor binding sites (*29*). At 10 mM $Mg^{2+}$, the Hill coefficient for inhibition approaches 2.2 for $Co(NH_3)_6^{3+}$ and 1.9 for $Ca^{2+}$. This number probably approaches the actual number of inhibitory binding sites of $Ca^{2+}$, because the extent of inhibition is large (97.4%–99.8%). The extent of inhibition by $Co(NH_3)_6^{3+}$ (60.2%–94.5%), however, may be insufficient to definitively conclude that there are only two inhibitory binding sites. The results of the two Hill analyses suggest

that $Ca^{2+}$ has two inhibitory binding sites but is able to productively fill some sites required for activity. $Co(NH_3)_6^{3+}$ may have more than two inhibitory binding sites, but whether it can fill any required sites is inconclusive.

*Structural metal ion requirements.* To investigate whether cations other than $Mg^{2+}$ can fill structural metal ion binding sites, we performed hydroxyl-radical mapping on the ligated product in the presence of various cations (Figure 5). This technique probes the global structure of nucleic acid molecules by preferentially cleaving solvent-accessible residues near the surface of a folded molecule (*35, 36*). In the absence of $Mg^{2+}$, 200 mM $K^+$ causes little, if any, of the native structure to form, while protection from cleavage decreases at 1 mM $Mg^{2+}$ if 200 mM $K^+$ is also present (Figure 5B). It is not clear whether this decrease is significant; if so, this result would correspond with the effect of $K^+$ on catalysis, in which the inhibition by $K^+$ at low $Mg^{2+}$ concentrations would result in structural destabilization. At higher levels of $Mg^{2+}$, $K^+$ makes little difference in the global structure of the ribozyme. Clearly, $K^+$ does not stabilize the structure or reduce the concentration of $Mg^{2+}$ required for folding.

In the presence of $Ca^{2+}$ alone, the ligase achieves the same global structure as in $Mg^{2+}$ (Figure 5C). Not only is the pattern of protection the same (except C68, which is not reproducibly protected in $Mg^{2+}$), but the concentration of metal ions required for folding is similar—$[M^{2+}]_{1/2}^{fold}$ for both metal ions is ≤1 mM (Figure 5D; unpublished results). The low metal ion concentration required for folding is in sharp contrast to the high concentration of

$Mg^{2+}$ optimal for catalysis. Apparently, the structural metal ion binding sites have a high affinity for divalent cations and can accommodate $Ca^{2+}$ as well as $Mg^{2+}$.

$Co(NH_3)_6^{3+}$ can also fold the ligase into a nativelike structure (Figure 5E). Much lower concentrations are required; $[Co(NH_3)_6^{3+}]_{1/2}^{fold}$ is between 10 and 30 $\mu$M. That a ~40-fold higher concentration of $Mg^{2+}$ is required to achieve the same degree of protection was not unexpected due to the stronger electrostatic interaction of the trivalent cobalt complex with the ribozyme (*17, 37, 38*). However, the maximum extent of protection is less than that of ribozymes folded in $Mg^{2+}$, and the level of protection decreases at higher $Co(NH_3)_6^{3+}$ concentrations. Hill analysis suggested that $Co(NH_3)_6^{3+}$ may be able to fill fewer required sites than $Ca^{2+}$ (Figure 4A,B). It is possible that appropriate binding of a couple sites can fold the ribozyme into a nativelike structure, but stable folding requires binding of one or more additional sites, which $Co(NH_3)_6^{3+}$ cannot bind. Alternatively, $Co(NH_3)_6^{3+}$ may bind to low-affinity sites in a manner that disrupts the structure.

*Kinetics of the ribozyme prefolded in $Ca^{2+}$.* The conclusion from hydroxyl-radical mapping studies is that $Ca^{2+}$ and $Co(NH_3)_6^{3+}$ can play structural roles in the ligase, even though they inhibit catalysis. A caveat to this conclusion is that these studies were performed on the product of the reaction, which may not be in the active conformation. In addition, hydroxyl-radical mapping is a low-resolution technique. Even though the structures folded by $Mg^{2+}$, $Ca^{2+}$, and $Co(NH_3)_6^{3+}$ are globally similar, there may be differences in the fine details. To determine whether the structure achieved by the ligase in the presence of $Ca^{2+}$ is the active one

or at least a productive intermediate, we examined the kinetics of self-ligation reactions in which the ribozyme was preincubated with $Ca^{2+}$ (Figure 6A).

In the absence of $Ca^{2+}$ preincubation, chemistry is rate-limiting and increases with increasing pH in a log-linear manner from pH 5.5 to pH 7.0 (*23*). At pH ≥8.0 the rate plateaus at 60 $min^{-1}$, indicating that another step in the reaction, presumably folding, becomes rate-limiting above pH 7.0 (Figure 6B, squares). When the ribozyme–substrate complex was preincubated with $Ca^{2+}$ prior to the addition of $Mg^{2+}$, the rate was somewhat faster, but not as fast as expected if chemistry were rate-limiting (data not shown). This might be because the ligase is inhibited by $Ca^{2+}$ or because the ribozyme prefolded with $Ca^{2+}$ requires a conformation change before chemistry can occur.

To prevent inhibition of the ligation reaction by the $Ca^{2+}$ used to prefold the ribozyme–substrate complex, the $Mg^{2+}$ used to initiate the reaction was supplemented with EGTA, which chelates $Ca^{2+}$ 100 times more readily than $Mg^{2+}$ (*39*). When the reaction was started by the simultaneous addition of $Mg^{2+}$ and EGTA (Figure 6A), ligation rates were log-linear with pH up to pH 8.5 (slope = 1.0; Figure 6B, circles). At pH >8.5, ligation rates rose more slowly with pH, possibly due to deprotonation of the bases and concomitant destabilization of the structure. The ability of $Ca^{2+}$-preincubation to extend the range in which ligation rates are log-linear with pH suggests that chemistry is rate-limiting at high pH for the prefolded ribozyme–substrate complex. Thus, the $Ca^{2+}$-prefolded ribozyme (and perhaps the

product folded in the presence of $Mg^{2+}$, $Ca^{2+}$, or $Co(NH_3)_6^{3+}$) must be very close to the active structure. Any slight, $Mg^{2+}$-induced conformation changes must be exceedingly fast.

## DISCUSSION

*Role of metal ions in structure and catalysis.* The class I ligase, like many other natural ribozymes and ribozymes selected in vitro, has multiple metal ion binding sites with differing roles, affinities, and specificities. Scheme 1 summarizes the relationships between metals and ribozyme folding and chemistry (where $R_u$, $R_f$, P, S, and $PP_i$ denote unfolded ribozyme, folded ribozyme, ligated ribozyme, substrate, and pyrophosphate, respectively). Folding of the ribozyme is rapid (60 $min^{-1}$), but can become rate-limiting at higher pH. Low metal ion concentrations are required for folding, which proceeds equally well in the presence of $Mg^{2+}$ or $Ca^{2+}$. In contrast, much higher $Mg^{2+}$ concentrations are required for catalysis, and only $Mg^{2+}$ is sufficient for optimal catalysis. Overall, $\geq 5$ $Mg^{2+}$ are required for folding and catalysis, and the presence of $Ca^{2+}$ reduces the number of required $Mg^{2+}$ to $\geq 2$, suggesting that two metal ions may be required for catalysis.

Scheme 1 was built with several simplifications. Folding is represented as a single step, representing attainment of catalytic competence. In reality, folding is probably several distinct steps, as has been seen for other RNAs (*40-43*). Furthermore, Scheme 1 begins with substrate bound to the ribozyme prior to folding and disregards the contribution of $Mg^{2+}$-dependent folding prior to substrate association. This is justified because at high pH, where folding is rate-limiting, incubation in $Mg^{2+}$ before adding high concentrations of substrate does not increase the ligation rate (data not shown), suggesting that the ligase does not stably exist in a

folded form without bound substrate. Thus, ligation rates largely reflect reactions in which folding is preceded by substrate association. The rate of unfolding is described as slow because the equilibrium between $R_u$•S and $R_f$•S appears to lie toward the folded form (at least at high $Mg^{2+}$ concentrations). This is based on the observation that reaction time courses for the $Ca^{2+}$-preincubation experiments fit a single-exponential curve, even at high pH, as expected for a homogenous population of molecules that are correctly folded at the moment of $Mg^{2+}$ addition. The chemical step ($R_f$•S ➡ P + $PP_i$) is represented as irreversible because previous work has shown that the reverse reaction (pyrophosphorolysis with formation of a triphosphate) is $10^7$-times slower than the forward reaction (*21*).

The $Mg^{2+}$ preference for catalysis is in stark contrast to RNA and DNA polymerases made of protein and most natural ribozymes, which have similar levels of activity in $Mg^{2+}$ and $Mn^{2+}$ (*1-6, 8, 9, 30*). One (or more) low-affinity $Mg^{2+}$ appears to be required for catalysis; $[Mg^{2+}]_{1/2}$ for folding is ≤1 mM, but $[Mg^{2+}]_{1/2}$ for activity is over 70 mM. $Co(NH_3)_6^{3+}$ and $Ca^{2+}$ probably compete with $Mg^{2+}$ for the low-affinity site (or fill nonoverlapping, mutually exclusive sites), both because they can fill some or all high-affinity structural sites and because they increase the apparent $K_d$ of the low-affinity site to ~200 mM. Likewise, $Mn^{2+}$ increases the apparent dissociation constant of the low-affinity site, due to suboptimal binding to this site or disruptive binding to inhibitory sites (data not shown).

While it is apparent from kinetic and structural studies that $Mn^{2+}$, $Ca^{2+}$, and $Co(NH_3)_6^{3+}$ can fill a subset of the required metal ion binding sites, one question that remains is whether

they inhibit by binding (or preventing $Mg^{2+}$ from binding) to structurally or catalytically required sites. $Mn^{2+}$ must be able to fulfill all of the structural and catalytic requirements to some degree, because the ribozyme is slightly active in the presence of $Mn^{2+}$ alone. However, if the ribozyme retains some activity even when not all stimulatory sites are bound, it is possible that $Mn^{2+}$ cannot fill all $Mg^{2+}$ binding sites. Inhibition by $Mn^{2+}$ could then be the result of disruptively overlapping some $Mg^{2+}$ binding sites, in addition to binding other sites with no affinity for $Mg^{2+}$.

$Co(NH_3)_6^{3+}$ and $Ca^{2+}$ must be able to fill some or all structurally required sites, because they both fold the ribozyme into nativelike structures. Hill analysis suggested that $Ca^{2+}$ may fill more of the required sites than $Co(NH_3)_6^{3+}$. Coupled to the lesser extent of hydroxyl-radical protection in the presence of $Co(NH_3)_6^{3+}$ relative to $Mg^{2+}$ or $Ca^{2+}$, this suggests that $Co(NH_3)_6^{3+}$ may not fill all structurally required sites. Nevertheless, the high level of structure formed in the presence of $Co(NH_3)_6^{3+}$, as well as its competition for the low-affinity, catalytic site suggests that $Co(NH_3)_6^{3+}$ also competes with $Mg^{2+}$ for catalytic metal ion binding sites (or binds mutually exclusive sites), rather than merely disrupting the ribozyme fold.

Unless $Ca^{2+}$ subtly perturbs the structure, it also inhibits by competing for catalytic sites (or mutually exclusive sites). Additional support for this comes from measuring the activity at high pH after preincubating the ribozyme–substrate complex with $Ca^{2+}$. Under these conditions, chemistry is rate-limiting up to approximately pH 8.5, so the rate of any conformation changes must significantly exceed 500 $min^{-1}$, the rate of the chemical step at pH

8.5. In comparison, tRNA folds 10-fold faster than this, the catalytic domain of RNase P folds at the same rate, the fast steps of Group I intron folding proceed 8-fold slower, and large conformational changes of several ribozymes occur on the order of minutes (*41, 44-49*).

While it is not possible to determine whether the inhibitory ions compete directly with $Mg^{2+}$ by binding overlapping sites or whether their binding to more distant sites alters the conformation and occludes $Mg^{2+}$ binding sites, their inability to adequately fill some required sites does suggest that certain chemical attributes of $Mg^{2+}$ are vital for ribozyme activity. $Ca^{2+}$, for instance, has a larger radius and a higher $pK_a$ than $Mg^{2+}$. Before assessing the importance of these characteristics for ligase activity, it will be useful to discuss the role of $Ca^{2+}$ in uncatalyzed RNA ligation and in some ribozyme-catalyzed reactions.

The uncatalyzed ligation reaction, in which two RNAs aligned by base-pairing to a template are ligated to form a 3′,5′-phosphodiester bond and release pyrophosphate, works with a wide variety of metal ions, and the rate varies inversely with the metal ion's $pK_a$ (*50*). The differences between the rates in the presence of different metal ions correlates with the difference between their $pK_a$s, such that the rate with $Mn^{2+}$ ($pK_a$ = 10.6) is 6–10-fold higher than with $Mg^{2+}$ ($pK_a$ = 11.4), and the rate with $Mg^{2+}$ is 13-fold higher than with $Ca^{2+}$ ($pK_a$ = 12.9). Likewise, a 2′,5′ ligase ribozyme, which was selected in vitro and performs a reaction similar to the class I ligase, demonstrates the same relationship between its activity and the $pK_a$ of the metal ion (*51*). Most in vitro-selected and natural ribozymes, however, tend to show more complex behavior. The Group I ribozyme prefers $Mn^{2+}$ over $Mg^{2+}$ in a class of binding

sites thought to be required for catalysis, but it is inactive in $Ca^{2+}$ (*2*). The 1.5 unit difference between the $pK_a$s of $Mg^{2+}$ and $Ca^{2+}$ cannot account for the total inability of $Ca^{2+}$ to support catalysis. Instead, it was shown that $Mn^{2+}$ and $Ca^{2+}$ bind to a site adjacent to the previously identified $Mg^{2+}$ binding sites. $Mn^{2+}$ at this site stimulates catalysis, while $Ca^{2+}$ inhibits the chemical step (*34*). Likewise, the slightly lower $pK_a$ of $Ca^{2+}$ is unlikely to account for its inhibition of the ligase ribozyme. It is more probable that $Ca^{2+}$ distorts the active site, either because it is too large or it binds to an incorrect site. However, whether $Mn^{2+}$ is preferred over $Mg^{2+}$ at the active site is impossible to discern because $Mn^{2+}$ inhibits the ligase ribozyme by binding to other sites.

In addition to a possible size requirement at certain metal ion binding sites, inhibition by $Co(NH_3)_6^{3+}$ suggests that inner sphere contact is required at some sites, since $Co(NH_3)_6^{3+}$ is exchange-inert and similar in size and shape to a fully hydrated $Mg^{2+}$ (*10, 11, 52*). Because the ribozyme is well-folded by $Ca^{2+}$ and mostly folded by $Co(NH_3)_6^{3+}$, it is likely they both disrupt $Mg^{2+}$ binding to catalytic sites, especially the low-affinity site. The correct positioning of the active site may therefore require binding by at least one metal ion with a radius smaller than that of $Ca^{2+}$ and inner-sphere coordination of one or more metal ions.

*Mechanism of the class I ligase.* $Ca^{2+}$ inhibits the ligase by binding to two sites, yet the ribozyme appears to be correctly folded. This raises the tantalizing possibility that the ligase uses a mechanism requiring two metal ions, similar to that of protein-enzyme polymerases (*22, 53-55*). In this mechanism, one metal ion coordinates the 3′-OH of the primer and the pro-$R_P$-

oxygen of the α-phosphate belonging to the incoming nucleotide (*54*), and the other contacts the nonbridging oxygens of the triphosphate, all through inner-sphere coordination (*53-55*). The first metal ion promotes the attack of the 3′-OH on the α-phosphate by lowering its affinity for hydrogen, and the second stabilizes the pyrophosphate leaving group (*22*). This mechanism is common to several nonhomologous polymerases (*22, 53-55*), as well as to enzymes that catalyze other phosphoryl transfer reactions (albeit with differing stereochemical preferences), including the 3′–5′ exonuclease of DNA polymerase I and alkaline phosphatase (*56-59*).

In the reaction catalyzed by the ligase ribozyme, one of the required metal ions is undoubtedly chelated to the triphosphate, with a $K_d$ between 10 and 100 μM, if binding to the triphosphate is not influenced by additional interactions with the ribozyme (*39, 50*). Given that $K^+$ has a $K_d$ of 100-300 mM for nucleoside triphosphates (*39*), the apparent dissociation constant of $Mg^{2+}$ would be 17-300 μM in the presence of 200 mM KCl, assuming competitive inhibition (eq 2).

$$K_d^{app} = K_d^{Mg^{2+}}\left(1 + \frac{[K^+]}{K_d^{K^+}}\right) \qquad (2)$$

If the $K_d$ of $Mg^{2+}$ is at the upper end of this range, competition by $K^+$ for this site should be represented in the higher Hill coefficient at low $Mg^{2+}$ concentrations. A second catalytically required $Mg^{2+}$ binds the low-affinity site, possibly coordinating the 3′-OH and the α-phosphate. The dissociation constant of $Mg^{2+}$ bound to this site is expected to be less than or equal to the $[Mg^{2+}]_{1/2}$ of the uncatalyzed reaction (150–300 mM) (*50*). Because the $[Mg^{2+}]_{1/2}$ is actually 70–100 mM, any additional ribozyme contacts may be weak, or the ribozyme may facilitate chemistry by a strategy involving $Mg^{2+}$ in substrate destabilization (*60*).

While there is no evidence supporting the specific metal contacts required by the two-metal-ion mechanism for the ligase reaction, there are several parallels between reactions catalyzed by the ligase ribozyme and proteinaceous polymerases that are consistent with this mechanism. The ligase, which forms the catalytic core of an RNA polymerase ribozyme, has the same stereochemical preference for sulfur substitution at the α-phosphate as polymerases composed of protein (*20, 21*), and both the ligase and polymerases made of protein can tolerate only $Mg^{2+}$ or $Mn^{2+}$ at their active sites (*30*). In addition, kinetic and structural analysis in the presence of $Mg^{2+}$ and $Ca^{2+}$ suggested that the ligase may require two metal ions for catalysis. If this model is correct and the polymerase ribozyme uses the same mechanism as the ligase, the two-metal-ion mechanism may be truly universal for polymerases (*22*).

*The class I ligase and the RNA world.* The ligase not only performs the chemistry of RNA polymerization, a reaction crucial to the RNA world hypothesis, but it is also one of the fastest ribozymes. Under optimal conditions, self-ligation proceeds at a rate of 800 $min^{-1}$, comparable to that of the RNA component of RNase P (360 $min^{-1}$) (*61*), and both ribozymes have rates 10 times faster than the observed rates of most other ribozymes (*17, 31, 51, 62-64*). However, the optimal conditions for these ribozymes differ from each other and from physiological conditions.

To make a fair comparison to natural ribozymes and evaluate the ligase as a reasonable facsimile of an RNA world enzyme, it would be useful to estimate its level of activity under

physiological and possible primordial conditions. The physiological concentration of free $Mg^{2+}$ is 1-2 mM, free $Ca^{2+}$ is approximately 0.1 μM, and total monovalent cations are about 200 mM (*33, 65-67*). Ignoring polycations that may inhibit the ribozyme further, the ligase would exhibit only 0.001 of its optimal activity, at best, under these conditions. With a $Mg^{2+}$ concentration of about 50 mM, seawater (a possible primordial environment) may seem more favorable (*33, 68*); however, the high concentration of $Ca^{2+}$ (~10 mM) would severely inhibit the ligase (*33*). While it is difficult to compare the ligase to natural ribozymes under physiological conditions, the Group I intron and RNase P are considerably more tolerant of $Ca^{2+}$ at high $Mg^{2+}$ concentrations (*2, 69, 70*), and the small self-cleaving ribozymes are very active in the presence of many cations, including $Ca^{2+}$ (*6-9*).

It is not surprising that the ligase is less tolerant of other metals than natural ribozymes, considering that it experienced only $Mg^{2+}$ and $K^+$ during its selection. The question then arises whether it would be worthwhile to perform in vitro selections in a broader range of metal ions. There is no clear trend demonstrating that the metal ion requirements of ribozymes are defined by the metal ions present during their selection. Although selected in the presence of only $Mg^{2+}$ and $K^+$, like the class I ligase, an acyl transferase and a 2′,5′ RNA ligase are active in a wide variety of metal ions (*17, 18, 51*). Other ribozymes selected in a mixture of metal ions are limited to activity in only a few and may be inhibited by some ions (*71-75*). If lower $Mg^{2+}$ concentrations and a wider variety of metal ions had been used in the ligase selection, the class I ligase would have been at a severe disadvantage. As it was, this ligase was the only member of its family isolated from a pool of $10^{15}$ sequences (*19*). Despite the lack of ionic stringency

during the selection, a ribozyme as catalytically efficient as many in nature was discovered. Selecting variants of the ligase with lower $Mg^{2+}$ optima and higher metal tolerance, as demonstrated previously with the Group I intron and RNase P (*76, 77*), may lead to the discovery of ribozymes that are active in conditions simulating intracellular or possible prebiotic environments.

**ACKNOWLEDGEMENT**

We thank Vernon Anderson for providing peroxynitrous acid used in hydroxyl radical mapping and Jamie Cate for helpful discussions. We also thank Alex Kravchuk and Ed Curtis for useful comments on the manuscript.

## REFERENCES


1. Pyle, A. M. (1996) *Met. Ions Biol. Sys. 32*, 479-520.
2. Grosshans, C. A., and Cech, T. R. (1989) *Biochemistry 28*, 6888-6894.
3. Deme, E., Nolte, A., and Jacquier, A. (1999) *Biochemistry 38*, 3157-3167.
4. Gardiner, K. J., Marsh, T. L., and Pace, N. R. (1985) *J. Biol. Chem. 260*, 5415-5419.
5. Guerrier-Takada, C., Haydock, K., Allen, L., and Altman, S. (1986) *Biochemistry 25*, 1509-1515.
6. Dahm, S. C., and Uhlenbeck, O. C. (1991) *Biochemistry 30*, 9464-9469.
7. Chowrira, B. M., Berzal-Herranz, A., and Burke, J. M. (1993) *Biochemistry 32*, 1088-1095.
8. Collins, R. A., and Olive, J. E. (1993) *Biochemistry 32*, 2795-2799.
9. Suh, Y.-A., Kumar, P. K. R., Taira, K., and Nishikawa, S. (1993) *Nucleic Acids Res. 21*, 3277-3280.
10. Hampel, A., and Cowan, J. A. (1997) *Chem. Biol. 4*, 513-517.
11. Young, K. J., Gill, F., and Grasby, J. A. (1997) *Nucleic Acids Res. 25*, 3760-3766.
12. Murray, J. B., Seyhan, A. A., Walter, N. G., Burke, J. M., and Scott, W. G. (1998) *Chem. Biol. 5*, 587-595.
13. Curtis, E. A., and Bartel, D. P. (2001) *RNA 7*, 546-552.
14. Pan, T., and Uhlenbeck, O. C. (1992) *Biochemistry 31*, 3887-3895.
15. Pan, T., Dichtl, B., and Uhlenbeck, O. C. (1994) *Biochemistry 33*, 9561-9565.
16. Sugimoto, N., and Ohmichi, T. (1996) *FEBS Lett. 393*, 97-100.
17. Suga, H., Cowan, J. A., and Szostak, J. W. (1998) *Biochemistry 37*, 10118-10125.

18. Lohse, P. A., and Szostak, J. W. (1996) *Nature 281*, 442-444.

19. Ekland, E. H., Szostak, J. W., and Bartel, D. P. (1995) *Science 269*, 364-370.

20. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001) *Science 292*, 1319-1325.

21. Glasner, M. E., Yen, C. C., Ekland, E. H., and Bartel, D. P. (2000) *Biochemistry 39*, 15556-15562.

22. Steitz, T. A. (1998) *Nature 391*, 231-232.

23. Bergman, N. H., Johnston, W. K., and Bartel, D. P. (2000) *Biochemistry 39*, 3115-3123.

24. Holmquist, B. (1988) *Methods in Enzymology 158*, 6-12.

25. Tanner, N. K., and Cech, T. R. (1985) *Nucleic Acids Res. 13*, 7741-7758.

26. Pan, T. (1995) *Biochemistry 34*, 902-909.

27. Celander, D. W., and Cech, T. R. (1991) *Science 251*, 401-407.

28. Ekland, E. H., and Bartel, D. P. (1996) *Nature 382*, 373-376.

29. Segel, I. H. (1975) *Enzyme Kinetics*, John Wiley and Sons, Inc., New York.

30. Mildvan, A. S., and Loeb, L. A. (1979) *CRC Crit. Rev. Biochem. 6*, 219-244.

31. Dahm, S. C., Derrick, W. B., and Uhlenbeck, O. C. (1993) *Biochemistry 32*, 13040-13045.

32. Cotton, F. A., and Wilkinson, G. (1988) *Advanced Inorganic Chemistry*, 5th ed., John Wiley and Sons, New York.

33. Pan, T., Long, D. M., and Uhlenbeck, O. C. (1993) in *The RNA World* (Gesteland, R. F., and Atkins, J. F., Eds.), pp 271-302, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

34. Shan, S.-O., and Herschlag, D. (2000) *RNA 6*, 795-813.

35. Tullius, T. D., Dombroski, B. A., Churchill, M. E. A., and Kam, L. (1987) *Methods in Enzymology 155*, 537-558.

36. Latham, J. A., and Cech, T. R. (1989) *Science 245*, 276-282.

37. Hampel, K. J., Walter, N. G., and Burke, J. M. (1998) *Biochemistry 37*, 14672-14682.

38. Nesbitt, S., Hegg, L. A., and Fedor, M. J. (1997) *Chem. Biol. 4*, 619-630.

39. Dawson, R. M. C., Elliott, D. C., Elliott, W. H., and Jones, K. M. (1986) *Data for Biochemical Research*, 3rd ed., Clarendon Press, Oxford.

40. Pan, T., and Sosnick, T. R. (1997) *Nat. Struct. Biol. 4*, 931-938.

41. Sclavi, B., Sullivan, M., Chance, M. R., Brenowitz, M., and Woodson, S. A. (1998) *Science 279*, 1940-1943.

42. Treiber, D. K., Rook, M. S., Zarrinkar, P. P., and Williamson, J. R. (1998) *Science 279*, 1943-1946.

43. Crothers, D. M., Cole, P. E., Hilbers, C. W., and Shulman, R. G. (1974) *J. Mol. Biol. 87*, 63-88.

44. Draper, D. E. (1996) *Nat. Struct. Biol. 3*, 397-400.

45. Fang, X.-W., Pan, T., and Sosnick, T. R. (1999) *Nat. Struct. Biol. 6*, 1091-1095.

46. Zhuang, X., Bartley, L. E., Babcock, H. P., Russell, R., Ha, T., Herschlag, D., and Chu, S. (2000) *Science 288*, 2048-2051.

47. Zarrinkar, P. P., and Williamson, J. R. (1994) *Science 265*, 918-924.

48. Zarrinkar, P. P., Wang, J., and Williamson, J. R. (1996) *RNA 2*, 564-573.

49. Walter, N. G., Hampel, K. J., Brown, K. M., and Burke, J. M. (1998) *EMBO J. 17*, 2378-2391.

50. Rohatgi, R., Bartel, D. P., and Szostak, J. W. (1996) *J. Am. Chem. Soc. 118*, 3332-3339.

51. Landweber, L. F., and Pokrovskaya, I. D. (1999) *Proc. Natl. Acad. Sci. U S A 96*, 173-178.

52. Jou, R., and Cowan, J. A. (1991) *J. Am. Chem. Soc. 113*, 6685-6686.

53. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) *Science 264*, 1891-1903.

54. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) *Nature 391*, 251-258.

55. Kiefer, J. R., Mao, C., Braman, J. C., and Beese, L. S. (1998) *Nature 391*, 304-307.

56. Steitz, T. A., and Steitz, J. A. (1993) *Proc. Natl. Acad. Sci. U S A 90*, 6498-6502.

57. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., and Steitz, T. A. (1988) *Proc. Natl. Acad. Sci. U S A 85*, 8924-8928.

58. Beese, L. S., and Steitz, T. A. (1991) *EMBO J. 10*, 25-33.

59. Kim, E. E., and Wyckoff, H. W. (1991) *J. Mol. Biol. 218*, 449-464.

60. Narlikar, G. J., Gopalakrishnan, V., McConnell, T. S., Usman, N., and Herschlag, D. (1995) *Proc. Natl. Acad. Sci. U S A 92*, 3668-3672.

61. Beebe, J. A., and Fierke, C. A. (1994) *Biochemistry 33*, 10294-10304.

62. Esteban, J. A., Banerjee, A. R., and Burke, J. M. (1997) *J. Biol. Chem. 272*, 13629-13639.

63. Herschlag, D., and Cech, T. R. (1990) *Biochemistry 29*, 10159-10171.

64. Olive, J. E., and Collins, R. A. (1998) *Biochemistry 37*, 6476-6484.

65. Snavely, M. D. (1990) *Met. Ions Biol. Sys. 26*, 155-175.

66. Maguire, M. E. (1990) *Met. Ions Biol. Sys. 26*, 135-153.

67. Lodish, H., Baltimore, D., Berk, A., Zipursky, S. L., Matsudaira, P., and Darnell, J. (1995) *Molecular Cell Biology*, 3rd ed., Scientific American Books, New York.

68. Lahav, N. (1999) *Biogenesis: Theories of Life's Origin*, Oxford University Press, New York.

69. McConnell, T. S., Herschlag, D., and Cech, T. R. (1997) *Biochemistry 36*, 8293-8303.

70. Smith, D., and Pace, N. R. (1993) *Biochemistry 32*, 5273-5281.

71. Seelig, B., and Jaschke, A. (1999) *Chem. Biol. 6*, 167-176.

72. Illangasekare, M., Sanchez, G., Nickles, T., and Yarus, M. (1995) *Science 267*, 643-647.

73. Illangasekare, M., and Yarus, M. (1997) *J. Mol. Biol. 268*, 631-639.

74. Huang, F., and Yarus, M. (1997) *Biochemistry 36*, 6557-6563.

75. Huang, F., and Yarus, M. (1997) *Biochemistry 36*, 14107-14119.

76. Lehman, N., and Joyce, G. F. (1993) *Nature 361*, 182-185.

77. Frank, D. N., and Pace, N. R. (1997) *Proc. Natl. Acad. Sci. U S A 94*, 14355-14360.

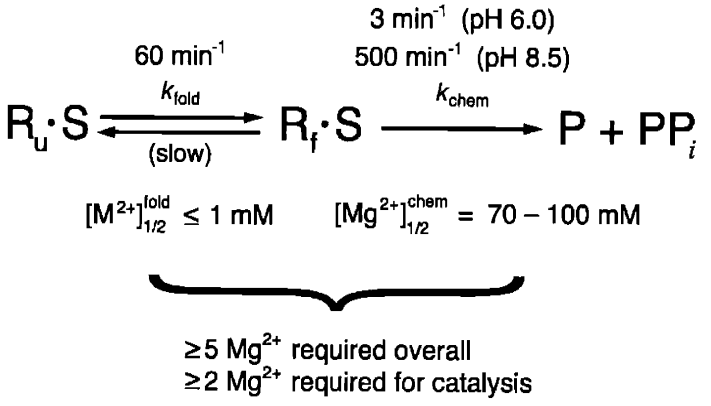
Table 1: Relative Reaction Rates in the Presence of Inhibitory Divalent Cations.

| | Relative Rate[a] | | |
|---|---|---|---|
| | 0.1 mM $M^{2+}$ | 1 mM $M^{2+}$ | 10 mM $M^{2+}$ |
| $Mn^{2+}$ | n.d.[b] | 0.56 | 0.032 |
| $Ca^{2+}$ | n.d. | 0.63 | 0.006 |
| $Sr^{2+}$ | n.d. | 0.95 | 0.116 |
| $Ba^{2+}$ | n.d. | 0.81 | 0.046 |
| $Zn^{2+}$ | 0.59 | 0.007 | <0.0006 |
| $Co^{2+}$ | 0.53 | 0.0009 | <0.0006 |
| $Cd^{2+}$ | 0.09 | <0.0006 | <0.0006 |
| $Pb^{2+}$ | 0.11 | <0.0006 | n.d. |

[a]Reactions were performed under standard conditions in 4 mM $MgCl_2$ plus 0.1, 1, or 10 mM additional divalent ion, and rates are reported relative to the reaction rate in 4 mM $MgCl_2$ only (0.25 $min^{-1}$).

[b]n.d., not determined.

Scheme 1


3 min-1 (pH 6.0)
60 min-1
500 min-1 (pH 8.5)
kfold
kchem
Ru·S
Rf·S
P + PPi
(slow)
[M2+]1/2fold ≤ 1 mM
[Mg2+]1/2chem = 70 – 100 mM
≥5 Mg2+ required overall
≥2 Mg2+ required for catalysis

## FIGURE LEGENDS

Figure 1. Secondary structure of the class I ligase. The ribozyme catalyzes the attack of the 3′-OH of the substrate on its 5′-triphosphate, forming a 3′,5′-phosphodiester bond and releasing pyrophosphate.

Figure 2. Dependence of activity on $Mg^{2+}$ at pH 6.0. (A) The $Mg^{2+}$ optimum is very high ($K_{1/2}$ = 70–100 mM). (B) Hill analysis demonstrates that the ligase is inhibited by $K^+$ at low $Mg^{2+}$ concentrations. In the absence of $K^+$ (O), the Hill coefficient is 3.5 for $Mg^{2+}$ concentrations between 0.5 and 2 mM and is ~1 at $Mg^{2+}$ concentrations above 8 mM. In the presence of 200 mM KCl (×), the Hill coefficient is 5.0 for $Mg^{2+}$ concentrations between 0.5 and 4 mM and is ~1 at $Mg^{2+}$ concentrations above 8 mM. The $k_{max}$ was estimated to be 2.8 $min^{-1}$ from the plateau of the curve (increasing its value 10-fold does not affect the Hill coefficients).

Figure 3. Very low ligase activity in $Mn^{2+}$ (O) compared to $Mg^{2+}$ (●). Reactions were at pH 6.0.

Figure 4. Hill analysis in the presence of the indicated inhibitory metals. (A,B) Conventional Hill plots in which the concentration of $Mg^{2+}$ was varied and $k_{max}$ = 2.8 $min^{-1}$. The slope decreases with increasing $Mg^{2+}$, and the Hill coefficient is given as a range calculated from lines fit to data from lower (solid line) or higher (dashed line) $Mg^{2+}$ concentrations. Data for $Mg^{2+}$ alone is shown with a dotted line for comparison. (A) In the presence of 60 μM $Co(NH_3)_6Cl_3$, the Hill coefficient ranges from 3.1 to 0.9. (B) In the presence of 5 mM $CaCl_2$,

the Hill coefficient ranges from 2.0 to 1.4. (C,D) Hill analysis in which the concentration of $Mg^{2+}$ was constant and the inhibitor concentration was varied. The rate of the reaction in 10 mM $MgCl_2$ in the absence of inhibitor, $k_o$, is equal to 0.6 $min^{-1}$, and $k_i$ is the rate of the reaction in 10 mM $MgCl_2$ plus the indicated concentration of inhibitor. The slope approaches the number of inhibitory cations bound (*29*). (C) Varying $Co(NH_3)_6Cl_3$ concentration results in a limiting slope (solid line) of –2.2. (D) Varying $CaCl_2$ concentration results in a limiting slope (solid line) of –1.9.

Figure 5. Hydroxyl-radical mapping of the class I ligase in the presence of different metal ions. Protection factors are relative to cleavage in the absence of metals at 60 °C. The average protection is the average of the protection factors of the most protected residues (A46–C48, G71–A74). (A) Pattern of protection in 10 mM $MgCl_2$, 22 °C (average of four experiments). (B) Comparison of $Mg^{2+}$-dependent folding in the absence (black) or presence (white) of 200 mM KCl. (C) Protection by low concentrations of $MgCl_2$ or $CaCl_2$. Only the region between nucleotides 45 and 76 is shown. (D) Comparison of folding in the presence of either $MgCl_2$ (black) or $CaCl_2$ (white). (E) Comparison of folding in $Mg^{2+}$ or $Co(NH_3)_6^{3+}$.

Figure 6. Reaction rates of the ligase prefolded in $CaCl_2$. (A) Scheme for prefolding experiment. $Ca^{2+}$ was added to the renatured substrate-ribozyme complex in water. The reaction was then started by the simultaneous addition of $MgCl_2$ and EGTA. (B) pH dependence of ligation rates with (●) or without (□) $CaCl_2$ preincubation. Reactions were in 60 mM $MgCl_2$ and 200 mM KCl.

Figure 1

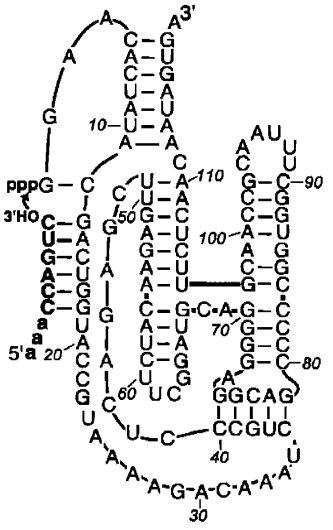


Figure 2

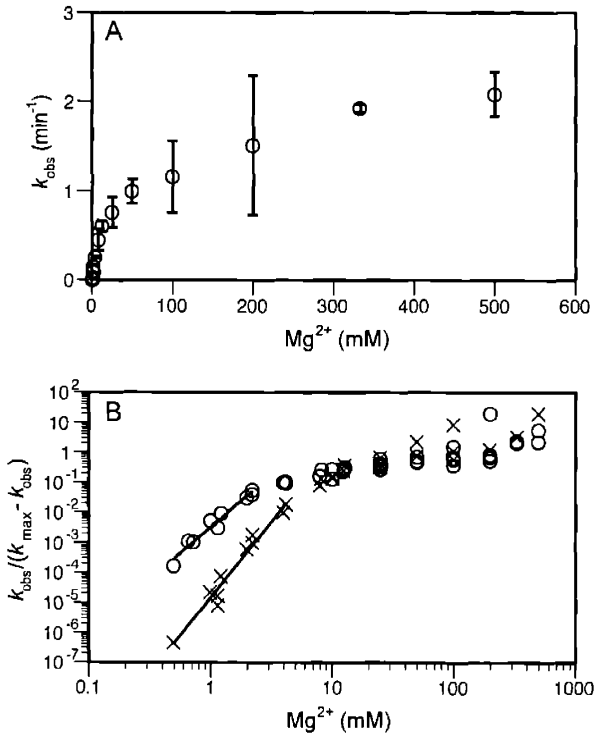


Figure 3

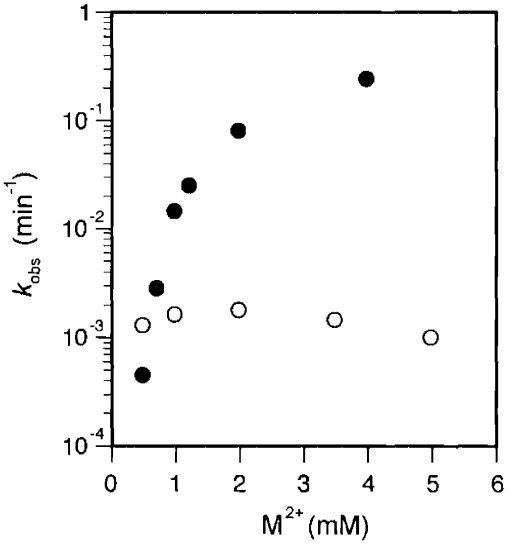


Figure 4

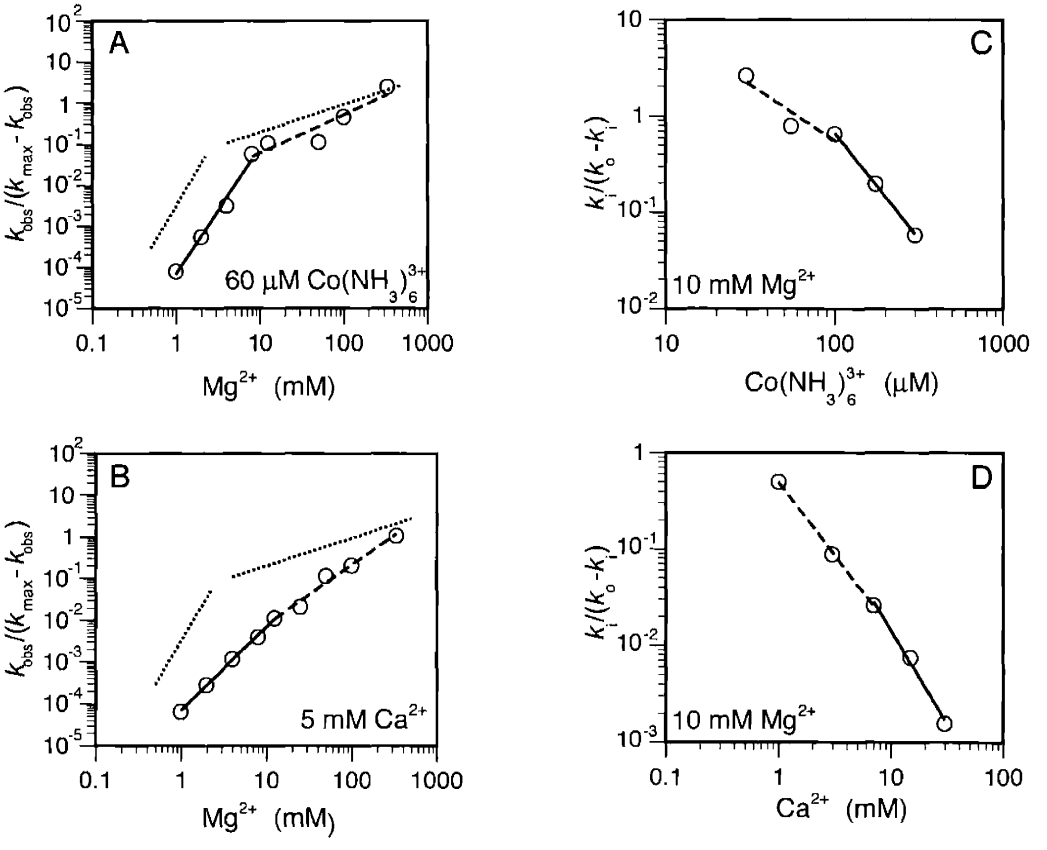

A
60 μM Co(NH3)6 3+
Mg2+ (mM)
B
5 mM Ca2+
Mg2+ (mM)
C
10 mM Mg2+
Co(NH3)6 3+ (μM)
D
10 mM Mg2+
Ca2+ (mM)

Figure 5

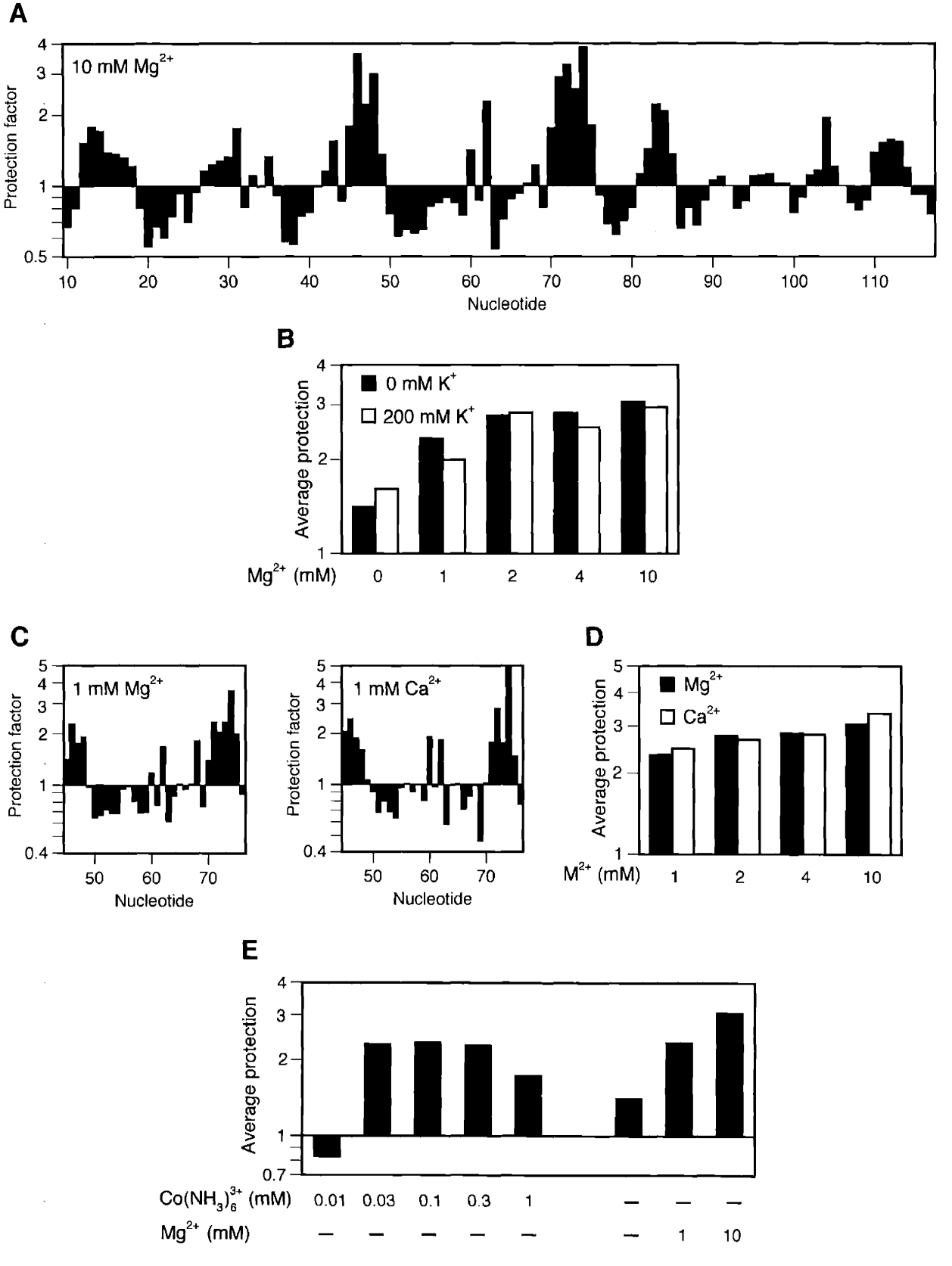
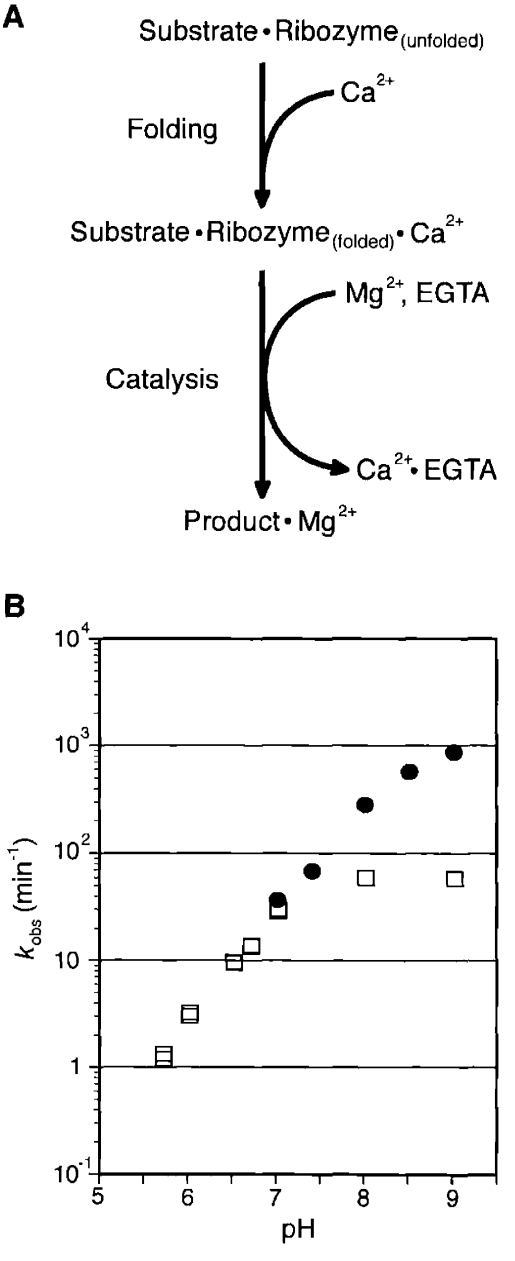

A
10 mM Mg2+
Protection factor
Nucleotide
B
0 mM K+
200 mM K+
Average protection
Mg2+ (mM)
C
1 mM Mg2+
1 mM Ca2+
Protection factor
Nucleotide
D
Mg2+
Ca2+
Average protection
M2+ (mM)
E
Average protection
Co(NH3)6 3+ (mM)
Mg2+ (mM)

Figure 6


A
Substrate • Ribozyme(unfolded)
Folding
Ca2+
Substrate • Ribozyme(folded) • Ca2+
Mg2+, EGTA
Catalysis
Ca2+ • EGTA
Product • Mg2+
B
kobs (min-1)
10^4
10^3
10^2
10
1
10^-1
5
6
7
8
9
pH

# Future Directions

This thesis has presented a characterization of the substrate and metal ion requirements of an RNA ligase ribozyme (Figure 1A,B). Several aspects of this ribozyme make it interesting in its own right, as well as a plausible starting point for an RNA polymerase ribozyme. Of primary interest is the mechanism of the ligase, as well as the precise functions and locations of the required metal ions. These experiments have also highlighted characteristics of the ligase ribozyme which, if improved, could greatly enhance the catalytic capabilities of the polymerase ribozyme, for which the ligase serves as the catalytic center.

***Reaction Mechanism of the Ligase Ribozyme***

Although the kinetics and structure of the ligase have been intensively studied, the most interesting question remains unanswered: what is the mechanism of the ligase? The reaction mechanisms of few ribozymes have been unequivocally demonstrated, and whether or not RNA, with its rather limited choice of functional groups, has an equally limited choice of catalytic strategies is an open question. In addition, all RNA and DNA polymerases made of protein use the same mechanism requiring two metal ions, even though some of them share no recognizable sequence similarity [1]. In this mechanism, one metal ion coordinates the 3′-OH of the primer and the pro-$R_P$-oxygen of the α-phosphate belonging to the incoming nucleotide [2], and the other contacts the nonbridging oxygens of the triphosphate, all through inner-sphere coordination [2-4]. Determining if the ligase uses this mechanism will demonstrate whether there is an alternative mechanism, or if it is likely that nonhomologous enzymes could have converged on the same mechanism.

Central to understanding the mechanism of the ligase ribozyme is determining the positions of the catalytic metal ions. One approach is to examine rescue of the activity of sulfur-substituted ribozymes by thiophilic metal ions [5]. Sulfur substitution disrupts $Mg^{2+}$ binding because $Mg^{2+}$ has a very low affinity for sulfur relative to its affinity for oxygen; transition metals such as $Mn^{2+}$, $Cd^{2+}$ and $Zn^{2+}$ have much higher affinities for sulfur [6]. The activity of ribozymes with sulfur substitutions at $Mg^{2+}$ binding sites is drastically reduced, but it can be partially restored in the presence of thiophilic metal ions which replace the required $Mg^{2+}$ [5].

Previous experiments demonstrated that the ligase and protein-enzyme polymerases have the same stereochemical preference for sulfur substitution at the α-phosphate of the triphosphate, which would be expected if the ligase uses the two-metal ion mechanism like polymerases made of protein [7]. The activity is slightly reduced if the substitution is at the pro-$S_P$ oxygen, as expected from an elemental effect on the transition state [8], but the activity is drastically reduced if the substitution is at the pro-$R_P$ oxygen. If a metal ion coordinates the pro-$R_P$ oxygen of the α-phosphate, then the activity of the pro-$R_P$ thio-substituted ribozyme may be partially restored if the reaction is carried out in the presence of a thiophilic metal ion; however, the thiophilic metal ion should not rescue the activity of the pro-$S_P$ thio-substituted ribozyme.

Initial attempts to observe metal ion rescue at the α-phosphate were unsuccessful (D. Shechner and D.P. Bartel, unpublished results). However, these experiments were performed at very high $Mn^{2+}$ concentrations. $Mn^{2+}$ inhibits the ribozyme severely at high concentration, probably disrupting its structure and making it difficult to detect small increases in the rate [9]. To thoroughly explore the possibility of $Mn^{2+}$ rescue, these

experiments could be repeated at low $Mn^{2+}$ concentrations, where inhibition by $Mn^{2+}$ is not as severe. These experiments could also be performed with low concentrations of $Cd^{2+}$ or $Zn^{2+}$, which are more thiophilic (although they are also more potent inhibitors than $Mn^{2+}$) [6, 9, 10].

***Localizing Metal Ion Binding Sites***

Another strategy for studying the role of metal ions in the structure and catalysis of the ligase is to select ribozyme variants that are active in different metal ions. The ligase ribozyme used in the experiments presented here is active almost exclusively in $Mg^{2+}$; there is only marginal activity in low $Mn^{2+}$ concentrations, and high $Mn^{2+}$ concentrations apparently disrupt the ribozyme's structure [9]. Selecting ligase variants that are active in the presence of $Mn^{2+}$ will have several benefits. First, variants that are not inhibited by $Mn^{2+}$ may offer a more successful alternative in the metal ion rescue experiments discussed above. Second, mutations in the metal-tolerant variants may suggest the locations of metal ion binding sites required for structure or catalysis. This method was used to identify a region of RNase P that helps form the active site, because variants active in $Ca^{2+}$ shared a specific mutation [11]. Because $Mn^{2+}$ disrupts the ribozyme's structure by binding sites which $Mg^{2+}$ does not bind [9], mutations in $Mn^{2+}$-tolerant ribozyme variants may be selected due to abrogation of $Mn^{2+}$-specific binding or subtle changes in required $Mg^{2+}$ binding sites to accommodate $Mn^{2+}$. To differentiate between these possibilities, a selection for $Ca^{2+}$-tolerant ligase variants could also be

performed. Mutations that enhance metal tolerance for both $Ca^{2+}$ and $Mn^{2+}$ may define the general location of required metal ion binding sites.

One difficulty in interpreting the meaning of mutations which alter metal specificity is that the mutation may be at some distance from the affected metal ion binding site. A direct method for determining the locations of metal ion binding sites is a variation on the metal ion rescue experiment described above, in which all possible phosphorothioate substitutions are simultaneously screened for activity [12]. $Mn^{2+}$-tolerant variants of the ligase ribozyme could be transcribed with the α-phosphorothioate version of each nucleotide at concentrations such that, on average, one phosphorothioate is incorporated per molecule. After reacting the thio-phosphorylated ribozymes with radiolabeled primer, the thiophosphate backbone could be cleaved with iodine [13]. If a particular thio substitution prevents the binding of a required metal ion (or otherwise disrupts the ribozyme's activity), that variant would react with the radiolabeled primer poorly, resulting in a faint band or a gap in the iodine cleavage ladder. To determine whether this gap represents a metal ion binding site, the experiment would be repeated in the presence of a thiophilic metal ion. If the thiophilic metal rescues activity, the ribozyme would react with the labeled primer to a greater extent, and there would be a stronger band at that position in the iodine cleavage ladder. The main disadvantage of this technique is that only the pro-$R_P$ oxygen can be substituted with sulfur co-transcriptionally [14]; as a result, some metal ion binding sites may be overlooked.

A simple method for determining the locations of metal ion binding sites regardless of their stereochemical preference is to analyze the pattern of metal-dependent hydrolysis [15, 16]. Observing $Mg^{2+}$-dependent cleavage is possible at elevated pH and

temperature; however, cleavage by $Mg^{2+}$ is weak. Cleavage by $Tb^{3+}$ is much stronger because its $pK_a$ is near neutral. It is also a good mimic of $Mg^{2+}$ because its radius, coordination geometry, and preference for oxygen are very similar to those of $Mg^{2+}$ [15]. If the ribozyme is properly folded by $Tb^{3+}$, strongly cleaved positions should indicate close proximity to a metal ion, especially if competition by $Mg^{2+}$ reduces cleavage. Some sites might still be missed, however, beacuase cleavage requires in-line attack geometry.

Discerning the roles of metal ions located by biochemical methods and their precise binding geometry will require a high resolution crystal structure. Visualizing the metal ions in the active site would be particularly enlightening, as it could definitively address the mechanism of the ligase. One problem might be the low electron density of $Mg^{2+}$, which is difficult to distinguish from water or hydrated $Na^+$ ions [17]. In addition, one of the $Mg^{2+}$ ions required for chemistry has a very high dissociation constant ($K_d$ = 70-100 mM), suggesting that this site might not be bound stably enough to be observed. However, the ribozyme folds in the presence of $Co(NH_3)_6^{3+}$, which bodes well for the prospect of incorporating electron-dense metals, at least in some structurally required sites [9]. Even though the ribozyme is inactive in $Co(NH_3)_6^{3+}$ (and other electron-dense metals), it may still bind the sites required for chemistry, albeit with an unknown level of distortion in the active site. If the active site metal ions cannot be observed, it may be necessary to select ligase variants which have greater affinity for the low-affinity catalytic $Mg^{2+}$ ion or variants that are active and well-folded in $Mn^{2+}$, which would be easier to observe crystallograpically.

*Improving the Polymerase Ribozyme*

A central tenet of the RNA world hypothesis is that the RNA genome should be able to replicate itself, and there has been considerable progress in efforts to discover a ribozyme capable of self-replication. The current leader is an RNA polymerase ribozyme which utilizes the ligase ribozyme as its catalytic core (Figure 1C) [18]. Although the polymerase ribozyme can extend a primer by an impressive 14 nucleotides, the reaction proceeds slowly, even at high $Mg^{2+}$ concentrations and pH. Detailed characterization of the ligase ribozyme suggests several ways in which the polymerase ribozyme might be improved.

The main reduction in the rate of the polymerase ribozyme was the result of deleting the second and third nucleotides of the ligase ribozyme (Figure 1A,B) [7]. These nucleotides served to orient the first nucleotide for attack by the 3′-hydroxyl group of the primer; in their absence the catalytic rate is 100-fold slower. In the context of the GTP ribozyme, which extends a primer by a single, templated nucleotide, these nucleotides were restored by using the trinucleotide pppGGA as the substrate. However, using trinucleotide substrates is not desirable for the polymerase, especially since the second and third nucleotides are untemplated and must be GA. A simple way to reincorporate these nucleotides into the polymerase ribozyme would be to extend the P2 oligonucleotide by adding GAA to its 5′ end (Figure 1D). Ideally, the stacking interaction between the terminal G and the incoming nucleotide would orient the nucleotide and increase the catalytic rate. Initial experiments show that this is too simplistic, however. The lengthened P2 oligonucleotide slightly increased the efficiency

of extension of the first few nucleotides, but extension was completely blocked when the polymerase reached a UC dinucleotide on the template which could pair with the GA at the end of the P2 oligonucleotide. This suggests that extending the P2 oligonucleotide will only be beneficial if the added nucleotides are docked into the ligase domain and cannot base-pair to the template. This could be tested by selecting improved polymerase ribozymes which use the lengthened P2. Because it may be difficult to adequately immobilize the GA dinucleotide when it is at the end of the oligonucleotide, another strategy would be to use a longer P2 with random sequence 5′ of the GA dinucleotide (Figure 1E). Selection could yield variants in which this randomized region forms secondary or tertiary interactions with the polymerase which help to constrain the movement of the GA dinucleotide.

Other aspects of the polymerase ribozyme that have not been optimized are its $Mg^{2+}$ requirement and tolerance of other metal ions. The $[Mg^{2+}]_{1/2}$ for the ligase ribozyme is 70-100 mM, due to the binding of a single, low-affinity $Mg^{2+}$ ion which is required in the chemical step, and the ligase is inhibited by all metal ions except $Mg^{2+}$ [9]. Although the metal requirements of the polymerase have not been studied in detail, its activity is best in 200 mM $Mg^{2+}$, suggesting that its $Mg^{2+}$ requirement has not changed substantially [18]. In addition, it is unlikely that the polymerase is more tolerant of other metal ions, since there were no substantial changes to the ligase domain during the selection of the polymerase. Increasing the ribozyme's affinity for $Mg^{2+}$ or altering its metal requirements may not greatly increase its reaction rate or efficiency under optimal conditions, but these improvements will be useful for the next generations of polymerase ribozymes. The ultimate goal is to use an RNA polymerase ribozyme capable of self-

replication as a simple model system for molecular evolution. The polymerase could even be incorporated into a lipid vesicle with other ribozymes to form a model "ribo-organism" with a minimal metabolism [19, 20]. If the polymerase ribozyme is to function in a biochemical pathway with other ribozymes that require different metal ions, it will have to function efficiently under a much wider set of conditions.

In total, the experiments characterizing the ligase have not uncovered any insurmountable obstacles that would render the current ligase and polymerase ribozymes dead ends in the continuing quest for a self-replicating ribozyme. Although the reaction rate of the polymerase is still quite slow compared to the self-ligating ribozyme (and polymerases made of protein), the increase in fidelity and extent of polymerization during its selection was remarkable. More remarkable is the fact that the polymerase shows some degree of processivity, even though this was not a criteria for its selection (M.S. Lawrence and D.P. Bartel, manuscript in preparation).

One trait necessary for a self-replicating ribozyme that may be difficult to attain without a carefully designed selection is helicase activity [21]. A helicase would be required for unraveling single-stranded templates and dissociating double-stranded RNA to free the nascent ribozyme. These activities might actually be combined: as the double-stranded RNA is being unraveled, one strand would be copied, while the other would be free to fold into an active ribozyme. A helicase could be selected from random sequence by first selecting a ribozyme that could catalyze strand exchange of a short helix (~5-10 nucleotides). For instance, the end of the pool could be base-paired to an oligonucleotide immobilized on beads, and active molecules could be selected based on their ability to dissociate themselves from the beads. The stringency could be increased

during the selection by increasing the length of the helix that must be dissociated. A helicase selected in this manner could then be linked to the polymerase ribozyme, and the three-domain polymerase (ligase domain, primer/template binding domain, and helicase domain) could be selected to optimize polymerization of long templates. A polymerase ribozyme selected in this step-wise manner will pave the way for the synthesis of minimal RNA-based life.

## REFERENCES


1. Steitz, T.A. (1998). A mechanism for all polymerases. *Nature* **391**, pp. 231-232.
2. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, pp. 251-258.
3. Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H. & Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science* **264**, pp. 1891-1903.
4. Kiefer, J.R., Mao, C., Braman, J.C. & Beese, L.S. (1998). Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* **391**, pp. 304-307.
5. Piccirilli, J.A., Vyle, J.S., Caruthers, M.H. & Cech, T.R. (1993). Metal ion catalysis in the Tetrahymena ribozyme reaction. *Nature* **361**, pp. 85-88.
6. Pecoraro, V.L., Hermes, J.D. & Cleland, W.W. (1984). Stability constants of $Mg^{2+}$ and $Cd^{2+}$ complexes of adenine nucleotides and thionucleotides and rate

constants for formation and dissociation of MgATP and MgADP. *Biochemistry* **23**, pp. 5262-5271.

7. Glasner, M.E., Yen, C.C., Ekland, E.H. & Bartel, D.P. (2000). Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. *Biochemistry* **39**, pp. 15556-15562.

8. Herschlag, D., Piccirilli, J.A. & Cech, T.R. (1991). Ribozyme-catalyzed and nonenzymatic reactions of phosphate diesters: rate effects upon substitution of sulfur for a nonbridging phosphoryl oxygen atom. *Biochemistry* **30**, pp. 4844-4854.

9. Glasner, M.E., Bergman, N.H. & Bartel, D.P. (2002). Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry* **41**, pp. 8103-8112.

10. Jaffe, E.K. & Cohn, M. (1978). Divalent cation-dependent stereospecificity of adenosine 5_-O-(2-thiotriphosphate) in the hexokinase and pyruvate kinase reactions. The absolute stereochemistry of the diastereoisomers of adenosine 5'-O-(2-thiotriphosphate). *J. Biol. Chem.* **253**, pp. 4823-4825.

11. Frank, D.N. & Pace, N.R. (1997). In vitro selection for altered divalent metal specificity in the RNase P RNA. *Proc. Natl. Acad. Sci. U.S.A.* **94**, pp. 14355-14360.

12. Christian, E.L. & Yarus, M. (1993). Metal coordination sites that contribute to structure and catalysis in the group I intron from Tetrahymena. *Biochemistry* **32**, pp. 4475-4480.

13. Gish, G. & Eckstein, F. (1988). DNA and RNA sequence determination based on phosphorothioate chemistry. *Science* **240**, pp. 1520-1522.

14. Eckstein, F. (1985). Nucleoside phosphorothioates. *Annu. Rev. Biochem.* **54**, pp. 367-402.

15. Sigel, R.K., Vaidya, A. & Pyle, A.M. (2000). Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.* **7**, pp. 1111-1116.

16. Hertweck, M. & Mueller, M.W. (2001). Mapping divalent metal ion binding sites in a group II intron by $Mn^{2+}$ and $Zn^{2+}$induced site-specific RNA cleavage. *Eur. J. Biochem.* **268**, pp. 4610-4620.

17. Holbrook, S.R., Sussman, J.L., Warrant, W., Church, G.M. & Kim, S.-H. (1977). RNA-ligand interactions: (I) magnesium binding sites in yeast $tRNA^{Phe}$. *Nucleic Acids Res.* **4**, pp. 2811-2820.

18. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, pp. 1319-1325.

19. Bartel, D.P. & Unrau, P.J. (1999). Constructing an RNA world. *Trends Cell Biol.* **9**, pp. M9-M13.

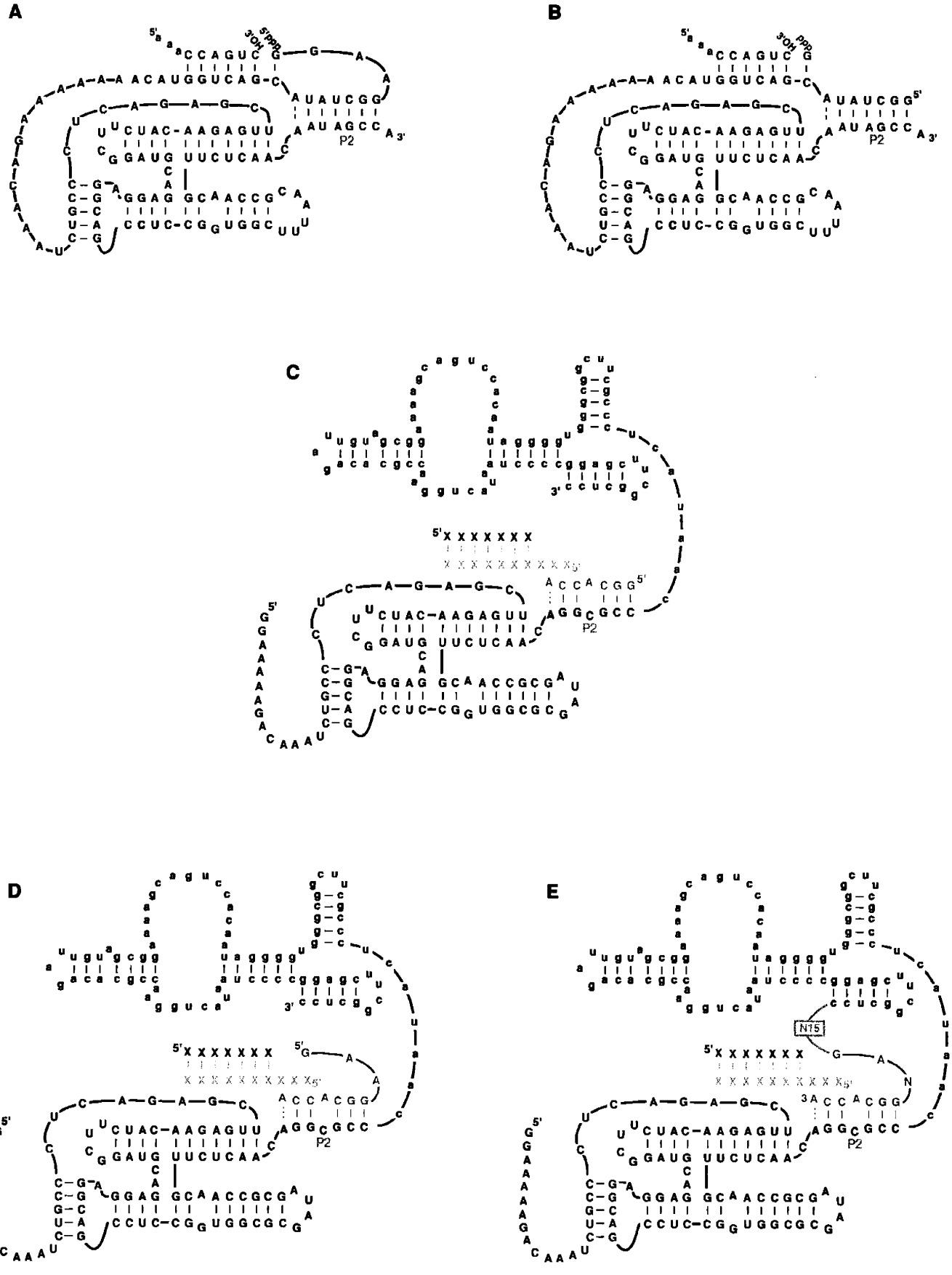
20. Szostak, J.W., Bartel, D.P. & Luisi, P.L. (2001). Synthesizing life. *Nature* **409 Suppl**, pp. 387-390.

21. Bartel, D.P. (1999). Re-creating an RNA replicase. In *The RNA World* (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds.), 2nd ed. pp. 143-181, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

## FIGURE LEGEND

Figure 1. The ligase and polymerase ribozymes. (A) The self-ligating ribozyme catalyzes the attack of the 3′-OH of the primer (red) on its 5′-triphosphate, forming a 3′,5′-phosphodiester bond and releasing pyrophosphate. (B) The GTP ribozyme was generated by deleting the first four nucleotides of the self-ligating ribozyme, and it extends the primer by a single, templated GTP (blue). (C) The polymerase ribozyme utilizes the ligase ribozyme as its catalytic core and has an additional domain (small letters) which is required for binding the primer-template complex (red and orange). The polymerase also requires an oligonucleotide (purple) to complete stem P2. (D) Supplying the polymerase ribozyme with a longer P2 oligonucleotide that restores the $2^{nd}$ and $3^{rd}$ nucleotides of the self-ligating ribozyme increases the efficiency of polymerization with some templates. (E) A pool in which the lengthened P2 is connected to the primer-template binding domain by a short random region (green) could be used to select polymerase ribozymes that utilize the lengthened P2 with any template.

# Figure 1



# Appendix A

# Cloning Developmentally Expressed microRNAs from Zebrafish

ABSTRACT: MicroRNAs (miRNAs) are a large class of noncoding RNAs that may be important regulators of gene expression in eukaryotes. In order to investigate their involvement in vertebrate development, we cloned small RNAs from zebrafish embryos at several different stages. We isolated ~140 miRNA candidates which are 21-23 nucleotides long and have the potential to form long bulged hairpin precursors with flanking genomic sequence. Most of these genes are conserved in several organisms, including the pufferfish *Tetraodon*, mouse, and human.

MicroRNAs (miRNAs) are a large class of noncoding RNAs that may be important regulators of gene expression in eukaryotes. The first two genes of this class to be characterized, *lin-4* and *let-7*, were initially discovered as temporally expressed genes that control the timing of larval development in *Caenorhabditis elegans* [1, 2]. These ~22 nucleotide (nt) RNAs are processed from larger precursors which form bulged hairpin loops [1, 3]. The enzyme that processes these small RNAs, Dicer, is also the ribonuclease which cleaves double-stranded RNA during RNA interference [4-7]. However, *lin-4* and *let-7* do not function in an RNAi-like manner; rather than causing the degradation of their mRNA targets, they bind to sites in the 3′-UTR and trigger translational repression [1, 2, 8, 9]. In addition, the processing of miRNAs differs from RNA degradation in RNA interference in that only one side of the hairpin precursor accumulates, while the other rapidly degrades [6].

Concerted efforts to clone small RNAs have identified more than 150 miRNAs in a wide variety of organisms, including *C. elegans*, *Drosophila melanogaster*, human, mouse, and *Arabidopsis thaliana* [10-15]. Although there is no sequence motif that unites this class, they all have the potential to pair with flanking genomic sequence to form long hairpin loops, or fold-back structures. In addition, miRNAs can be either developmentally or constitutively expressed [10, 11]. This suggests that miRNAs may regulate a wide variety of genes by a mechanism that is ubiquitous among animals and plants.

The identification of these abundant noncoding RNAs raises many questions regarding the functions of miRNAs. One important question is whether miRNAs play a vital role in vertebrate development, like *lin-4* and *let-7* in *C. elegans*. As a first step

toward answering this question, we cloned small RNAs from zebrafish embryos. Zebrafish was chosen because there is about 85% coverage of its genomic sequence, which would allow us to assess whether the cloned sequences form the fold-back structures with flanking genomic sequence that are hallmarks of miRNAs. In addition, zebrafish offers many prospects for functional studies. Small RNAs were isolated from several developmental stages, including two stages that occur prior to the mid-blastula transition when embryonic transcription is initiated [16]. Out of 4438 small RNAs that were sequenced, there were ~140 miRNA candidates which have potential secondary structural elements typical of other identified miRNAs. The number of expressed miRNAs appeared to increase during development, with very few miRNAs cloned before the mid-blastula transition. In addition, most of them were conserved in other organisms, suggesting that they serve important functions. Analysis of the frequency of the miRNA candidates cloned from each stage and the timing of their expression will allow us to select miRNAs with possible roles in development for functional studies.

## MATERIALS AND METHODS

*Isolating and cloning small RNAs.* Total RNA was isolated from zebrafish embryos (TLF strain) by incubating them in a 1 mg/ml solution of pronase (Sigma) for 5 min to remove the chorion and extracting the RNA using the Tri-Reagent protocol (Molecular Research Center, Inc.). Embryos were collected from 8 stages: zygote, 256 cell, germ ring, 75% epiboly, tailbud, 11 somite, 24 hours, and 48 hours. In order to maximize the chance of isolating rare miRNAs and to make predictions about their

developmental expression, the cloning procedure was performed separately for each stage.

Small RNAs were cloned by the procedure developed by Lau, et al. [10]. Briefly, RNAs of 18-28 nt were isolated from 200 μg (zygote and 256 cell) or 50 μg (all other stages) total RNA on 15% polyacrylamide gels. Purified RNA was ligated to a 3′-adaptor oligonucleotide using a pre-adenylated DNA oligonucleotide (A-5′-pp-5′-CTGTAGGCACCATCAAT-$OCH_3$) and RNA ligase (Amersham Pharmacia Biotech). After gel purification on a 10% polyacrylamide gel, the RNA was ligated to a 5′-adaptor oligonucleotide (ATCGTaggcaccugaaa; uppercase, DNA; lowercase RNA; Dharmacon Research) using RNA ligase and ATP. The ligated RNA was gel purified on a 10% polyacrylamide gel, reverse transcribed, and amplified by PCR. The DNA was digested with the nonpalindromic restriction enzyme BanI and ligated to form concatamers of 200-700 nucleotides. The DNA was isolated from a 2% agarose gel and cloned using the TOPO TA cloning kit (Invitrogen). Clones were submitted for sequencing to Agencourt Bioscience Corporation.

*Sequence analysis.* Sequences were analyzed using a suite of programs (uRNAs.pl) which automatically compares the cloned sequences to selected genomes and uses the Zuker folding algorithm to produce secondary structures of possible miRNA precursors [10] (L.P. Lim, N.C. Lau, E.G. Weinstein, A. Abdelhakim, C. Burge, and D.P. Bartel, in preparation). Clones were compared to the genome sequence available for zebrafish (ftp.ensembl.org/pub/traces/zebrafish and ftp.sanger.ac.uk/pub/zebrafish); *Tetraodon nigroviridis* (ftp-genome.wi.mit.edu/pub/annotation/tetraodon); *Fugu rubripes* (http://fugu.hgmp.mrc.ac.uk/Download); human, mouse, and *Drosophila melanogaster*

(www.ncbi.nlm.nih.gov:80/PMGifs/Genomes); *C. elegans* (www.wormbase.org, release WS45); and *C. briggsae* (www.ncbi.nlm.nih.gov/Traces).

## RESULTS AND DISCUSSION

In order to investigate the roles of miRNAs in vertebrate development, small RNAs (~18-28 nt) were isolated and sequenced from 8 developmental stages (zygote, 256 cell, germ ring, 75% epiboly, tailbud, 11 somite, 24 hours, and 48 hours). From 777 concatamers there were 4438 cloned sequences, of which 2770 were unique (Table 1). The unique sequences were compared to genomic databases from zebrafish, the pufferfishes *Tetraodon nigroviridis* and *Fugu rubripes*, human, mouse, *Drosophila*, *C. elegans* and *C. briggsae*. Approximately 140 miRNA candidates were selected based on their potential to form long hairpin secondary structures with flanking genomic sequences (Table 2).

As observed with miRNAs from other species, zebrafish microRNAs have an average length of 21-23 nt and can occur on either the 5′ or 3′ side of the hairpin. About half of them start with a U, and the rest start with an A or C, which correlates with the sequence specificity of Dicer [10, 17]. Many of the miRNAs also exhibit heterogeneity in their lengths, both at the 5′ and 3′ ends. Although the miRNAs precursors share a common structure overall, the size and positions of the bulges is variable (Figure 1A).

Because we were comparing the cloned RNAs to incomplete shotgun genomic sequences for zebrafish and also wanted to look for homologs in other organisms, we accepted matches of ≥90% identity in our sequence comparisons. In a few cases, we were unable to find the miRNA candidate in the zebrafish genome, but were able to

identify matches in other genomes, which have the expected secondary structure of miRNAs (for example, P1_A02-1 in Table 2). The sequences in Table 2 have perfect matches in one or more genomes, and there were many more sequences with only one or two mismatches. While a few of these may represent sequencing errors, many of them appear to represent paralogous gene families in zebrafish or possible homologs in other organisms. For instance, F1a_A05-2 and F2_B07-2 differ by only the first nucleotide, and F1a_C06-6, S1_D02-4, and T1c_C03-1 also differ by a single nucleotide. There were perfect zebrafish matches to all of these, as well as perfect hits to other genomes by at least one member of the family. It is unlikely that the family members with only perfect hits to zebrafish are sequencing errors, because they were cloned multiple times. In one unusual case, clone F1a_B05-7 had perfect matches to zebrafish, *Tetraodon*, human, and mouse, and one similar match to *Drosophila*. The *Drosophila* sequence differs at a single position, where A5 is replaced by a GG dinucleotide, but the putative miRNA is on the opposite side of the fold-back (Figure 1B). If this *Drosophila* sequence is a miRNA, this would be the first example of possible homologs having different precursors, although it is possible that such short sequences could have arisen by convergent evolution.

Overall, the miRNAs were extremely well-conserved. Only one of them had no similar match to a genome other than zebrafish. These homologs tend to have similar, but not necessarily identical structures (Figure 1B,C). This suggests that sequence divergence of the strand opposite the miRNA is constrained and that the secondary structure might guide correct cleavage by Dicer.

The results presented here indicate that miRNAs are expressed throughout zebrafish development, and that miRNA expression increases as development proceeds (data not shown). Indeed, the RT-PCR of the zygote and 256 cell stages appeared as an even, light smear corresponding to inserts of ~18-28 nt ligated to the adaptor oligonucleotides, whereas there was a fairly intense band corresponding to inserts of 21-25 nt in later stages (data not shown). Developmental expression patterns for most *C. elegans* miRNAs suggest that once miRNA expression is activated, expression is usually not turned off (L. Lim et al., in preparation). It will be interesting to determine if the same is true for zebrafish, or if the expression of some miRNAs is restricted to particular times. The presence of miRNA candidates in stages preceding the onset of embryonic transcription (zygote and 256 cell) is particularly intriguing. The processing of these miRNAs from maternally transcribed precursors could be crucial steps in events just after fertilization or just prior to the onset of embryonic transcription.

The isolation of ~150 miRNA candidates is the first step in understanding their roles in vertebrate development. The frequency of cloning certain miRNAs in each stage promises to provide an indication of their developmental expression, and this expression must be verified on Northern blots to confirm that the cloned sequences are miRNAs. In addition, these miRNAs are excellent candidates for functional studies because the phenotypes of underexpression or misexpression are easily observed in zebrafish embryos.

ACKNOWLEDGEMENTS

We are especially grateful to Elizabeth Wiellette and Hazel Sive for providing zebrafish embryos and their expertise. We also thank Nelson Lau for providing reagents and protocols for the cloning procedure and to Lee Lim for assistance with bioinformatics.

REFERENCES


1. Lee, R.C., Feinbaum, R.L. & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, pp. 843-854.
2. Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. & Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**, pp. 901-906.
3. Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E. & Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, pp. 86-89.
4. Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, pp. 363-366.
5. Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G. & Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. *Cell* **106**, pp. 23-34.
6. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T. & Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, pp. 834-838.
7. Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J. & Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev* **15**, pp. 2654-2659.
8. Wightman, B., Ha, I. & Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* **75**, pp. 855-862.
9. Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R. & Ruvkun, G. (2000). The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol Cell* **5**, pp. 659-669.
10. Lau, N.C., Lim, L.P., Weinstein, E.G. & Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294**, pp. 858-862.
11. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* **294**, pp. 853-858.
12. Lee, R.C. & Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. *Science* **294**, pp. 862-864.
13. Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. & Tuschl, T. (2002). Identification of Tissue-Specific MicroRNAs from Mouse. *Curr Biol* **12**, pp. 735-739.

14. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. & Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* **16**, pp. 720-728.
15. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. & Bartel, D.P. (2002). MicroRNAs in plants. *Genes Dev* **16**, pp. 1616-1626.
16. Kane, D.A. & Kimmel, C.B. (1993). The zebrafish midblastula transition. *Development* **119**, pp. 447-456.
17. Zamore, P.D., Tuschl, T., Sharp, P.A. & Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**, pp. 25-33.

**Table 1. Number of Sequences Cloned.**

| Stage | Concatamers | Sequences | Unique Sequences |
|---|---|---|---|
| Zygote | 67 | 263 | 183 |
| 256 cell | 85 | 293 | 270 |
| Germ ring | 87 | 423 | 187 |
| 75% epiboly | 91 | 681 | 416 |
| Tailbud | 95 | 504 | 326 |
| 11 somite | 90 | 645 | 362 |
| 24 hours | 92 | 660 | 431 |
| 48 hours | 170 | 969 | 595 |
| **Total** | **777** | **4438** | **2770** |

**Table 2. MicroRNA Candidates Cloned from Zebrafish.**

| miRNA name[a] | Clone name | Sequence[b] | Length | Fold-back Arm | Organisms[c] |
|---|---|---|---|---|---|
| | F1a_A05-2 | CACATTCATTGCTGTCGGTGGGTT | 21-25 | 5' | z+ t- f- |
| | F2_B07-2 | AACATTCATTGCTGTCGGTGGGTT | 24 | 5' | z+ h+ t+ f+ m+ |
| tom-156 | F2_D03-2 | AACATTCAACGCTGTCGGTGAGT | 22-23 | 5' | z+ f+ t+ h+ m+ |
| | F1b_F05-3 (F2_D03-2as) | ACCATCGACCGTTGACTGTACC | 21-22 | 3' | z+ f+ t- |
| | P1_C10-4 | ACCATCGACCGTTGATTGTACC | 22 | 3' | t+ f+ h+ m+ |
| | F1b_E01-2 | ACCATCGAGTGTTGAGTGTACC | 22 | 3' | z+ t- f- |
| *miR-21* | F1a_A10-5 | TAGCTTATCAGACTGGTGTTGGC | 22-23 | 5' | z+ t+ f+ m- h- |
| miR-125b (~lin-4) | F1a_A12-4 | TCCCTGAGACCCTAACTTGTGA | 21-22 | 5' | z+ t+ f+ h+ d+ |
| | F1b_E04-5 | TCCCTGAGACCCTAACTCGTGA | 22 | | z+ |
| *miR-125a (~lin-4)* | F1b_F03-4 | TCCCTGAGACCCTTAACCTGTGA | 22-23 | 5' | z+ f+ t+ h- |
| miR-20 | F1a_B03-2 | TAAAGTGCTTATAGTGCAGGTAG | 23 | 5' | z+ h+ t+ f+ m+ |
| | F1a_B11-8 | CAAAGTGCTTACAGTGCAGGTAG | 22-24 | 5' | z+ t+ f+ m+ h+ |
| *miR-91* | F1a_B05-8 | CAAAGTGCTCACAGTGCAGGTAG | 21-23 | 5' | z+ |
| *miR-93, miR-106* | P1_F10-3 | AAAAGTGCTGTTTGTGCAGGTAG | 23 | 5' | f+ t+ h- m- |
| *miR-20* | F2_F09-1 | AAAGTGCTATCAAGTTGGGGTAC | 23 | 3' | z+ |
| | P1_F09-6 | AAAGTGCTATCAAGTTGGGGTAG | 20-23 | 3' | z+ |
| | S1_B05-1 | AAAGTGCTATCAAGTTGGGGTAT | 23 | 3' | z+ |
| | T2_G01-2 | AAAAGTGCTATCAAGTTGGGGTAG | 22-24 | 3' | z+ |
| | T2_A12-2 | AAAGTGCTATCAAGCTGGGGTA | 22 | 3' | z+ |
| | G2_H06-1 | AAAGTGCTATCAAGTTGAGGTAG | 21-23 | 5' | z+ |
| | P1_A01-7 | AAAGTGCTATCAAGTTGGGGTAA | 23-24 | 3' | z+ |
| | T2_C08-4 | AAAGTGCTAACAAGTTGGGGTAG | 23 | 3' | z+ |
| | E1_C12-10 | ATAAGTGCTATTTGTTGGGGTAG | 23 | 3' | z+ |
| | F1a_D08-3 | TAAGTGCTATTTGTTGGGGTAG | 20-23 | 3' | z+ f+ t- d- |
| | F2_A03-4 | TAAGTGCTATTTGTTGGGGTAT | 22-23 | 3' | z+ |
| | S1_G05-9 | TAAGTGCTATTTGTTGGGGTAC | 22-23 | 3' | z+ f- |
| | G1_C02-3 | TAAGTGCTATTTGTTGAGGTAG | 22 | 3' | z+ |
| | S1_H01-3 | TAAGTGCTATTTGTTGGGGTAGA | 23 | 3' | z+ t- |
| | T2_F07-5 | TAAGTGCCTATTTGTTGGGGTNG | 23 | 3' | z- |
| | E1_D04-9 | TAAGTGCTATTTGTCGGGGTAG | 22 | 3' | z+ |
| | T2_C05-4 | TTAAGTGCTATTTGTTCGGNGTAG | 24 | 3' | z- |
| | G1_B05-5 | TAAGTGCTATCTGTTGGGGTAG | 22 | 3' | z+ |
| | G1_E01-1 | TAAGTGCTATTTGTTGGGGTAA | 22 | 3' | z+ |
| | E2_A07-1 | TAAAGTGCTATTTGTTGGGGTAG | 23 | 3' | z- f- t- |
| | P1_B02-4 | TAAGTGCTTCTCTTTGGGGTAG | 20-23 | 3' | z+ f- t- |
| | E1_B11-6 | TAAGTGCTTCTCTTTGGGGT | 20 | 3' | z+ t- |
| | S1_A02-2 | TAAGTGCTTCTCTTTGGGGTAC | 22 | 3' | z+ |
| | T2_A12-5 | TAAGTGCTTCTGTTTGGGGTAG | 22 | 3' | z+ t- |
| | G1_H11-6 | CAAGTGCTATTTGTTGGGGTAG | 22 | 3' | t- |
| | G1_B05-2 | CAAGTGCTTCTCTTTGGGGTAG | 22 | 3' | z+ |

[a] MicroRNAs listed in regular font are identical to the sequence cloned from zenrafish, and those listed in italics have at least one nucleotide difference from the zebrafish sequence.
[b] For sequences that were cloned multiple times, the most frequently cloned sequence is listed, and sequences are grouped into families based on similarity.
[c] Genomes are abbreviated as zebrafish (z), *Tetraodon* (t), *Fugu* (f), human (h), mouse (m), *Drosophila* (d), *C. elegans* (e), and *C. briggsae* (b). Genomes marked with "+" indicate a perfect match; genomes marked with a "-" indicate a hit with ≥90% identity.

**Table 2. MicroRNA Candidates Cloned from Zebrafish (continued).**

| miRNA name[a] | Clone name | Sequence[b] | Length | Fold-back Arm | Organisms[c] |
|---|---|---|---|---|---|
| *miR-16* | F1a_B04-1 | TAGCAGCACGTAAATATTGGAG | 21-22 | 5' | z+ t+ f+ m- h- e- |
| | S1_C08-6 | TAGCAGCACGTAAATATTGAT | 21 | 5' | f- t- |
| | F1a_B05-7 | ACAGCAGGCACAGACAGGCAG | 21-22 | 3' | z+ t+ f+ h+ m+ d- |
| | F1a_B07-1 | CCCAGTGTTCAGACTACCTGTTC | 22-23 | 5' | z+ t+ f+ h+ m+ |
| | F1a_C06-6 | TAGGTAGTTTCATGTTGTTGGG | 22 | 5' | z+ t+ f+ h+ m+ |
| | S1_D02-4 | TAGGTAGTTTCAAGTTGTTGGG | 22-23 | 5' | z+ t+ f+ |
| | T1c_C03-1 | TAGGTAGTTTGATGTTGTTGGG | 22 | 5' | z+ |
| *miR-1* | F1a_C07-2 | TGGAATGTAAGGAAGTGTGTGG | 21-23 | 3' | z+ t+ h+ |
| miR-19b | F2_F02-4 | TGTGCAAATCCATGCAAAACTGA | 23 | 3' | z+ t+ f+ h+ m+ |
| | F2_A07-7 | TGTGCAAATCCATGCAAAACTC | 22 | 3' | z+ |
| miR-19a | S1_F03-1 | TTGTGCAAATCTATGCAAAACTG | 23 | 3' | z+ f+ t+ h+ m+ |
| tom-160 | F1a_D08-5 | TGGACGGAGAACTGATAAGGG | 21-22 | 3' | z+ t+ f+ h+ m+ d+ |
| | F1b_F08-7 | CCGGGAGTGGGATGTTTGCGCT | 22 | 3' | z+ |
| | S1_A12-3 | CTGGGCGGAGGGTGTTTGCTGT | 22 | 3' | z+ |
| miR-92 | P1_D05-5 | TATTGCACTTGTCCCGGCCT | 20-23 | 3' | z+ f+ t+ h+ m+ d- |
| | F1a_A10-9 | TATTGCACTCGTCCCGGCCTCC | 21-22 | 3' | m+ |
| miR-25 | F2_D07-2 | CATTGCACTTGTCTCGGTCTGA | 22 | 3' | z+ f+ t+ h+ m+ |
| tom-164 | F1b_H09-7 | TAACACTGTCTGGTAACGATGTT | 23 | 3' | z+ f+ t+ h+ m+ |
| | P1_G04-9 | TAACACTGTCTGGTAACGATGCA | 23 | 3' | z+ |
| | F1c_E09-4 | TAATACTGTCTGGTAATGCCGT | 22 | 3' | z+ t+ f+ m+ |
| | F2_D11-1 | TAATACTGCCTGGTAATGATGA | 21-22 | 3' | z+ f+ t+ h+ m+ |
| | F2_C06-3 | TAATACTGCCTGGTAATGATGCA | 21-23 | 3' | z+ t- m- |
| *miR-100, miR-99a* | F2_D04-1 | AACCCGTAGATCCGATCTTGTG | 21-22 | 5' | z+ t- f- m- h+ d- |
| | F2_D02-1 | TACCCTGTAGAACCGAATTTGT | 22-23 | 5' | z+ f+ t+ h+ m+ |
| | P1_E12-8 | TACCCTGTAGAACCGAATGTGT | 22-23 | 5' | z+ f+ t+ |
| | S1_F01-3 | TACCCTGTAGATCCGAATTTGTG | 21-23 | 5' | z+ f+ t+ h+ m+ d- |
| *miR-57* | P1_E12-2 | TACCCTGTAGATCCGGATTTGTG | 21-23 | 5' | z+ f+ t+ |
| *miR-18* | F1b_H09-8 | TAAGGTGCATCTAGTGCAGATAG | 23 | 5' | z+ f+ t+ h+ m+ |
| | F2_B04-1 | TAAGGTGCATCTTGTGTAGTTAG | 22-23 | 5' | z+ f- t- |
| | S1_F03-3 (F1b_H09-8as) | ACTGCCCTAAGTGCTCCTTCTGG | 23 | 3' | z+ f+ t+ h+ m+ |
| miR-30c | F1c_G01-5 | TGTAAACATCCTACACTCTCAGCT | 24 | 5' | z+ f+ t+ h+ m+ |
| miR-30b | G1_A10-7 | TGTAAACATCCTACACTCAGCT | 22 | 5' | z+ t+ h+ m+ |
| miR-30d | S1_B08-4 | TGTAAACATCCCCGACTGGAAGCT | 20-24 | 5' | z+ f+ t+ h+ m+ |
| | Z2b_H07-3 | TCCTCAGGTTGAGTGCGCGATGGC | 24 | 5' | z+ |
| | F2_A10-3 | TGGAAGTGAATGGGTACAGGTT | 22 | 3' | z+ |

**Table 2. MicroRNA Candidates Cloned from Zebrafish (continued).**

| miRNA name[a] | Clone name | Sequence[b] | Length | Fold-back Arm | Organisms[c] |
|---|---|---|---|---|---|
| *miR-122a* | F2_B01-4 | ATGGAGTGTCAATGGTTACAGGT | 23 | 3' | z+ |
| | F2_B04-5 | AGCTACATTGTCTGCTGGGTTTC | 23 | 3' | z+ f+ t+ h+ m+ |
| | P1_E05-9 | AGCTACATCTGGCTACTGGGTCTC | 23-24 | 3' | z+ f+ t+ h+ m+ |
| miR-108as | F2_C08-4 | TAATGCCCCTAAAAATCCTTA | 21 | 3' | z+ f+ t+ h+ m+ |
| miR-143 | F2_C08-6 | TGAGATGAAGCACTGTAGCTC | 21 | 3' | z+ f+ h+ m+ |
| | F2_A11-4 | ACTCTAACTTTAGCATCTTTCT | 22 | 5' | z+ |
| | G1_G07-4 | TCTCTAACTTTAGCATCTTTCT | 22 | 5' | z+ |
| | F2_E03-3 | TATAGGGATGGAAGCCATGCA | 21 | 3' | z+ f+ t+ |
| let-7c | F2_F03-5 | TGAGGTAGTAGGTTGTATGGTT | 22 | 5' | z+ f- t- m+ h+ d- e- b- |
| miR-128 | F2_F05-1 | TCACAGTGAACCGGTCTCTTT | 21 | 3' | z+ f+ t+ h+ m+ |
| | F2_H01-7 | TCCTTCATTCCACCGGAGTCTG | 22 | 5' | z+ f+ t+ h+ m+ |
| miR-138 | F2_B01-3 | AGCTGGTGTTGTGAATCAGGCCG | 23 | 5' | z+ f+ t+ h+ m+ |
| | F2_B08-1 | GCTATTTCACAACACCAGGGT | 21 | 3' | z+ f+ t+ m- |
| | F2_H01-4 | GTGAAATGTTTAGGACCACTTG | 22-23 | 3' | z+ f+ t+ h- |
| | S1_A07-12 | GTGAAATGTTCAGGACCACTTG | 22 | 3' | z+ |
| miR-133 | F2_C05-3 | TGGTCCCCTTCAACCAGCTGT | 21 | 3' | z+ f+ t+ h+ m+ d+ |
| | F2_C11-1 | TTTGGTCCCCTTCAACCAGCT | 21 | 3' | z+ t+ h+ m+ |
| | F2_H06-1 | CTGTGCGTGTGACAGCGGCTA | 21 | 3' | z+ f+ t+ h- m- d- |
| | F2_B07-5 | TTCCCTTTGTCATCCTATGCCT | 22 | 5' | f+ t+ m+ h+ |
| tom-170 | P1_F10-6 | ACCACAGGGTAGAACCACGGAC | 20-23 | 3' | z+ f+ t+ h+ |
| miR-140 | P1_E12-5 (P1_F10-6as) | CAGTGGTTTTACCCTATGGTA | 21 | 5' | z+ f+ t+ h+ |
| miR-124a | P1_G04-3 | TAAGGCACGCGGTGAATGCCAA | 21-23 | 3' | z+ f+ t+ h+ m+ d+ e- |
| *miR-34* | P1_G03-2 | TGGCAGTGTCTTAGCTGGTTGT | 22-23 | 5' | z- f+ t+ h+ m+ d- |
| | P1_D02-1 | CGCGTGCGTCGACCGGAACGGA | 22 | 5' | z+ |
| miR-139 | P1_B01-6 | TCTACAGTGCATGTGTCTCCAGT | 23 | 5' | f+ t+ h- m- |
| | P1_A02-1 | CACAGCGCCTGCAATGTGGAGG | 22 | 3' | f+ t+ |
| miR-126 | P2_C11-4 | TCGTACCGTGAGTAATAATGCA | 22 | 3' | z+ f+ t+ h- m- |
| | P1_G05-4 | AGAAAACCGGTAACCATTGACTT | 23 | 5' (one 3') | z+ |
| | S1_E02-2 | ACCGGTAACCATTGACTTCTA | 21 | 5' (one 3') | z+ |

**Table 2. MicroRNA Candidates Cloned from Zebrafish (continued).**

| miRNA name[a] | Clone name | Sequence[b] | Length | Fold-back Arm | Organisms[c] |
|---|---|---|---|---|---|
| *miR-30* | P1_B07-7 | CAGTGCAATATTAAAAGGGCAT | 22 | 3' | z+ f+ h- m- |
| | P1_E03-3 | CAGTGCAATAATGAAAGGGCAT | 22 | 3' | z+ f+ t- h- m- |
| *miR-130* | F2_A05-2 | TAGTGCAATATTGCTTATAGGGTC | 24 | 3' | z+ f+ |
| | P1_H07-2 | ACTCTTTCCCTGTTGCACTAC | 21 | 5' | z+ f- t- h+ m+ |
| | P2_B10-2 | TGGGGAGCCTGAGCTCGGTGGAT | 23 | 3' | z+ |
| | Z2a_E05-4 | TGGGAATACCAGGTGCTGTAAGCTA | 23-25 | 3' | z+ f? m? |
| | Z2a_A06-2 | TGGGAATACCAGGTGCTGCAAGCTT | 25 | 3' | z+ t+ |
| | Z2a_C10-1 | TGGGAATACCAGGTGCTGCAAGCTA | 25 | 3' | z+ |
| | Z2a_D11-3 | GGAATACCAGGTGCTGTAGGCTA | 23 | 3' | z- |
| | P2_E02-2 | CCATCCGGGTGCGGCGTGGGC | 21 | 5' | z+ f- |
| | F1b_F09-3 | CCCCGACTCTTCCCCTCCGCCC | 20-28 | 5' | z+ |
| | T1c_E04-1 | CCCCCCGACTCTTCCCCTCC | 20 | 5' | z+ |
| | S1_A04-6 | GCCCGCCTCTTGCTCGCGTCCGTCCGG | 22-28 | 5' | z+ |
| | S1_C02-4 | CGCCTCTTGCTCGCGTCCGTCC | 22 | 5' | z+ f- |
| *miR-15a* | S1_C06-5 | TAGCAGCACATCATGGTTTGTA | 21-22 | 5' | z+ f- t- m- |
| miR-142as | T1c_A06-1 | TGTAGTGTTTCCTACTTTATGG | 22 | 3' | z+ f+ t+ h+ |
| | T2_C03-2 | ACCCTCACAAAAGCACTGACT | 20-21 | 5' | z+ |
| | T2_D07-1 | ACCCTCACAAAGGCACTGACT | 21 | 5' | z+ |
| | E1_A12-1 | CTGGGTGGTATGCTTCTTTGGG | 22 | 5' | z+ |
| | E1_C12-7 | TAGCTGGTAATTTGTAGATGGTA | 23 | 3' | z+ |
| | E1_F02-5 | TACGTTGGGCCGACGTCGTGCCG | 23 | 5' | z+ |
| | B1_G01-3 | AAGGTCCAACCTCACATGTCCT | 22 | 5' | z+ f+ t+ |
| | F1a_A10-10 | AGCCACTGACTAACGCACATTG | 22 | 5' | z+ f+ t+ |
| miR-23a | B1_D12-4 | TCACATTGCCAGGGATTT | 18 | 3' | f+ |
| | S1_C10-2 | CGGGGCCGTGGCACTGTATGAGA | 23 | 5' | z+ m- |
| miR-9, miR-131as | F2_E05-2 | TCTTTGGTTATCTAGCTGTATGT | 23 | 5' | z- f- t- h- m- d- |
| | P1_E03-2 | TACTGCATCAGGAACTGATTGGAT | 24 | 5' | f- t- h+ m- |
| miR-29a | T1c_A06-3 | TAGCACCATCTGAAATCGGTTT | 22 | 3' | f- t- h- m- |
| *miR-22* | B1_G04-4 | AAGCTGCCAGCTGAAGAACTCC | 22 | 3' | z- f- |

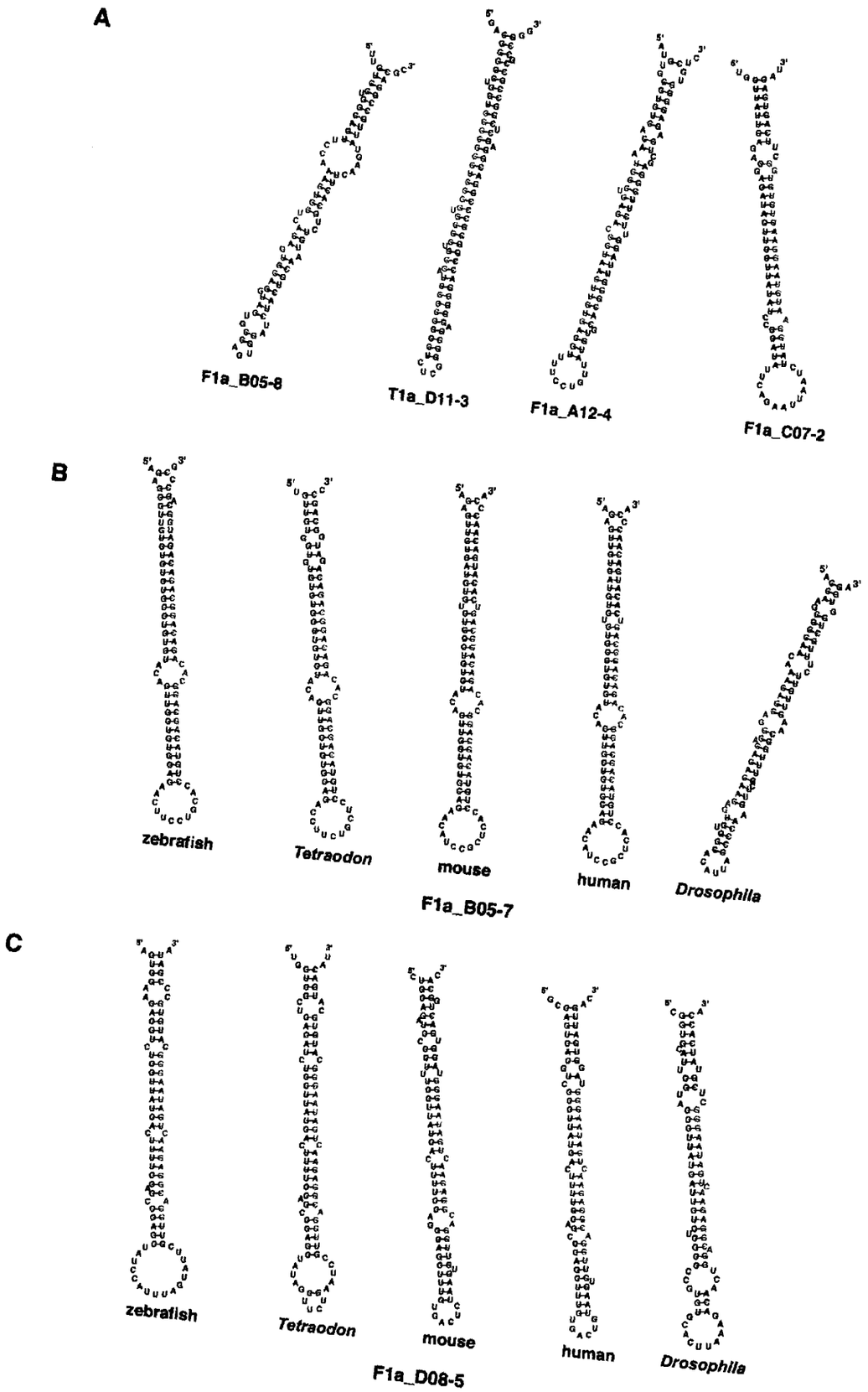
**Table 2. MicroRNA Candidates Cloned from Zebrafish (continued).**

| miRNA name[a] | Clone name | Sequence[b] | Length | Fold-back Arm | Organisms[c] |
|---|---|---|---|---|---|
| | F1b_G08-4 | CCACATCGCACAGCTACTCATG | 22 | 3' | z- |
| | F1a_D04-6 | TGNGAATGACGTATTTGCCGCTGCC | 25 | 3' | z- |
| | F1b_G06-8 | CAGGGGGTCGGGAAGCACTGT | 21 | 3' | z- |
| | F1b_H11-7 | CCTCCAATATTGCTCGTGCTGC | 22 | 3' | f- |
| | F2_F02-6 | AACAATATCCTGGCGCTGCCTGAGT | 25 | 5' | f- t- |
| | T1a_D12-4 | CGGGGGTGCGCGCGCGGTGGC | 21 | 5' | f- t- |
| | E1_D11-2 | TCGACTCCCGGTATGGGAACCA | 22 | 5' | f- |
| | G1_G04-2 | TATGGAAGTCACTAGCTCCTGCA | 23 | 3' | z- |
| | F1c_E03-2 | CGGGTCCCTTCGTGGTCGCCA | 21 | 5' | t- |

FIGURE LEGEND

Figure 1. Secondary structures of possible miRNA precursors. A) Four typical miRNAs from zebrafish. B) F1a_B05-7, a well-conserved zebrafish miRNA with a possible homolog in *Drosophila*. C) F1a_D08-5, another well-conserved miRNA.

# Figure 1


A
F1a_B05-8
T1a_D11-3
F1a_A12-4
F1a_C07-2
B
zebrafish
Tetraodon
mouse
human
Drosophila
F1a_B05-7
C
zebrafish
Tetraodon
mouse
human
Drosophila
F1a_D08-5

# Appendix B

# RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension

# RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension

**Wendy K. Johnston, Peter J. Unrau,* Michael S. Lawrence, Margaret E. Glasner, David P. Bartel†**

**The RNA world hypothesis regarding the early evolution of life relies on the premise that some RNA sequences can catalyze RNA replication. In support of this conjecture, we describe here an RNA molecule that catalyzes the type of polymerization needed for RNA replication. The ribozyme uses nucleoside triphosphates and the coding information of an RNA template to extend an RNA primer by the successive addition of up to 14 nucleotides—more than a complete turn of an RNA helix. Its polymerization activity is general in terms of the sequence and the length of the primer and template RNAs, provided that the 3′ terminus of the primer pairs with the template. Its polymerization is also quite accurate: when primers extended by 11 nucleotides were cloned and sequenced, 1088 of 1100 sequenced nucleotides matched the template.**

The RNA world hypothesis states that early life forms lacked protein enzymes and depended instead on enzymes composed of RNA (*1*). Much of the appeal of this hypothesis comes from the realization that ribozymes would have been far easier to duplicate than proteinaceous enzymes (*2*–*5*). Whereas coded protein replication requires numerous macromolecular components [including mRNA, transfer RNAs (tRNAs), aminoacyl-tRNA synthetases, and the ribosome], replication of a ribozyme requires only a single macromolecular activity: an RNA-dependent RNA polymerase that synthesizes first a complement, and then a copy of the ribozyme. If this RNA polymerase were itself a ribozyme, then a simple ensemble of molecules might be capable of self-replication and eventually, in the course of evolution, give rise to the protein-nucleic acid world of contemporary biology. Finding a ribozyme that can efficiently catalyze general RNA polymerization would support the idea of the RNA world (*1*, *6*, *7*) and would provide a key component for the laboratory synthesis of minimal life forms based on RNA (*8*, *9*).

Although progress has been made in finding ribozymes capable of template-directed RNA synthesis, none of these ribozymes has the sophisticated substrate-binding properties needed for general polymerization (*7*). Derivatives of self-splicing introns are able to join oligonucleotides assembled on a template (*10*–*12*). However, the templates that can be copied are limited to those that match the oligonucleotide substrates, and it has not been possible to include sufficient concentrations of all the oligonucleotide substrates needed for a general copying reaction. More recently, efforts have shifted to derivatives of an RNA-ligase ribozyme that was isolated from a large pool of random RNA sequences (*13*–*15*). Some derivatives are capable of template-directed primer extension using nucleoside triphosphate (NTP) substrates, but their reaction is also limited to a small subset of possible template RNAs (*15*). These ligase derivatives recognize the primer-template complex by hybridizing to a particular unpaired segment of the template (Fig. 1A). In using a short segment of a special template to direct primer extension, these ribozymes resemble telomerases more than they resemble the enzymes that replicate RNA and DNA by means of general polymerization.

**Polymerase Isolation.** We have used the catalytic core of the ligase ribozyme (*14*, *16*) as a starting point for the generation of a ribozyme with general RNA polymerization activity. The new polymerase ribozyme was isolated from a pool of over $10^{15}$ different RNA sequences. Sequences in the starting pool contained a mutagenized version of the parental ligase (Fig. 1B). To sample a broad distribution of mutagenesis levels, the starting pool comprised four subpools in which the core residues of the ligase domain were mutagenized at levels averaging either 0, 3, 10, or 20% (*17*). Two loops within the ligase domain, both unimportant for ligase function, were replaced with 8-nucleotide (nt) random-sequence segments (Fig. 1B). The 5′ terminus of the ligase domain was covalently attached to an RNA primer so that molecules able to catalyze primer extension could be selected by virtue of their attachment to the primer that they extended.

In contrast to the parental ribozyme, which hybridizes to the template, a ribozyme capable of general polymerization must recognize the primer-template complex without relying on sequence-specific interactions. Therefore, the template RNA was designed to be too short for extensive hybridization with the ribozyme (Fig. 1B). For the parental ribozyme (Fig. 1A), the pairing between the ribozyme and the template also comprised a stem known to be necessary for ligase function (*16*). To restore this stem, a short RNA, GGCACCA (purple RNA in Fig. 1B), was introduced to hybridize to the segment of the ligase domain that formerly paired with the template. Finally, because a more general mode of primer-template recognition might require the participation of an additional RNA domain, a 76-nt random-sequence segment was appended to the 3′ terminus of the degenerate ligase domain (Fig. 1B).

Sequence variants able to recognize the primer-template in this new configuration and then extend the primer with tagged nucleotides were enriched by repeated rounds of in vitro selection and amplification (Table 1). RNAs that extended their primer by using 4-thioUTP were isolated on APM gels (urea denaturing gels cast with a small amount of *N*-acryloylaminophenylmercuric acetate, which impedes migration of RNA containing 4-thioU) (*18*). To favor variants that recognize generic rather than sequence-specific features of the primer-template, different primer-template sequences and lengths were used in different rounds of selection (Table 1). To favor the more efficient ribozyme variants, the stringency of the selection was increased in later rounds by requiring addition of two tagged nucleotides, such as biotinylated A and 4-thioU (*19*), and by decreasing the time of incubation with the tagged NTPs (Table 1).

After eight rounds of selection and amplification, desirable variants had increased in abundance to the point that a detectable fraction of the pool molecules could extend their primer by using both 4-thioUTP and radiolabeled ATP in a template-dependent fashion (Fig. 2). Other variants able to tag themselves were detected as early as round four, but most of these ribozymes added tagged nucleotides in the absence of the template oligonucleotide, or they decorated themselves at sites other than the 3′ terminus of the primer (*20*). Seventy-four variants from rounds 8 through 10 were cloned and found to represent 23 sequence families, each family having descended from a different ancestral sequence of the starting pool. Ribozymes from two families extended their

Whitehead Institute for Biomedical Research, and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.

*Present address: Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada.

†To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu

primer by using both 4-thioUTP and radiolabeled ATP in a template-dependent fashion. These two families are represented by isolates 9.1 and 10.2 (Fig. 2). Isolate 10.2 (Fig. 1C), from round 10, was from the more prevalent and active of these two families and was chosen for further study.

**Polymerization with multiple turnover.** Having isolated a ribozyme that did not rely on forming base pairs with the template RNA during primer extension, we next determined whether it instead recognized the particular sequence used to link the primer to the ribozyme. We were pleased to discover that the round-10 ribozyme did not require this sequence. In fact, it did not require any covalent attachment to the primer. When incubated with a 10-fold excess of both a 6-nt primer and a 9-nt template, as well as the appropriate nucleoside triphosphate, the round 10 polymerase fully extended the primer, with multiple turnover (Fig. 3). The primer and template sequences in this experiment were designed to differ from those used most frequently during the selection (Table 1, aligning the sequences relative to the primer 3′ termini). Despite the complete change in the sequence of the primer-template complex, the polymerase isolate was able to recognize the complex and extend the primer (*21*). This indicates that the ribozyme binds the primer-template without relying upon recognition of a particular sequence.

The new mode of primer-template recognition appears to be conferred by the accessory domain derived from the 76-nt random-sequence segment and the 3′-terminal segment that binds the primer used for the reverse transcription–polymerase chain reaction (RT-PCR). Without the accessory domain, the ligase domain of the round-10 ribozyme, like the parental ligase itself (Fig. 3B), displayed no activity in polymerization assays requiring general primer-template recognition. Indeed, deleting only 9 nt from the 3′ terminus of the round-10 ribozyme severely diminished activity (*20*). It is interesting that the core ligase residues emerged unchanged in this round-10 isolate (compare Fig. 1, B and C), and the GGCACCA oligonucleotide designed to complete the ligase domain proved to be necessary for polymerase function (*20*). This suggests that the parental ligase did not need to adapt in order to accommodate the more general primer-template recognition afforded by the accessory domain.

The round-10 ribozyme was tested with numerous other primer-template pairs. In all cases

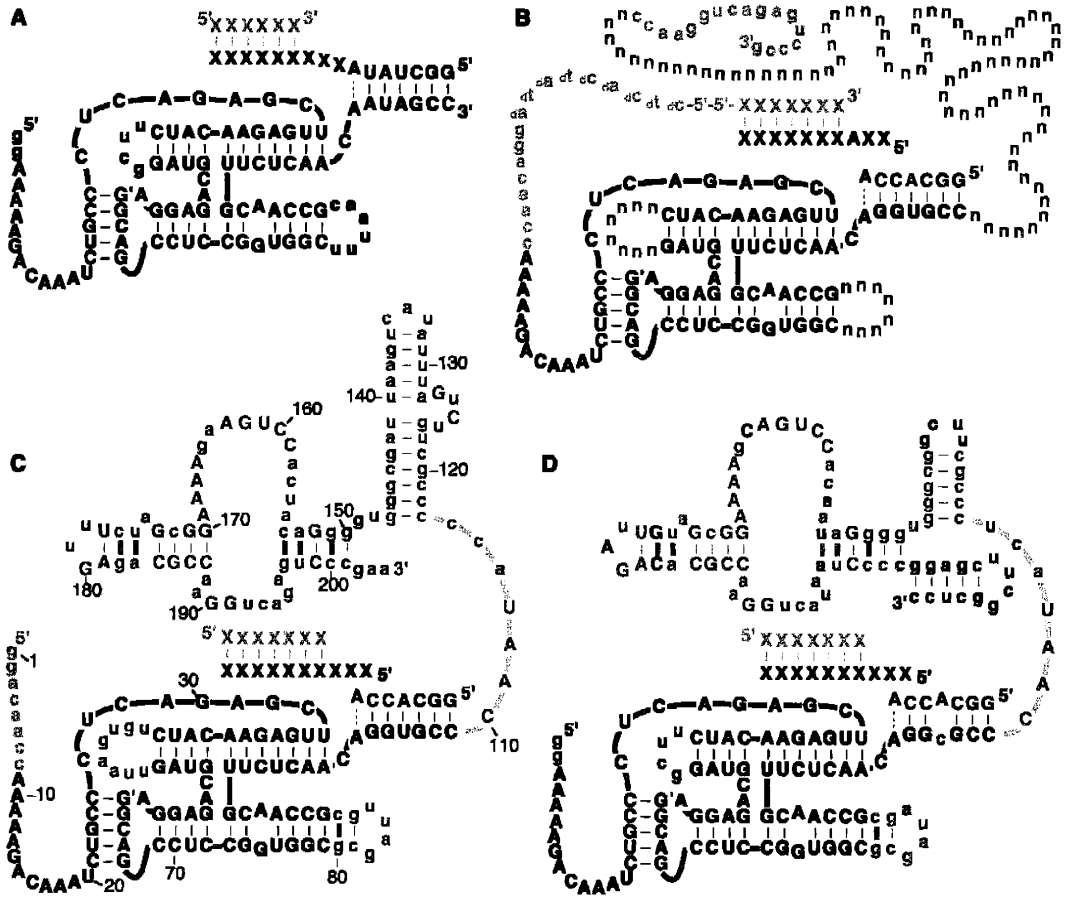



**Fig. 1.** Secondary structure models of the ribozymes. Short dashes indicate base pairs. (**A**) A ribozyme (black strand) able to promote limited RNA polymerization (*15*). It extends an RNA primer (orange strand) by using nucleoside triphosphates and coding information from an appropriate RNA template (red strand). The ribozyme can accommodate any of the four RNA nucleotides at residues indicated by an X, provided that the primer pairs with the template. However, the 5′ portion of the template must pair with the ribozyme. The depicted ribozyme was derived from an RNA ligase ribozyme; black uppercase residues and defined residues of the template comprise the core of the ligase ribozyme. (**B**) A pool of RNA sequences based on the ligase ribozyme (*17*). Colors differentiate residues representing the ligase core (black, purple), random sequence (blue), primer (orange), template (red), and RT-PCR primer-binding sites (green). Residues prefixed by "d" are DNA; all others are RNA. The 5′ end of the RNA primer is covalently joined to the 5′ end of each pool molecule via a phosphodiester linkage (-5′-5′-) (*38*). The sequence of the primer-template (X) in a given round usually differed from that of the previous round (Table 1). (**C**) The round-10 ribozyme (isolate 10.2). Residues derived from the random-sequence segments or the 3′ RT-PCR primer-binding site of the starting pool are colored blue; other drawing conventions are as in (B). Comparative sequence analysis of improved isolates from rounds 14 and 18 (*23*) supports the importance, as well as the proposed secondary structure, of the accessory domain (residues 110 to 204), particularly within the 3′ region of this domain (residues 150 to 201). Blue uppercase residues were invariant among all 22 improved isolates. Because the chance conservation of a residue not important for activity is low ($P = 0.0074$ for conservation in 22 of 22 isolates), nearly all 29 of these residues must be important for ribozyme function. Thick blue dashes mark covarying pairs, five of which (G151:C200, A153:U198, C154:G197, U175:A183, and C176:G182) support the proposed pairing within the 3′ region of the accessory domain. (**D**) The round-18 ribozyme, a shortened derivative of an improved isolate from round 18. Nucleotide changes from the round-10 isolate that arose from combinatorial mutagenesis are in pink; changes engineered when reducing the ribozyme's size are in gray (*23*). The four changes consistently found among the improved round-14 and round-18 isolates are in uppercase pink. Other drawing conventions are as in (C).

it was able to recognize the primer-template complex and to extend the primer by the Watson-Crick match to the template. Extension across from C or G template residues was usually more efficient and accurate than extension across from A or U. Extension was also much more efficient when the unpaired portion of the template was shorter than 5 nt (Fig. 4B).

**Sequence optimization of the polymerase.** To find improved polymerase variants, the in vitro selection procedure was continued for another eight rounds of selection and amplification, starting with a newly synthesized pool of variants based on the round-10 isolate (*22*). In this pool, the accessory domain, as well as the two 8-nt segments at the loops within the ligase domain, were mutagenized at an average level of 20%. Because this mutagenesis included the former RT-PCR primer-binding sequence, a new RT-PCR primer-binding sequence was appended to each pool molecule. In half the pool molecules, most of the ligase core residues were mutagenized at an average level of 3%, whereas in the other half, they were not intentionally mutagenized.

The additional rounds of selection-amplification were performed with three noteworthy modifications to the protocol of the first 10 rounds (Table 1). First, longer template RNAs were used to favor those variants better able to recognize primer-template complexes with long, unpaired template segments. Second, selection was based on the ability to add two tagged U's rather than one tagged U and one tagged A. This change was implemented after learning that the round-10 ribozyme uses biotinylated ATP much less efficiently than unmodified ATP. Third, high concentrations of unmodified ATP, CTP, and GTP were included to disfavor those variants prone to incorporating these competitor nucleotides instead of the tagged Watson-Crick match to the template.

Isolates from rounds 14 to 18 were screened for the ability to fully extend a 10-nt primer (GAAUCAAGGG) on an 18-nt template (3′-CUUAGUUCCCGCCCGGCC, underline indicates segment that pairs with the primer). Most isolates from round 18 had disrupted ligase domains and showed no sign of polymerase activity when assayed individually. They presumably were selected because of a parasitic ability to deliver their primer to the active site of a different molecule. Other isolates had polymerase activity and were much more active than the round-10 parental ribozyme. Comparative sequence analysis of the 22 most active isolates (*23*) identified conserved residues and structural features, which clustered in the 3′-terminal half of the accessory domain (Fig. 1C) and are likely to be critical for its function. This analysis also suggested a model for the secondary structure of the accessory domain (Fig. 1C) and identified four residues in the domain that consistently differed from the round-10 ribozyme (*23*). These mutations likely conferred increased polymerase activity.

One isolate from round 18 was particularly adept at using long templates. To investigate features of the ribozyme needed for activity, derivatives of this isolate were constructed and tested (*23*). A 189-nt derivative (Fig. 1D) that retains the polymerization activity of the full-length round-18 isolate has been most extensively characterized. This derivative (hereafter referred to as the round-18 ribozyme) has all the features of the accessory domain that were conserved among the 25 most active isolates, including the four mutations thought to confer improved activity (Fig. 1D). Additionally, it has

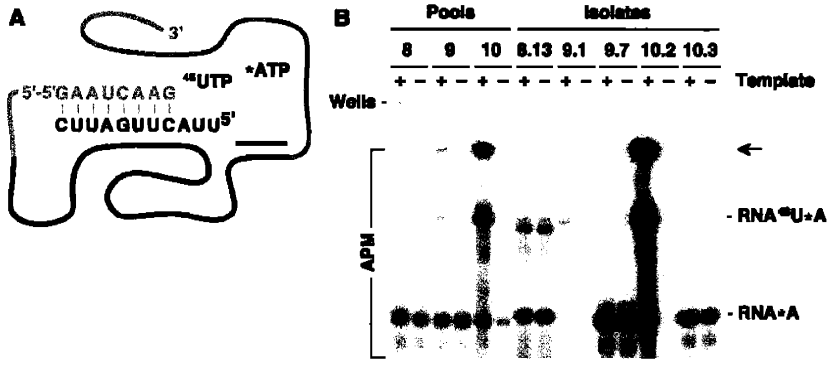



**Fig. 2.** Detection of ribozymes able to extend an attached primer by two nucleotides in a template-dependent manner. (**A**) Schematic of the RNAs in this experiment. Ribozymes attached to an RNA primer (orange) were incubated with 1 mM 4-thioUTP ($^{4S}$UTP) and trace [$\alpha$-$^{32}$P]ATP (*ATP), in the presence or absence of an RNA template (red) that codes for the addition of U and A. (**B**) Activities of ribozyme pools and isolates after 8 to 10 rounds of selection. Extension reactions were for 12 hours, under the conditions used during the rounds of selection (*18*). Shown is a PhosphorImager scan of an APM denaturing gel separating RNAs extended with radiolabeled A (RNA*A) from those extended with both 4-thioU and radiolabeled A (RNA$^{4S}$U*A). The arrow points to RNA$^{4S}$U*A extended by a second 4-thioU, which did not migrate into the APM portion of the gel. Note that addition of the second $^{4S}$U could not have been directed by an A in the template because only one of the template coding residues is an A; some misincorporation of $^{4S}$U was expected in this experiment because of the very large excess of $^{4S}$UTP over ATP. The sequence families represented by 9.1 and 10.2 add both 4-thioU and radiolabeled A in a template-dependent manner. The round-10 isolate (10.2) was chosen for further analysis and is shown in Fig. 1C.

**Table 1.** Parameters and substrates for in vitro selection. For each round of selection, pool RNA with covalently attached primer (*38*) was incubated with the indicated template RNA and NTPs for the indicated time (*18*). Nucleotide analogs 4-thioU, biotin-N6-A, and 2-aminopurine (*39*) are abbreviated $^{4S}$U, $^{B}$A, and $^{2N}$P, respectively. The primer attached to the pool molecules was complementary to the underlined segment of the template. Variants with polymerase activity were selected based on their primer being extended with the tagged nucleotides indicated in the Selection criteria column (*18*). In late rounds, 2 mM ATP, 2 mM CTP, and 2 mM GTP were included as competitor NTPs (Comp. NTPs). Pool mutagenesis was either during chemical synthesis (Synthesis) or during error-prone amplification (PCR) of the template DNA (*17*, *22*, *40*).

| Round | Mutagenesis | Template RNA | NTPs | Time (hour) | Selection criteria |
|---|---|---|---|---|---|
| 1 | Synthesis | 3′-GGUCAGAUU | $^{4S}$UTP (2 mM) | 36 | $^{4S}$U |
| 2 | None | 3′-GGUCAGAACC | $^{4S}$UTP (2 mM) | 20 | $^{4S}$U |
| 3 | None | 3′-GGUCAGAA | $^{4S}$UTP (2 mM) | 20 | $^{4S}$U |
| 4 | None | 3′-CUUAGUUCAUU | $^{4S}$UTP (2 mM) | 19 | $^{4S}$U |
| 5 | None | 3′-CUUAGUUCAUU | $^{4S}$UTP (2 mM) | 1 | $^{4S}$U |
| 6 | None | 3′-GGUCAGAUU | $^{4S}$UTP, $^{B}$ATP (1 mM each) | 14 | $^{B}$A, $^{4S}$U |
| 7 | None | 3′-CUUAGUUCAUU | $^{4S}$UTP, $^{B}$ATP (1 mM each) | 17 | $^{B}$A, $^{4S}$U |
| 8 | None | 3′-GGUCAGAUU | $^{4S}$UTP, $^{B}$ATP (1 mM each) | 17 | $^{B}$A, $^{4S}$U |
| 9 | None | 3′-GGUCAGAUU | $^{4S}$UTP, $^{B}$ATP (1 mM each) | 4 | $^{B}$A, $^{4S}$U |
| 10 | None | 3′-CUUAGUUCAUU | $^{4S}$UTP (1 mM) | 20 | $^{4S}$U |
| 11 | Synthesis | 3′-UCGACGGAACC | $^{4S}$UTP (1 mM) | 4 | 2 $^{4S}$U |
| 12 | None | 3′-ACCUGAGAAGG | $^{4S}$UTP (1 mM) | 4 | 2 $^{4S}$U |
| 13 | None | 3′-CAAGUCCAACC | $^{4S}$UTP (1 mM) | 0.2 | 2 $^{4S}$U |
| 14 | None | 3′-UCGACGGAACC | $^{4S}$UTP (1 mM) | 0.2 | 2 $^{4S}$U |
| 15 | PCR | 3′-UCGACGG$^{2N}$p$^{2N}$PCCUGCGUC | $^{4S}$UTP (0.1 mM), Comp. NTPs | 20 | 2 $^{4S}$U |
| 16 | PCR | 3′-CAAGUCC$^{2N}$p$^{2N}$PUGAUCGUA | $^{4S}$UTP (0.1 mM), Comp. NTPs | 4 | 2 $^{4S}$U |
| 17 | PCR | 3′-ACCUGAG$^{2N}$p$^{2N}$PGUGUAUGU | $^{4S}$UTP (0.1 mM), Comp. NTPs | 2 | 2 $^{4S}$U |
| 18 | None | 3′-UCGACGG$^{2N}$p$^{2N}$PCCUGCGUC | $^{4S}$UTP (0.1 mM), Comp. NTPs | 0.1 | 2 $^{4S}$U |

a U-to-C mutation within the ligase domain in the segment designed to pair with the 7-nt RNA that completes the ligase domain (Fig. 1D). Reversing this point mutation diminished activity, and omitting the 7-nt RNA abolished activity altogether. We therefore speculate that, although the 7-nt RNA must still pair to this segment, a non–Watson-Crick distortion of the helix better accommodates long template RNAs. It is noteworthy that the four other active round-18 isolates also had point mutations within the segment designed to pair with the 7-nt RNA (*23*).

**Extensive and accurate RNA polymerization.** Although the round-18 ribozyme was only marginally improved over the round-10 ribozyme when short templates were used (Fig. 3), it was much better when longer templates were used (Fig. 4). With templates coding for 4, 8, 11, and 14 nucleotides, the round-18 ribozyme extended the primer by the corresponding number of residues (Fig. 4B). Normal RNA linkages were synthesized, as determined by nuclease analysis of the extension product (*23*). Furthermore, extension was predominantly by the Watson-Crick match to the template. When primers that were fully extended using the template coding for 11 nt were cloned and sequenced, 89 of 100 sequences precisely matched the template. Of the 1100 residues sequenced, only 12 were mismatches (Fig. 4C), implying an overall Watson-Crick error rate of 0.011 per nucleotide. Thus, the round-18 ribozyme can accurately use information from an RNA template and all four nucleoside triphosphates to extend an RNA primer by a complete turn of an RNA helix.

To examine the accuracy of polymerization more systematically, we measured the efficiency of matched and mismatched extension using four templates that differed only at the first coding nucleotide. For each template, the Watson-Crick match was added most efficiently (Table 2). The best fidelity was with the -C- template, for which the error rates ranged from 0.00004 to 0.0002. Fidelity was lower for the -G- and -U- templates, primarily because extension by the two G:U wobble mismatches had error frequencies of 0.044 and 0.085. The overall fidelity was 0.967 per residue. In other words, with all four NTPs supplied at equimolar concentrations, extension by the matched nucleotide typically would be 96.7% of the total extension.

A fidelity of 96.7% (Table 2) is somewhat lower than the 99% fidelity inferred from sequencing fully extended primer molecules (Fig. 4C). Two factors contribute to the higher fidelity observed in Fig. 4C. The first is the influence of sequence context on fidelity (*24*). The second arises from the fact that, after a mismatch was incorporated, further extension of the growing chain was less efficient because the 3′ terminus of the primer no longer paired with the template (*25*). Thus, the full-length product of Fig. 4C was enriched in molecules with few misincorporated nucleotides. Mismatch incorporation also reduces the extension efficiency of proteinaceous polymerases, a property particularly important for certain DNA polymerases because it facilitates exonucleolytic proofreading (*26*).

Polymerase fidelity is most simply expressed by assuming that all four NTPs are present at equal concentration, even though cellular NTP concentrations are not equimolar (*27*). For the round-18 ribozyme, certain asymmetric NTP ratios would produce observed fidelities significantly greater than 0.967. For example, lowering the GTP concentration to one-tenth that of the other NTPs would decrease G misincorporation by 10-fold, while only lowering the fidelity of extension across from C from 0.9996 to 0.996. Because G misincorporation was the major source of error, this would increase the overall fidelity from 0.967 to 0.985 with the templates in Table 2.

A Watson-Crick fidelity of 0.985 is still lower than the ≥0.996 fidelity seen with viral polymerases that replicate RNA by using RNA templates (*28*, *29*), and it is much lower than that seen for polymerases that replicate DNA (*30*). Nevertheless, the Watson-Crick fidelity of the round-18 ribozyme compares favorably to that of other ribozymes. Previously, the best ribozyme fidelity had been obtained with the engineered ligase derivative (Fig. 1A), which has an overall fidelity of 0.85 with equimolar NTPs and observed fidelities of 0.88 to 0.92 with more favorable

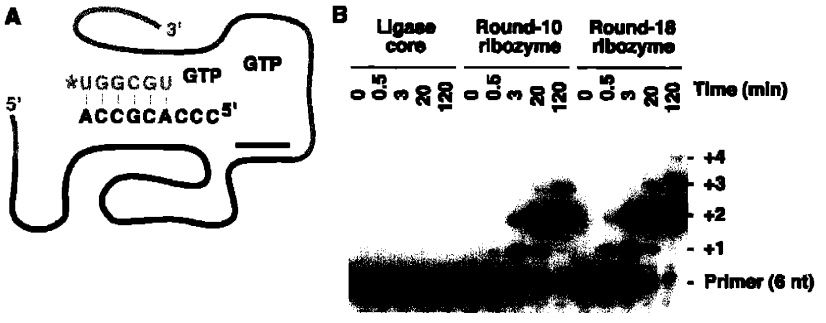



**Fig. 3.** Intermolecular primer extension using a short primer-template. (**A**) Schematic of the RNAs used in these polymerization reactions. Drawing conventions are as in Fig. 2A. Note that the primer is $^{32}$P end-labeled and that neither the primer RNA nor the template RNA is tethered or hybridized to the ribozyme. (**B**) PhosphorImager scan of a denaturing gel separating primer-extension products of the indicated ribozymes. Reactions included 1 μM ribozyme, 10 μM primer, 10.5 μM template, and 4 mM GTP, and were incubated for the indicated time in polymerization assay conditions (*33*). "Ligase core" refers to a ribozyme identical to that of Fig. 1A (black strand), except that its 3′ terminus was modified to pair with the 7-nt RNA (GGCACCA) that completes the ligase core; no extension was observed with this ribozyme. The round-10 and round-18 ribozymes are depicted in Figs. 1C and D, respectively. After long incubation times, some of the primer was extended with three templated residues plus one nontemplated residue. Many proteinaceous polymerases, including Qβ replicase (*42*) and *Taq* DNA polymerase (*43*), also tend to add an extra, nontemplated residue.

**Table 2.** Watson-Crick fidelity of RNA polymerization. For each template-NTP combination, the efficiency of extension by at least 1 nt was determined. For each template, the four efficiencies were normalized to that of the matching NTP, yielding the relative efficiencies of extension. The relative efficiency of extension for a mismatch is the same as its error rate (*27*) and misinsertion ratio (*26*). Fidelities were calculated as the efficiency for the match, divided by the sum of the efficiencies for all four NTPs. The average fidelity is the geometric average of the fidelities for each template (*41*). For each Watson-Crick match, the absolute efficiency (per molar per minute) is also shown in parentheses. It is reported as the observed rate constant of primer extension divided by NTP concentration, from polymerase assays (*33*) using 5 μM ribozyme, 2 μM primer (CUGCCAACCG), and 2.5 μM template (3′-GACGGUUGGCXCGCUUCG, where X is the indicated template residue). In these assays, NTPs were supplied at concentrations well below half-saturating.

| Template | Relative efficiency of extension | | | | Fidelity |
|---|---|---|---|---|---|
| | ATP | CTP | GTP | UTP | |
| -A- | 0.0034 | 0.0014 | 0.0043 | 1.0 (5.3) | 0.991 |
| -C- | 0.0002 | 0.0002 | 1.0 (5.4) | 0.00004 | 0.9996 |
| -G- | 0.0002 | 1.0 (41) | 0.0006 | 0.044 | 0.957 |
| -U- | 1.0 (87) | 0.0001 | 0.085 | 0.0002 | 0.921 |
| | | | | | Average = 0.967 |

NTP ratios (*15*). Moreover, the fidelity of the round-18 ribozyme approaches that of one proteinaceous polymerase, pol η, a eukaryotic polymerase needed for accurate replication of UV-damaged DNA (*31*). Yeast pol η has an overall fidelity of 0.984, which would increase to 0.989 with an optimal NTP ratio (*32*).

**General RNA-dependent RNA polymerization.** The round-18 ribozyme worked with every primer-template tested. As the primer-templates have no sequence features in common, the ribozyme does not rely on any sequence-specific contacts. Additionally, because the primer-template complex must shift in register relative to the polymerase active site each time another nucleotide is added, every polymerization experiment actually examines primer extension in a series of differing sequence contexts, demonstrating further that the polymerization is general with respect to nucleotide sequence. Granted, the efficiency of nucleotide addition varied depending on the sequence context, as evidenced by an uneven distribution of extension intermediates (Fig. 4), but this phenomenon is also observed with protein polymerases (*26*, *27*).

All templates used heretofore were less than 21 nt long, leaving open the question of whether the ribozyme could accommodate longer primer-template substrates, as would be required of an RNA replicase. To address this question, three related substrates were tested. The first was a short substrate, with a 10–base pair primer-template duplex and a 10-nt template coding region. The second substrate was the same, except its template coding region was lengthened from 10 to 100 nt. The ribozyme extended this substrate by as many as 9 nt in 23 hours, although somewhat less efficiently than it extended the short version (Fig. 5). The third substrate was the same as the second, except that its primer-template duplex was lengthened from 10 to 60 base pairs. The ribozyme extended this substrate just as efficiently as the second substrate. Thus, the ribozyme is free from steric constraints that would preclude polymerization using long templates or long primer-template helices.

Given this general recognition of primer-templates, the range for primer extension, currently just beyond one helical turn, is limited merely by the ribozyme's efficiency. Polymerization is too slow for more extension to be observed within 24 hours, and longer incubations yield limiting returns, because buffer and ionic conditions optimal for polymerization (*33*) also promote ribozyme and template degradation. Reactions would have to be supplemented periodically with fresh ribozyme to achieve polymerization substantially beyond one helical turn. Nonetheless, in initiating extension of a long primer hybridized to a long template, the round-18 ribozyme demonstrates polymerization that is

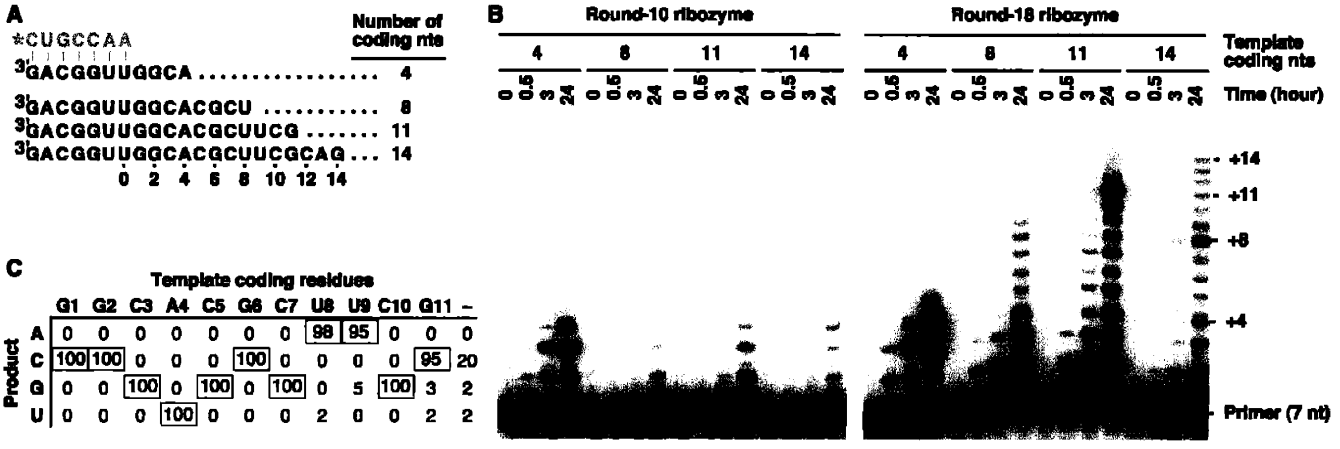



**Fig. 4.** Improved RNA polymerization. (**A**) Schematic of the 7-nt primer (orange) and the templates (red) with 4 to 14 coding residues used in these polymerization reactions. The ribozymes (not depicted) were as in Fig. 3. (**B**) PhosphorImager scan of a denaturing gel separating primer-extension products of the indicated ribozymes. Reactions included 5 μM ribozyme, 2 μM primer, 2.5 μM template, and 4 mM each NTP, and were incubated for the indicated time in polymerization assay conditions (*33*). (**C**) Tabulation of nucleotides inserted across from template residues during primer-extension by the round-18 ribozyme. Full-length primer-extension products encoded by the template with 11 coding residues were cloned, and 100 full-length clones were sequenced (*23*). For each template residue (red), the number of clones that had an A, C, G, or U at the corresponding position was tabulated. Tallies representing Watson-Crick matches to the template are boxed. The column below the red dash reports the identity of the nontemplated nucleotide (Fig. 3B, legend) added to the 3′ terminus of 24 of the sequenced primer-extension products. Coding residues are numbered as in (A).

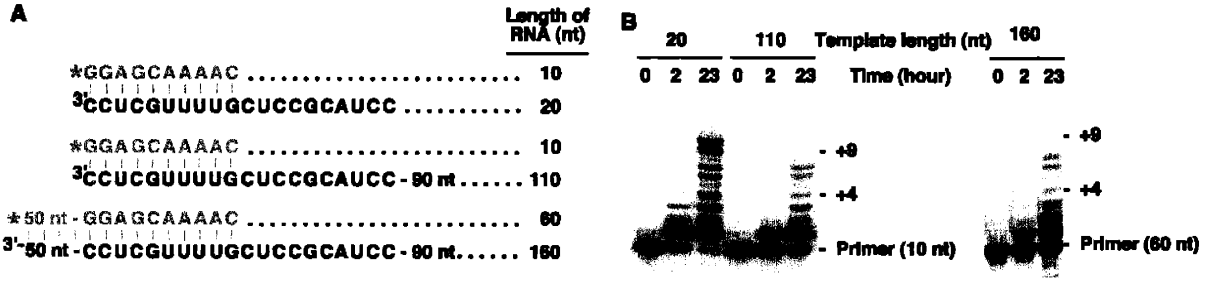



**Fig. 5.** RNA polymerization using long primers and long templates. (**A**) Schematic of the three primer-template combinations examined (*23*). Each combination included a 10- or 60-nt primer (orange) and a 20-, 110-, or 160-nt template (red). All three combinations have the same sequence near the site of polymerization (residues defined). (**B**) PhosphorImager scan of denaturing gels separating primer-extension products. Reactions included 5 μM round-18 ribozyme, 0.5 μM primer, 1 μM template, and 4 mM each NTP, and were incubated for the indicated time in polymerization assay conditions (*33*).

general not only with respect to the nucleotide sequence but also to the length of the primer-template complex.

General template-directed RNA polymerization requires recognition of the generic features of a primer-template complex in addition to ever-changing NTP specificity, as dictated by the next template residue. It is a complex reaction—one of the more sophisticated reactions catalyzed by single polypeptides. The demonstration that such an activity can be generated de novo, without reference to any biological ribozyme or structure, is a testament both to the catalytic abilities of RNA, as well as to modern combinatorial and engineering methodology. Key to this success may have been the stepwise procedure of first, isolating from random sequences an appropriate catalytic core in the context of a simple reaction (*13*, *14*); second, optimizing the sequence of the catalytic core (*16*); third, determining the limits of the core activity (*15*, *34*, *35*); and fourth, flanking the core with additional random sequence and selecting for more sophisticated substrate binding. Thus far, efforts to select for polymerization activity in a single step directly from random-sequence RNA have yielded only ribozymes that decorate themselves inappropriately with tagged nucleotides (*36*).

How could general polymerase activity have arisen on early Earth? If emergence of the first RNA replicase ribozyme coincided with the origin of life, it would have had to arise in a single step from prebiotically synthesized RNA, without the benefit of Darwinian evolution. Our shortest construct retaining activity was 165 nt, with about 90 nt involved in important Watson-Crick pairing and at least another 30 critical nucleotides (*23*). Ribozymes with the efficiency, accuracy, and other attributes of an RNA replicase might have to be even larger than this. However, current understanding of prebiotic chemistry argues against the emergence of meaningful amounts of RNA molecules even a tenth this length (*1*). This difficulty is anticipated by those who propose that life, and Darwinian evolution, began before RNA. Some speculate that in this "pre-RNA world," life was based on an RNA-like polymer, yet to be identified, that possessed the catalytic and templating features of RNA but also a more plausible prebiotic synthesis (*1*). The pre-RNA life forms presumably later developed the ability to synthesize RNA, facilitating the emergence of an RNA replicase ribozyme, which in turn enabled the transition to the RNA world.

It will be interesting to examine the extent to which continued mutation and selection can improve the activity of the polymerase ribozyme. Perhaps ribozymes with accuracy and efficiency sufficient for self-replication can be generated. The requisite fidelity may be close at hand, possibly only requiring a reduction of the overall error rate to one-third its current value, thereby increasing fidelity observed with unequal NTP concentrations from 0.985 to 0.995 (*37*). The increase in polymerization efficiency would need to be more substantial (at least 100-fold), although not beyond the degree of optimization achieved previously with in vitro evolution experiments. Other important issues will need to be addressed, including strand dissociation after polymerization. Nevertheless, the general polymerization activity of the round-18 ribozyme offers support for the idea of autocatalytic RNA replication in the distant past, as well as a new starting point for its demonstration in the not-so-distant future.

### References and Notes


1. G. F. Joyce, L. E. Orgel, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, New York, 1999), pp. 49–77.
2. N. R. Pace, T. L. Marsh, *Origins Life* **16**, 97 (1985).
3. P. A. Sharp, *Cell* **42**, 397 (1985).
4. T. R. Cech, *Proc. Natl. Acad. Sci. USA* **83**, 4360 (1986).
5. L. E. Orgel, *J. Theor. Biol.* **123**, 127 (1986).
6. A. J. Hager, J. D. Pollard, J. W. Szostak, *Chem. Biol.* **3**, 717 (1996).
7. D. P. Bartel, in *The RNA World* R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999), pp. 143–162.
8. D. P. Bartel, P. J. Unrau, *Trends Cell Biol.* **9**, M9 (1999).
9. J. W. Szostak, D. P. Bartel, P. L. Luisi, *Nature* **409**, 387 (2001).
10. J. A. Doudna, J. W. Szostak, *Nature* **339**, 519 (1989).
11. J. A. Doudna, S. Couture, J. W. Szostak, *Science* **251**, 1605 (1991).
12. R. Green, J. W. Szostak, *Science* **258**, 1910 (1992).
13. D. P. Bartel, J. W. Szostak, *Science* **261**, 1411 (1993).
14. E. H. Ekland, J. W. Szostak, D. P. Bartel, *Science* **269**, 364 (1995).
15. E. H. Ekland, D. P. Bartel, *Nature* **382**, 373 (1996).
16. ———, *Nucleic Acids Res.* **23**, 3231 (1995).
17. Each subpool had the structure shown in Fig. 1B but a different percent mutagenesis of ligase core residues. For a residue mutagenized at 20%, the nucleotide precursor reservoir was doped such that the parental base would be added at a frequency of 0.8, whereas each of the other three bases would be added at a frequency of 0.067. Each subpool was constructed starting with two synthetic single-stranded pools (*16*). The only black residues of Fig. 1B that were not mutagenized were the six residues at the 3′ terminus of the ligase core domain. These six residues (GGUGCC) corresponded to the Ban I restriction site used to construct the subpools and were designed to pair with the 7-nt RNA that completes the ligase domain (GGCACCA, purple RNA in Fig. 1B). An average of three RNA copies of each subpool were combined to generate a starting pool with a sequence complexity of $2 \times 10^{15}$.
18. Pool extension reactions typically included 0.2 to 0.5 μM pool RNA linked to primer [Fig. 1B; (*38*)], 1.0 μM template RNA (Table 1), 1.0 μM GGCACCA RNA (Fig. 1B), and 0.1 to 2 mM tagged NTPs (Table 1) in pool extension buffer (60 mM $MgCl_2$, 200 mM KCl, 50 mM EPPS, pH 8.0, 22°C). In rounds 10 to 18, incubations also included 1.0 μM reverse-transcription primer (TTCAGATTGTAGCCTTC) and lacked KCl. The pool RNA and oligonucleotides were mixed together in water, incubated at 80°C for 3 min, then allowed to cool to room temperature for 10 min. Pool extension reactions were started by the simultaneous addition of pool extension buffer and NTPs. Reactions were stopped with addition of 80 mM EDTA after 0.1 to 36 hours (Table 1). Excess 4-thioUTP was removed with a Centricon YM-10 centrifugal filter device (Millipore), and molecules that had been tagged with 4-thioU were isolated on APM gels (*44*). APM gels were prepared by using 10 to 80 μM APM in the lower portion of the gel, no APM in the top portion of the gel, and 8M urea per 5% acrylamide throughout the gel (Fig. 2). For isolation of RNA extended with at least one 4-thioU (rounds 1 to 10, Table 1), RNA was eluted from a gel slice extending from the APM interface to the migration of pool RNA with a single 4-thioU. For isolation of RNA extended by at least two 4-thioUs (Table 1), RNA was eluted from a slice containing only the APM interface. In rounds 2 to 5, 9, 12 to 14, and 17, RNA was further purified on a second APM gel. Eluted RNA was precipitated and reverse-transcribed (*16*) by using primer CGGGACTCTGACCTTGG (rounds 1 to 3, 6) AAACGGGACTCTGACCTTG (rounds 4, 5, 7, 9), TTCGGGACTCTGACCTT (rounds 8, 10), or TTCAGATTGTAGCCTTC (rounds 11 to 18). RT primers were alternated in rounds 4 to 10 to vary the 3′-terminal residues of the pool RNA, disfavoring the selection of molecules that extend their 3′ terminus rather than extending the attached primer. In rounds including biotinylated ATP (biotin-N6-ATP, New England Nuclear), RNA-cDNA duplex molecules were also purified by capture on streptavidin magnetic beads (*16*). PCR amplification of the cDNA used a primer that completed the T7 RNA polymerase promoter sequence, (TTC*TAATACGACTCACTATAGGACAACC*, italics, T7 promoter sequence; underline, 5′ primer-binding site, Fig. 1B). PCR DNA was transcribed (*16*) to generate RNA for the next round of selection or cloned (Topo cloning kit, Invitrogen) for sequencing and further analysis.
19. Throughout this manuscript, added nucleotide residues (pA, pC, pG, and pU) are abbreviated by the corresponding nucleoside symbols (A, C, G, and U).
20. W. K. Johnston, D. P. Bartel, data not shown.
21. To examine whether the ribozyme preferred substrate sequences used during its selection, the primer-template from selection rounds 1, 6, 8, and 9 (Table 1) was also tested and found not to be used any more efficiently than the substrate of Fig. 3.
22. The starting pool for round 11 was constructed as in (*17*), except: (i) it was based on the round-10 ribozyme; (ii) it was a combination of two subpools, with the ligase core positions mutagenized at 0 or 3%; (iii) blue positions of Fig. 1C were mutagenized at 20%; and (iv) a new RT-PCR primer-binding site (GAAGGCUACAAUCUGAA) was appended to the 3′ end of the pool molecules. About three RNA copies of each subpool were combined to generate a starting pool with a sequence complexity of $10^{15}$.
23. Supplemental material describing comparative sequence analysis of ribozyme variants, engineering of shorter and more active polymerase constructs, nuclease analysis of the primer-extension product, methods for Fig. 4C, and RNAs of Fig. 5 is available on *Science* online at www.sciencemag.org/cgi/content/full/292/5520/1319/DC1
24. The contribution of sequence context [as opposed to sequence composition (*45*)] to the higher fidelity in Fig. 4C is particularly apparent with the extension across from G1, G2, and G6, which was by C at all three positions in all 100 products sequenced. In contrast, Table 2 shows that when coding residue 4 was changed to G, 4% of the extension across from that G was by the U mismatch. To confirm this apparent effect of sequence context, the relative efficiencies of matched and mismatched extension across from the G6 coding residue were determined as in Table 2, by using the primer CUGCCAACCGUG and the template, ribozyme, and conditions of Fig. 4C. The efficiency with the CTP match was 27 $M^{-1}$ $min^{-1}$, and relative efficiencies were 0.0007, 1.0, 0.002, and 0.017 with ATP, CTP, GTP, and UTP, respectively, for a fidelity at G6 of 0.981—somewhat better than the 0.957 fidelity seen across from a G at coding residue four (Table 2). Thus, sequence context can influence accuracy of polymerization, a phenomenon also observed with protein polymerases (*26*, *27*).
25. For example, in one context, relative extension efficiencies were 1.0, 0.044, 0.78, or 0.056 when the previous nucleotide was a U incorporated across from an A, C, G, or U template residue, respectively.
26. H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **60**, 477 (1991).

27. T. A. Kunkel, *Bioessays* **14**, 303 (1992).
28. D. A. Steinhauer, J. J. Holland, *J. Virol.* **57**, 219 (1986).
29. C. D. Ward, J. B. Flanegan, *J. Virol.* **66**, 3784 (1992).
30. L. A. Loeb, T. A. Kunkel, *Annu. Rev. Biochem.* **51**, 429 (1982).
31. M. T. Washington, R. E. Johnson, S. Prakash, L. Prakash, *J. Biol. Chem.* **274**, 36835 (1999).
32. When using $k_{cat}/K_m$ values from (*31*), the optimal ratio of dNTP concentrations for yeast pol η would be 100:23:90:35 for dATP, dCTP, dGTP, and TTP, respectively. When using the values from Table 2 for the round-18 ribozyme, a ratio of 50:100:4:25 for ATP, CTP, GTP, and UTP would result in an observed fidelity of 0.990.
33. Ribozyme RNAs were transcribed from PCR-generated DNA templates, which were produced by using plasmid DNA and the appropriate primers. The DNA template for the round-18 ribozyme had penultimate 2′-methoxyl substitutions to reduce the addition of nontemplated residues at the ribozyme 3′ terminus (*46*). Standard polymerization assays were in 200 mM $MgCl_2$, 50 mM tris, pH 8.5, at 22°C. Although polymerization using an intermolecular primer was readily observed in the conditions of the selection (60 mM $MgCl_2$, 200 mM KCl, 50 mM EPPS, pH 8.0, 22°C), the standard assay conditions were more optimal. Ribozyme, primer, template, and NTPs were included at the concentrations indicated, and the GGCACCA RNA that completes the ligase domain was in 1.25-fold molar excess over the ribozyme concentration. These RNAs were mixed together in water, incubated at 80°C for 3 min, then incubated at 22°C for at least 5 min before starting the reaction by simultaneous addition of buffer, salt, and NTPs. Heating the ribozyme separately from the primer and template RNAs did not affect the polymerization reaction kinetics. Reactions were stopped by addition of 1.6 volumes of 145 mM EDTA/6 M urea, heated (90°C, 5 min) in the presence of competitor RNA designed to hybridize to the template RNA, then analyzed on sequencing gels. Incubation with competitor RNA, added in 10-fold molar excess over template RNA, led to better gel resolution because it prevented reassociation of the extended primer with the template.
34. N. H. Bergman, W. K. Johnston, D. P. Bartel, *Biochemistry* **39**, 3115 (2000).
35. M. E. Glasner, C. C. Yen, E. H. Ekland, D. P. Bartel, *Biochemistry* **39**, 15556 (2000).
36. P. J. Unrau, D. P. Bartel, unpublished results.
37. Consider hypothetical replicases about the size of the round-18 ribozyme (~200 nt) that synthesize an average of five 200-nt RNA strands within their lifetime. For replication through complementary-strand then second-strand synthesis, two of the five strands produced by each polymerase must have the correct residues at all of the positions that contribute to function. If the identities of effectively 80% of the 200 nucleotides contribute to function, then these replicases would require a fidelity of at least 0.995 ($0.995^{160} = 0.45 \approx 2/5$). Note that this scenario does not account for the dilution of productive ribozyme that would occur as active ribozyme variants replicate an increasing number of inactive variants; compartmentalization and selection would be needed to achieve sustainable replication (*7*, *9*).
38. The 5′ terminus of gel-purified pool RNA was ligated (*47*, *48*) to the DNA portion of a DNA-RNA chimeric oligonucleotide, 3′-ATATCACTC-5′-5′-$X_{(6-8)}$ (residues defined, deoxynucleotides synthesized by using "reverse" phosphoramidites from Glenn Research; 5′-5′, phosphodiester linkage joining the 5′ terminus of the DNA with the 5′ terminus of the RNA; $X_{(6-8)}$, RNA primer segment 6 to 8 nt in length complementary to the template indicated in Table 1). The pool was then gel-purified to remove the splint oligonucleotide needed for ligation.
39. An A analog, 2-aminopurine pairs with U and 4-thioU, but unlike A, its Watson-Crick pairing with 4-thioU does not involve the sulfur atom. Because the sulfur atom subtly distorts pairing geometry, templates with 2-aminopurine were used in later rounds of selection.
40. Repeated error-prone PCR was performed as described (*13*), except serial dilutions were after every 6 cycles of PCR. RNA from four subpools with expected mutagenesis levels of 0, 2, 4, and 6%, was mixed and used for round 15. RNA from two subpools with expected mutagenesis levels of 0 and 1% was used for round 16. RNA from two subpools with expected mutagenesis levels of 0 and 2% was used for round 17.
41. Because the net accuracy of RNA polymerization is represented by the product of the fidelities for each added nucleotide, geometric averages are reported throughout this study when describing overall fidelity.
42. J. N. Bausch, F. R. Kramer, E. A. Miele, C. Dobkin, D. R. Mills, *J. Biol. Chem.* **258**, 1978 (1983).
43. J. M. Clark, *Nucleic Acids Res.* **16**, 9677 (1988).
44. P. J. Unrau, D. P. Bartel, *Nature* **395**, 260 (1998).
45. Sequence composition of the template explains practically none of the higher fidelity in Fig. 4C. The template of Fig. 4C has one A, four C, four G, and two U coding residues. Thus, the expected fidelity per nucleotide, calculated by using the values of Table 2, is $[0.991 \times (0.9996)^4 \times (0.957)^4 \times (0.921)^2]^{(1/11)} = 0.969$, a value very close to the generic overall fidelity of 0.967 calculated in Table 2.
46. C. Kao, M. Zheng, S. Rudisser, *RNA* **5**, 1268 (1999).
47. M. J. Moore, P. A. Sharp, *Science* **256**, 992 (1992).
48. T. Tuschl, M. P. Ng, W. Pieken, F. Benseler, F. Eckstein, *Biochemistry* **32**, 11658 (1993).
49. We thank members of the laboratory and F. Solomon for helpful comments on this manuscript. Supported by grants from the NIH (D.P.B.), a Medical Research Council (Canada) postdoctoral fellowship (P.J.U.), and a Howard Hughes Medical Institute predoctoral fellowship (M.E.G.).

16 March 2001; accepted 10 April 2001