

Searching for ribozymes with new and improved properties

by

Edward A. Curtis

B.A., Ecology and Evolutionary Biology
Princeton University, 1997

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

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2006

© 2006 Edward A. Curtis. All Rights Reserved.

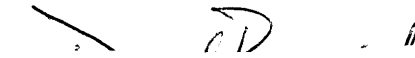
The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part.

Signature of Author:




Edward A. Curtis
Department of Biology
(September 11, 2006)

Certified by:

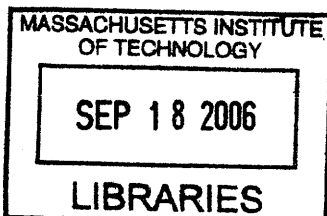


David P. Bartel
Professor of Biology
Thesis Supervisor

Accepted by:



Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Committee



ARCHIVES

Searching for ribozymes with new and improved properties

by

Edward A. Curtis

Submitted to the Department of Biology
on September 11, 2006
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Biology

ABSTRACT

The technique of *in vitro* selection was used to learn more about the properties of RNA sequence space. In Chapter 1 I investigated how readily ribozymes with new catalytic activities and folds could arise from an existing ribozyme scaffold. *In vitro* selection was used to isolate 23 kinase ribozymes from a pool of more than 10^{14} variants of an aminoacylase parent ribozyme. The density of kinase ribozymes in the pool increased dramatically as the mutational distance from the starting ribozyme increased, suggesting a need to escape the fold of the parent. Consistent with this idea, the folds of two kinases characterized in detail were different from that of the starting aminoacylase scaffold. In Chapter 2 I investigated the extent to which one of these kinase ribozymes could be optimized using a method of recombination called synthetic shuffling. Point mutations from previously isolated sequence variants of this ribozyme were shuffled in more than 10^{14} different combinations, and active variants isolated by *in vitro* selection. The rate of the most efficient ribozyme identified was 30-fold faster than that of the most efficient ribozyme used to build the pool, with a second order rate enhancement approaching 10^{10} -fold. Further analysis revealed two groups of mutations, derived from two different ribozymes used to build the shuffled pool, that each increased the rate of the ribozyme by approximately 30-fold. The effects of these mutations were independent of one another, and when combined produced a ribozyme with a rate 600-fold faster than that of the initial isolate. Together, these experiments provide insight into how RNA sequence space can be searched for ribozymes with new and improved properties.

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

Dedication

To my parents, Ed and Joy
My brother and sister-in-law, Matt and Amy
And my girlfriend Tamara

Acknowledgements

Thanks to my advisor David Bartel for teaching me about *in vitro* selection
Mike Lawrence and Suzanne Nguyen for being great friends and colleagues
Uli Muller for encouragement and the Wall of Interest
And members of the lab for advice and support

Table of Contents

Abstract	2
Dedication	3
Acknowledgements	4
Table of Contents	5
Introduction	6
Chapter One	24
New catalytic structures from an existing ribozyme	
Chapter 1 has been published previously as: Curtis, E.A. and Bartel, D.P. New catalytic structures from an existing ribozyme. <i>Nat.</i> <i>Struct. Mol. Biol.</i> 12 , 994-1000 (2005).	
Chapter Two	66
Modularity of a catalytic RNA	
Future Directions	105
Appendix	109
The hammerhead cleavage reaction in monovalent cations	
The appendix has been published previously as: Curtis, E.A. and Bartel, D.P. The hammerhead cleavage reaction in monovalent cations. <i>RNA</i> 7 , 546-552 (2001).	

INTRODUCTION

The ability of living cells to catalyze transformations such as fermentation has been recognized for thousands of years, but the power to recapitulate such processes *in vitro* is a more recent discovery¹. Buchner was the first to demonstrate that a yeast cell extract could be used to ferment sugar into alcohol, and he noted that “With respect to the theory of fermentation, the following statement can now be made: it is not necessary to have an apparatus as complicated as the living yeast cell for fermentation to take place. Rather a soluble ferment, or enzyme, is to be considered as the carrier of the fermentation activity”¹. As purification methods became more rigorous, it soon became apparent that enzymes were proteins. By the late 1960's, however, scientists such as Crick, Orgel, and Woese had started to speculate that RNA molecules might also be capable of enzymatic activity²⁻⁴. This speculation was based in part on the idea that RNA was capable of forming complex and highly ordered structures², as well as the observation that RNA was an especially effective template^{2,4}. Orgel put it this way: "Could polynucleotide chains with well-defined secondary structures act as primitive enzymes? I doubt that they alone could exhibit extensive catalytic activity, although one cannot be quite sure"². The crystal structure of tRNA, unveiled in the mid 1970's, was consistent with the catalytic RNA hypothesis: the three-dimensional fold of this molecule was well-defined, and contained pockets that appeared capable of binding and orienting substrates so that catalysis could occur^{5,6}.

The hypothesis that RNA could act as an enzyme was confirmed in the early 1980s, when the Cech group showed that a purified intron from the 26S rRNA gene of the ciliate *Tetrahymena thermophila* could splice itself from its RNA transcript in the absence of any added protein⁷. An intron synthesized by *in vitro* transcription was also

capable of self-splicing, providing strong evidence that this reaction was not being catalyzed by contaminating protein⁸. The reaction required moderate concentrations of monovalent and divalent metal ions, which is a typical requirement for RNA enzymes, and an exogenous guanosine cofactor, which is used as a nucleophile in the first step of splicing⁷. This self-splicing RNA motif became known as the Group I intron ribozyme.

At about the same time the Group I intron was discovered, Altman, Pace and colleagues found that the RNA component of RNase P was also a ribozyme. Previous analysis had shown that both the RNA and protein components of the enzyme were required for activity in *Escherichia coli*, and initial analysis had suggested that this was also the case *in vitro*⁹. At elevated Mg²⁺ concentrations, however, they found that the RNA components of the enzyme from both *Escherichia coli* (M1 RNA) and *Bacillus subtilis* (P-RNA) were catalytically active by themselves⁹. Several control experiments suggested that this catalytic activity was not due to contaminating protein: M1 RNA was still catalytically active after protease treatment, and protein could not be detected in the sample by Coomassie staining. RNase P exhibited two properties of a classical enzyme that were not shared by the Group I intron: it catalyzed multiple turnover RNA cleavage, and it remained unchanged over the course of the reaction⁹.

In the years following these discoveries several other naturally occurring ribozymes were identified¹⁰. The reactions catalyzed by these ribozymes include RNA splicing (the Group I¹¹ and Group II¹² introns), cleavage of tRNA and rRNA (RNase P¹³), and self-cleavage (the hairpin¹⁴, HDV¹⁵, hammerhead¹⁶, VS¹⁷, glmS¹⁸, and beta-globin¹⁹ ribozymes). Although each of these reactions involves phosphoryl transfer, the catalytic repertoire of naturally occurring ribozymes is not limited to such reactions: peptide bond

formation in the ribosome is also catalyzed by RNA. Early support for this idea came from biochemical studies, in which purified ribosomes from *Thermus aquaticus* were shown to catalyze peptide bond formation after being stripped of most protein²⁰, and further evidence was obtained when the ribosome was crystallized in the presence of several transition state analogs of peptide bond formation^{21,22}. These analogs were bound to 23S rRNA, and were about 18 angstroms away from the closest side-chain atoms of ribosomal proteins²². Unlike other naturally occurring ribozymes, synthetic 23S rRNA has not been observed to catalyze peptide bond formation by itself²³, suggesting that this ribozyme requires some combination of ribosomal proteins, other ribosomal RNAs, or nucleotide modifications to adopt its catalytically active conformation.

Although the discovery of naturally occurring ribozymes provided important insight into the catalytic abilities of RNA, it also raised many new questions. With regard to the evolutionary potential of RNA an especially important question was whether RNA could catalyze reactions other than phosphoryl transfer and peptide bond formation, and more generally whether RNA could adopt a wide range of phenotypes. In some respects these questions were first addressed by the pioneering experiments of Spiegelman and colleagues, although at the time of this work it had not been shown that RNA could function as a catalyst. In these experiments a protein called Q β -replicase was used to copy the RNA genome of bacteriophage Q β *in vitro*. By serially diluting the reaction mixture, replication of Q β -RNA could be followed for many generations. This system was used to investigate how Q β -RNA would evolve in response to various selective pressures. In the first study, Mills et al.²⁴ progressively shortened the time of the replication reaction, and after 74 serial dilutions obtained a Q β -RNA variant that

could replicate 15-fold faster than the initial Q β -RNA. In subsequent experiments Q β -RNA variants were isolated that could replicate more efficiently at reduced CTP concentrations²⁵, in the presence of the replication inhibitor tubercidin triphosphate²⁵, and in the presence of the denaturant ethidium bromide^{26,27}. These studies showed that, with respect to a given process such as replication, RNA molecules could adopt many types of phenotypes.

By the late 1980s the convergence of several technologies made it possible to explore the range of phenotypes available to RNA molecules in much greater detail. Improved methods to chemically synthesize DNA had been developed, allowing large DNA pools of defined sequence to be built (reviewed in 28), and purified T7 RNA polymerase could be used to generate μg quantities of RNA *in vitro*²⁹. Using these two technologies pools containing many different RNA sequences could be constructed. In 1970 reverse transcriptase was isolated, an enzyme that could generate cDNA from an RNA template³⁰⁻³², and by the mid 1980s the development of PCR allowed workers to exponentially amplify DNA molecules^{33,34}. By combining reverse transcription and PCR, RNA molecules, such as those that survived a selection step, could be exponentially amplified. Finally, using methods to clone^{35,36} and sequence^{37,38} DNA that had been developed in the 1970s, workers could determine the sequences of RNA molecules with interesting phenotypes.

These techniques were combined to produce a method called *in vitro* selection³⁹⁻⁴². The basic idea involves three steps. First, a diverse pool of RNA sequences is enriched for molecules with some interesting property. For example, a pool could be searched for RNAs that bind a small molecule immobilized on a column. Second, molecules that

survive the enrichment process are amplified by reverse transcription and PCR. Third, PCR products are transcribed using T7 RNA polymerase to generate RNA for the next round of selection. Because this process can be iterated, even RNA molecules with extremely rare phenotypes can be isolated from large pools of inactive sequences. Soon after the idea of *in vitro* selection was developed, a series of landmark papers demonstrated the power of this approach.

In the first published study, Tuerk and Gold⁴⁰ set out to better characterize the binding interaction between T4 RNA polymerase and the mRNA that encodes it. Translation of this mRNA is inhibited when it is bound to T4 RNA polymerase, and this limits the amount of T4 RNA polymerase that is synthesized when it is abundant in the cell. Tuerk and Gold⁴⁰ generated all $4^8 = 66,536$ possible variants of an 8 nucleotide site in the 5' leader sequence of the mRNA, and used *in vitro* selection to isolate variants that could bind T4 RNA polymerase protein immobilized on a filter. Two of the recovered sequences bound equally well to the protein: one corresponded to the wild-type sequence, and the other differed at 4 of 8 positions and had not previously been described.

Soon afterwards, several studies investigated whether *in vitro* selection could be used to isolate variants of existing ribozymes with altered phenotypes. In one set of experiments, variants of the Group I intron that could cleave a DNA substrate more efficiently were isolated^{41,43}, and in a different study variants that functioned more efficiently in the presence of Ca^{2+} were obtained⁴⁴. In each case these variants differed from the prototype Group I intron sequence at only a few nucleotides, and the fold of the intron appeared to be preserved.

Other workers set out to determine whether RNAs with interesting phenotypes could be isolated from large pools of random RNA sequences. From such a pool containing more than 10^{13} different sequences, Ellington and Szostak⁴² isolated RNA molecules that could bind to small organic dyes immobilized to agarose. Approximately one in 10^{10} random sequences was capable of binding. These RNAs, called aptamers, bound their targets with dissociation constants in the high micromolar range. Some were specific for particular target molecules, while others could not discriminate among the various dyes used in the selection.

Soon afterwards, Bartel and Szostak⁴⁵ demonstrated that ribozymes with new catalytic activities could also be isolated from large pools of random sequences. They generated a pool containing 10^{15} different sequences, in which a 5' constant region that could base pair with a substrate oligonucleotide was attached to a random sequence domain flanked by a 3' constant region. Variants that could covalently attach the substrate to themselves were isolated using an oligonucleotide affinity column complementary to the substrate. These molecules were then amplified by reverse transcription and PCR using a primer complementary to the 3' constant region of the pool and a primer corresponding to the sequence of the substrate oligonucleotide, so that only molecules that had attached the substrate to themselves would be amplified. About 65 different RNA ligase ribozymes were isolated from the pool, indicating that at least one in 2×10^{13} random RNA sequences of this length can catalyze RNA ligation. Although not as efficient as protein enzymes, some of these ribozymes enhanced the rate of template-directed RNA ligation by up to 7×10^6 -fold, and further optimization yielded variants that enhanced the reaction by 10^9 -fold^{46,47}.

In the following years many types of new ribozymes were isolated by *in vitro* selection, typically from large pools of random sequence. Longer pools are more difficult and more expensive to make than shorter pools, but they may be enriched for superior catalysts: more efficient ribozymes (and aptamers) tend to have higher information contents than less efficient ribozymes (and aptamers)⁴⁸, and more complex folds are thought to be more abundant in longer pools than in shorter ones^{49,50}. Several studies suggest that, when searching for a completely new activity, starting with an existing RNA scaffold rather than random sequence may actually be harmful. For example, in Chapter 1 of this thesis I found that, as the mutational distance from an aminoacylase ribozyme scaffold increased, the density of kinase ribozymes in sequence space also increased⁵¹. Furthermore, the folds of these kinase ribozymes were different from that of the starting aminoacylase scaffold⁵¹. Similar results were obtained in several studies using aptamers⁵²⁻⁵⁴, although this does not appear to be true when an aptamer scaffold is used to isolate RNAs that bind a similar target molecule. For example, variants of a citrulline aptamer that could bind arginine were found to have the same fold as that of the starting citrulline aptamer^{55,56}.

Other studies have investigated how to optimize existing ribozymes. In many cases the sequence neighborhood of a ribozyme is enriched for variants with improved catalytic activities, and these variants can be isolated by randomly mutagenizing the ribozyme and re-selecting for more efficient variants^{42,46}. Mutagenesis is accomplished by either synthesizing a partially degenerate oligonucleotide template encoding the ribozyme of interest^{42,46} or by amplifying the template by mutagenic PCR⁵⁷. In most cases, the fastest ribozymes isolated using this approach are 10 to 50-fold faster than the

starting sequence (Curtis, Lawrence, and Bartel, in preparation). Another method of ribozyme optimization, based on a recombination method called synthetic shuffling⁵⁸, is described in Chapter 2 of this thesis (Curtis and Bartel, in preparation). Point mutations from sequence variants of a kinase ribozyme that had been previously obtained by random mutagenesis and re-selection⁵¹ were shuffled in more than 10^{14} different arrangements, and improved ribozyme variants were isolated by *in vitro* selection. The rate of the fastest kinase isolated was approximately 30-fold faster than that of the fastest ribozymes used to build the pool. In a different approach, a technique called nonhomologous random recombination can be used to generate pools of topological variants of ribozymes or aptamers⁵⁹. In one experiment this method was used to generate an aptamer with a binding affinity 46-fold greater than that of the aptamers used to build the pool⁵⁹. In another case, this method yielded variants of a pyrimidine synthase ribozyme that were smaller, but not substantially faster than the ribozyme used to generate the pool⁶⁰.

Most selection schemes used to isolate new ribozymes require molecules to covalently modify themselves (in the presence of a substrate) in such a way that they can be separated from inactive members of the pool. In one example of this approach, Lorsch and Szostak isolated kinase ribozymes by incubating pool RNA with ATP γ S, and capturing thiophosphorylated molecules using a thiopropyl column⁶¹. In a second, less common type of selection scheme, a pool is incubated with an immobilized transition state analog of the reaction of interest (analogous to the approach used in the isolation of catalytic antibodies⁶²), and RNAs that can bind this transition state analog are isolated, some of which can typically catalyze the reaction. The use of transition state analogs has

expanded the known scope of RNA-mediated catalysis by facilitating the isolation of ribozymes that could not be readily identified using other approaches. Examples include a ribozyme that catalyzes bridged biphenyl isomerization⁶³ and several that catalyze porphyrin metalation^{64,65}.

A limitation of these approaches is that they do not permit direct selection for multiple turnover. This issue has been recently addressed by several groups using a technique called *in vitro* compartmentalization^{66,67}. In one of these studies⁶⁷, DNA templates encoding variants of a ligase, as well as one of the two RNA substrates of the ligase, were immobilized on microbeads. Each microbead was linked to a single template sequence and a large excess of substrate. The microbeads were then emulsified in the presence of components required for *in vitro* transcription, as well as an excess of the second (fluorophore-containing) substrate of the ligase. Ribozymes that could catalyze multiple turnover ligation labeled the microbead attached to their template with multiple copies of the fluorophore containing substrate, and these microbeads were isolated by FACS. Although the technique is currently limited to pools containing about 10^9 different sequences, it may represent a general way to select for ribozymes that catalyze multiple reactions.

A goal of many *in vitro* selection experiments has been to identify types of reactions that can be catalyzed by RNA. This has been motivated in part by the RNA World hypothesis⁶⁸, which postulates that ribo-organisms once ruled the earth. At the very least this would require a ribozyme that can catalyze its own replication, and it might appear more plausible if ribozymes were shown to catalyze a wide range of reactions^{69,70}. These studies indicate that RNA is a versatile catalyst⁶⁹. Ribozymes have

been isolated that catalyze RNA phosphorylation⁶¹, RNA ligation^{45,71-75}, RNA cleavage⁷⁶⁻⁷⁸, RNA aminoacylation⁷⁹, RNA alkylation,⁸⁰ peptide bond formation⁸¹, glycosidic bond formation^{82,83}, carbon-carbon bond formation^{84,85}, and oxidization-reduction⁸⁶. Although these studies indicate that ribozymes catalyzing simple chemical reactions can be isolated from large pools of random sequences, ribozymes with more sophisticated activities are not always as readily accessible. For example, it has not been possible to directly isolate a template-dependent RNA polymerase ribozyme from a random-sequence pool⁸⁷. In several instances such ribozymes have been isolated by building them one domain at a time. In the case of the polymerase ribozyme isolated by Bartel and co-workers, this involved first isolating a ribozyme that could catalyze the appropriate chemistry, and then selecting for a second domain that facilitated more efficient template-dependent polymerization^{45-47,87,88}.

Another major goal has been to generate ribozymes, either by engineering or *in vitro* selection, that can be used as tools in medicine, industry, and basic research⁸⁹⁻⁹¹. In one example of this approach, an engineered version of the Group I intron was used to repair mutant β -globin mRNA transcripts in erythrocyte precursor cells⁹². Engineered hammerhead and hairpin ribozymes have also been used to suppress the expression of mRNAs in cells^{93,94}. Because these ribozymes recognize their substrates by base-pairing, and can cleave virtually any sequence, designing variants that target a specific mRNA is relatively straightforward. An extension of this approach has been used to identify genes associated with particular phenotypes⁹⁵⁻⁹⁷. In one study a plasmid library encoding hammerhead ribozymes with randomized substrate-binding arms was introduced into MCF-7 cells, which were then chemically treated to induce apoptosis⁹⁶. Many of the

surviving cells contained hammerhead variants with substrate-binding arms complementary to genes known to be involved in apoptosis, while other contained hammerheads with substrate-binding arms complementary to possibly new apoptotic genes. In another interesting application, *in vitro* selection was used to generate self-cleaving ribozymes that are allosterically activated by small molecules such as cAMP⁹⁸¹⁰⁰. These ribozymes can be used to measure the concentration of various small molecules in environmental or biological samples, and might also be used to monitor the concentration of small molecules *in vivo*.

In conclusion, the technique of *in vitro* selection has significantly expanded the known scope of RNA-mediated catalysis. Much has been learned about how to search sequence space for new ribozymes, as well as optimized versions of existing ribozymes, and these efforts have led to the isolation of ribozymes that perform complex tasks such as RNA-dependent RNA polymerization. The continued development and application of such techniques will hopefully lead to the isolation of even faster and more sophisticated ribozymes that can be used in medicine, industry, and basic research, and that bring great pleasure to their creators.

REFERENCES

1. Dressler, D. & Potter, H. *Discovering Enzymes*, (W.H. Freeman and Company, New York, New York, 1990).
2. Orgel, L.E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381-393 (1968).
3. Crick, F.H. The origin of the genetic code. *J. Mol. Biol.* **38**, 367-379 (1968).
4. Woese, C.R. The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives. *Proc. Natl. Acad. Sci. USA* **59**, 110-117 (1968).
5. Suddath, F.L. *et al.* Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0 angstroms resolution. *Nature* **248**, 20-24 (1974).
6. Robertus, J.D. *et al.* Structure of yeast phenylalanine tRNA at 3 angstroms resolution. *Nature* **250**, 546-551 (1974).
7. Cech, T.R., Zaug, A.J. & Grabowski, P.J. *In vitro* splicing of the ribosomal RNA precursor of *Tetrahymena*: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487-496 (1981).
8. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**, 147-157 (1982).
9. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857 (1983).
10. Doudna, J.A. & Cech, T.R. The chemical repertoire of natural ribozymes. *Nature* **418**, 222-228 (2002).
11. Cech, T.R. Self-splicing of group I introns. *Annu. Rev. Biochem.* **59**, 543-568 (1990).
12. Fedorova, O., Su, L.J. & Pyle, A.M. Group II introns: highly specific endonucleases with modular structures and diverse catalytic functions. *Methods* **28**, 323-335 (2002).
13. Frank, D.N. & Pace, N.R. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.* **67**, 153-180 (1998).
14. Fedor, M.J. Structure and function of the hairpin ribozyme. *J. Mol. Biol.* **297**, 269-291 (2000).
15. Shih, I.H. & Been, M.D. Catalytic strategies of the hepatitis delta virus ribozymes. *Annu. Rev. Biochem.* **71**, 887-917 (2002).
16. Blount, K.F. & Uhlenbeck, O.C. The structure-function dilemma of the hammerhead ribozyme. *Annu. Rev. Biophys. Biomol. Struct.* **34**, 415-440 (2005).
17. Collins, R.A. The *Neurospora* Varkud satellite ribozyme. *Biochem. Soc. Trans.* **30**, 1122-1126 (2002).
18. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. & Breaker, R.R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281-286 (2004).
19. Teixeira, A. *et al.* Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature* **432**, 526-530 (2004).
20. Noller, H.F., Hoffarth, V. & Zimniak, L. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**, 1416-1419 (1992).

21. Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905-920 (2000).
22. Nissen, P., Hansen, J., Ban, N., Moore, P.B. & Steitz, T.A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-930 (2000).
23. Khaitovich, P., Tenson, T., Mankin, A.S. & Green, R. Peptidyl transferase activity catalyzed by protein-free 23S ribosomal RNA remains elusive. *RNA* **5**, 605-608 (1999).
24. Mills, D.R., Peterson, R.L. & Spiegelman, S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. USA* **58**, 217-224 (1967).
25. Levisohn, R. & Spiegelman, S. Further extracellular Darwinian experiments with replicating RNA molecules: diverse variants isolated under different selective conditions. *Proc. Natl. Acad. Sci. USA* **63**, 805-811 (1969).
26. Saffhill, R., Schneider-Bernloehr, H., Orgel, L.E. & Spiegelman, S. *In vitro* selection of bacteriophage Q-beta ribonucleic acid variants resistant to ethidium bromide. *J. Mol. Biol.* **51**, 531-539 (1970).
27. Kramer, F.R., Mills, D.R., Cole, P.E., Nishihara, T. & Spiegelman, S. Evolution *in vitro*: sequence and phenotype of a mutant RNA resistant to ethidium bromide. *J. Mol. Biol.* **89**, 719-736 (1974).
28. Caruthers, M.H. Gene synthesis machines: DNA chemistry and its uses. *Science* **230**, 281-285 (1985).
29. Milligan, J.F., Groebe, D.R., Witherell, G.W. & Uhlenbeck, O.C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783-8798 (1987).
30. Baltimore, D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209-1211 (1970).
31. Temin, H.M. & Mizutani, S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211-1213 (1970).
32. Baltimore, D. Discovery of the reverse transcriptase. *FASEB J.* **9**, 1660-1663 (1995).
33. Saiki, R.K. *et al.* Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354 (1985).
34. Mullis, K.B. The unusual origin of the polymerase chain reaction. *Sci. Am.* **262**, 56-61, 64-55 (1990).
35. Jackson, D.A., Symons, R.H. & Berg, P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **69**, 2904-2909 (1972).
36. Cohen, S.N., Chang, A.C., Boyer, H.W. & Helling, R.B. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl. Acad. Sci. USA* **70**, 3240-3244 (1973).
37. Sanger, F. & Coulson, A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 441-448 (1975).

38. Maxam, A.M. & Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**, 560-564 (1977).
39. Joyce, G.F. Amplification, mutation and selection of catalytic RNA. *Gene* **82**, 83-87 (1989).
40. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510 (1990).
41. Robertson, D.L. & Joyce, G.F. Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467-468 (1990).
42. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
43. Beaudry, A.A. & Joyce, G.F. Directed evolution of an RNA enzyme. *Science* **257**, 635-641 (1992).
44. Lehman, N. & Joyce, G.F. Evolution *in vitro* of an RNA enzyme with altered metal dependence. *Nature* **361**, 182-185 (1993).
45. Bartel, D.P. & Szostak, J.W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411-1418 (1993).
46. Eklund, E.H. & Bartel, D.P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231-3238 (1995).
47. Eklund, E.H., Szostak, J.W. & Bartel, D.P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364-370 (1995).
48. Carothers, J.M., Oestreich, S.C., Davis, J.H. & Szostak, J.W. Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.* **126**, 5130-5137 (2004).
49. Sabeti, P.C., Unrau, P.J. & Bartel, D.P. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem. Biol.* **4**, 767-774 (1997).
50. Knight, R. & Yarus, M. Finding specific RNA motifs: function in a zeptomole world? *RNA* **9**, 218-230 (2003).
51. Curtis, E.A. & Bartel, D.P. New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.* **12**, 994-1000 (2005).
52. Mannironi, C., Scerch, C., Fruscoloni, P. & Tocchini-Valentini, G.P. Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif. *RNA* **6**, 520-527 (2000).
53. Held, D.M., Greathouse, S.T., Agrawal, A. & Burke, D.H. Evolutionary landscapes for the acquisition of new ligand recognition by RNA aptamers. *J. Mol. Evol.* **57**, 299-308 (2003).
54. Huang, Z. & Szostak, J.W. Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer. *RNA* **9**, 1456-1463 (2003).
55. Famulok, M. Molecular recognition of amino acids by RNA-aptamers: an L-citrulline binding RNA motif and its evolution into an L-arginine binder. *J. Am. Chem. Soc.* **116**, 1698-1706 (1994).
56. Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, M. Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* **272**, 1343-1347 (1996).

57. Cadwell, R.C. & Joyce, G.F. Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28-33 (1992).
58. Ness, J.E. *et al.* Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* **20**, 1251-1255 (2002).
59. Bittker, J.A., Le, B.V. & Liu, D.R. Nucleic acid evolution and minimization by nonhomologous random recombination. *Nat. Biotechnol.* **20**, 1024-1029 (2002).
60. Wang, Q.S. & Unrau, P.J. Ribozyme motif structure mapped using random recombination and selection. *RNA* **11**, 404-411 (2005).
61. Lorsch, J.R. & Szostak, J.W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31-36 (1994).
62. Hilvert, D. Critical analysis of antibody catalysis. *Annu. Rev. Biochem.* **69**, 751-793 (2000).
63. Prudent, J.R., Uno, T. & Schultz, P.G. Expanding the scope of RNA catalysis. *Science* **264**, 1924-1927 (1994).
64. Conn, M.M., Prudent, J.R. & Schultz, P.G. Porphyrin metalation catalyzed by a small RNA molecule. *J. Am. Chem. Soc.* **118**, 7012-7013 (1996).
65. Kawazoe, N., Teramoto, N., Ichinari, H., Imanishi, Y. & Ito, Y. *In vitro* selection of nonnatural ribozyme-catalyzing porphyrin metalation. *Biomacromolecules* **2**, 681-686 (2001).
66. Agresti, J.J., Kelly, B.T., Jaschke, A. & Griffiths, A.D. Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using *in vitro* compartmentalization. *Proc. Natl. Acad. Sci. USA* **102**, 16170-16175 (2005).
67. Levy, M., Griswold, K.E. & Ellington, A.D. Direct selection of trans-acting ligase ribozymes by *in vitro* compartmentalization. *RNA* **11**, 1555-1562 (2005).
68. Gilbert, W. The RNA world. *Nature* **319**, 618 (1986).
69. Bartel, D.P. & Unrau, P.J. Constructing an RNA world. *Trends Cell Biol.* **9**, M9-M13 (1999).
70. Muller, U.F. Re-creating an RNA world. *Cell Mol. Life Sci.* **63**, 1278-1293 (2006).
71. Yoshioka, W., Ikawa, Y., Jaeger, L., Shiraishi, H. & Inoue, T. Generation of a catalytic module on a self-folding RNA. *RNA* **10**, 1900-1906 (2004).
72. Hager, A.J. & Szostak, J.W. Isolation of novel ribozymes that ligate AMP-activated RNA substrates. *Chem. Biol.* **4**, 607-617 (1997).
73. Robertson, M.P. & Ellington, A.D. *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat. Biotechnol.* **17**, 62-66 (1999).
74. Landweber, L.F. & Pokrovskaya, I.D. Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: the spandrels of RNA evolution. *Proc. Natl. Acad. Sci. USA* **96**, 173-178 (1999).
75. Jaeger, L., Wright, M.C. & Joyce, G.F. A complex ligase ribozyme evolved *in vitro* from a group I ribozyme domain. *Proc. Natl. Acad. Sci. USA* **96**, 14712-14717 (1999).
76. Pan, T. & Uhlenbeck, O.C. *In vitro* selection of RNAs that undergo autolytic cleavage with Pb²⁺. *Biochemistry* **31**, 3887-3895 (1992).
77. Williams, K.P., Ciafre, S. & Tocchini-Valentini, G.P. Selection of novel Mg(2+)-dependent self-cleaving ribozymes. *EMBO J.* **14**, 4551-4557 (1995).

78. Jayasena, V.K. & Gold, L. *In vitro* selection of self-cleaving RNAs with a low pH optimum. *Proc. Natl. Acad. Sci. USA* **94**, 10612-10617 (1997).
79. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643-647 (1995).
80. Wilson, C. & Szostak, J.W. *In vitro* evolution of a self-alkylating ribozyme. *Nature* **374**, 777-782 (1995).
81. Zhang, B. & Cech, T.R. Peptide bond formation by *in vitro* selected ribozymes. *Nature* **390**, 96-100 (1997).
82. Unrau, P.J. & Bartel, D.P. RNA-catalysed nucleotide synthesis. *Nature* **395**, 260-263 (1998).
83. Lau, M.W., Cadieux, K.E. & Unrau, P.J. Isolation of fast purine nucleotide synthase ribozymes. *J. Am. Chem. Soc.* **126**, 15686-15693 (2004).
84. Tarasow, T.M., Tarasow, S.L. & Eaton, B.E. RNA-catalysed carbon-carbon bond formation. *Nature* **389**, 54-57 (1997).
85. Seelig, B. & Jaschke, A. A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **6**, 167-176 (1999).
86. Tsukiji, S., Pattnaik, S.B. & Suga, H. An alcohol dehydrogenase ribozyme. *Nat. Struct. Biol.* **10**, 713-717 (2003).
87. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319-1325 (2001).
88. Eklund, E.H. & Bartel, D.P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **383**, 192 (1996).
89. Hesselberth, J., Robertson, M.P., Jhaveri, S. & Ellington, A.D. *In vitro* selection of nucleic acids for diagnostic applications. *J. Biotechnol.* **74**, 15-25 (2000).
90. Sullenger, B.A. & Gilboa, E. Emerging clinical applications of RNA. *Nature* **418**, 252-258 (2002).
91. Breaker, R.R. Natural and engineered nucleic acids as tools to explore biology. *Nature* **432**, 838-845 (2004).
92. Lan, N., Howrey, R.P., Lee, S.W., Smith, C.A. & Sullenger, B.A. Ribozyme-mediated repair of sickle beta-globin mRNAs in erythrocyte precursors. *Science* **280**, 1593-1596 (1998).
93. Sarver, N. *et al.* Ribozymes as potential anti-HIV-1 therapeutic agents. *Science* **247**, 1222-1225 (1990).
94. Couture, L.A. & Stinchcomb, D.T. Anti-gene therapy: the use of ribozymes to inhibit gene function. *Trends Genet.* **12**, 510-515 (1996).
95. Kawasaki, H. & Taira, K. Discovery of functional genes in the post-genome era by novel RNA-protein hybrid ribozymes. *Nucleic Acids Res. Suppl.*, 133-134 (2001).
96. Kawasaki, H., Onuki, R., Suyama, E. & Taira, K. Identification of genes that function in the TNF-alpha-mediated apoptotic pathway using randomized hybrid ribozyme libraries. *Nat. Biotechnol.* **20**, 376-380 (2002).
97. Kato, Y., Tsunemi, M., Miyagishi, M., Kawasaki, H. & Taira, K. Functional gene discovery using hybrid ribozyme libraries. *Methods Mol. Biol.* **252**, 245-256 (2004).

98. Koizumi, M., Soukup, G.A., Kerr, J.N. & Breaker, R.R. Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat. Struct. Biol.* **6**, 1062-1071 (1999).
99. Seetharaman, S., Zivarts, M., Sudarsan, N. & Breaker, R.R. Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nat. Biotechnol.* **19**, 336-341 (2001).
100. Breaker, R.R. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.* **13**, 31-39 (2002).

CHAPTER 1

New catalytic structures from an existing ribozyme

ABSTRACT

Although protein enzymes with new catalytic activities can arise from existing scaffolds, less is known about the origin of ribozymes with new activities. Furthermore, mechanisms by which new macromolecular folds arise are not well characterized for either protein or RNA. Here we investigate how readily ribozymes with new catalytic activities and folds can arise from an existing ribozyme scaffold. Using *in vitro* selection, 23 distinct kinase ribozymes were isolated from a pool of sequence variants of an aminoacylase parent ribozyme. Analysis of these new kinases demonstrated that ribozymes with new folds and biochemical activities can be found within a short mutational distance of a given ribozyme. However, the probability of finding such ribozymes increases considerably as the mutational distance from the parental ribozyme increases, indicating a need to escape the fold of the parent.

INTRODUCTION

Since 1958, when the structure of myoglobin was determined at atomic level resolution¹, the folds of thousands of protein enzymes have been characterized, providing a wealth of information about protein structure, mechanism, and evolution^{2,3}. With regard to evolution, a number of these studies suggest that protein enzymes with new biochemical activities can arise from existing protein scaffolds^{4,5}. For instance, enzymes in the α/β hydrolase superfamily catalyze a wide range of reactions, and include an acetylcholinesterase, a semialdehyde dehalogenase, and a haloalkane dehalogenase^{5,6}. Similarities in the global folds of these enzymes, local details of their structures, and the positioning and identity of catalytic residues suggest that at least some members of this superfamily were derived from a common ancestor, perhaps following a gene duplication event⁷. In some cases, such a transformation has been directly observed in the laboratory. For example, a single amino acid change is sufficient to change sheep blowfly carboxylesterase into an organophosphorus hydrolase⁸. It has also been noted that protein enzymes with particular activities sometimes catalyze unrelated reactions at low levels, suggesting a possible starting point for the evolution of new catalysts (reviewed in 9). Furthermore, with few exceptions^{10,11}, proteins with new catalytic functions or binding properties generated using *in vitro* evolution have been isolated in the context of existing scaffolds (reviewed in 12).

Comparatively less is known about the origin of ribozymes with new activities, although it has been suggested that starting with a ribozyme scaffold might provide an advantage when isolating new ribozymes by *in vitro* selection^{13,14}. Furthermore, the origin of new macromolecular folds is not well understood for either protein¹⁵ or RNA.

Here, the technique of *in vitro* selection was used to investigate how readily ribozymes with new activities could arise from an previously isolated aminoacylase ribozyme^{16,17}, and whether the folds of such ribozymes were likely to be new as well. Our results indicate that ribozymes with new biochemical activities and folds can be found within a short mutational distance of a given ribozyme, but the probability of finding such ribozymes increases considerably as the mutational distance from the starting ribozyme increases.

RESULTS

Kinase ribozymes from an aminoacylase ribozyme

Our first objective was to determine whether ribozymes with new catalytic activities can be found in the sequence neighborhood of an existing parent ribozyme. As a starting point for our experiments we chose a previously isolated and well characterized self-aminoacylating ribozyme called Isolate 77 (references 16,17). This ribozyme aminoacylates its 3' terminus using adenylated phenylalanine as a substrate (Fig. 1a), and under optimized conditions catalyzes this reaction with a k_{cat}/K_m of $6 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$ and a second order rate enhancement of 6×10^6 -fold (relative to the non-enzymatic hydrolysis rate of adenylated phenylalanine¹⁸). To generate variants of the parent, RNA was transcribed from a partially degenerate DNA template in which 65 of the 90 positions encoding this ribozyme were partially randomized at an average rate of approximately 11% per position (i.e., at each partially randomized position, the parental base was present at a frequency of 0.89, and each of the other 3 bases was present at a frequency of ~ 0.04).

Several factors were considered in the design of this pool. First, we wanted most of the sequences in the pool to be similar to the parent ribozyme. About 95% of the unique sequences in a pool mutagenized at this level will be within 12 mutations of the parent ribozyme, and essentially every possible sequence within 8 mutations of the parent will be present at least once¹⁹. At the same time, we wanted our pool to be diverse enough to contain new ribozymes. At this level of mutagenesis, the pool of 2×10^{15} double-stranded DNA templates contained about 4×10^{14} unique sequences, which would

likely be sufficient if kinases were being selected from a randomized nucleic acid pool^{20,21}. Finally, we wanted our pool to be enriched for ribozymes that are especially close to the parent. In a pool designed in this manner, sequences closer to the parent are present at higher initial copy numbers than are sequences farther away. For our pool, this was true for sequences within 8 mutations of the parent, while sequences farther away were likely present as single copies, if at all. Thus, a ribozyme within 8 mutations of the parent ribozyme was expected to be over-represented in this pool relative to a ribozyme farther from the parent.

An *in vitro* selection protocol was used to isolate rare individuals from this pool that thiophosphorylate themselves in the presence of the GTP analog GTP γ S (Fig. 1b). We were interested in the transition from an aminoacylase ribozyme to kinase ribozymes for several reasons. First, phosphorylation is chemically distinct from aminoacylation, with a transition state that is trigonal bipyramidal rather than tetrahedral. Furthermore, GTP γ S differed from the adenylated phenylalanine substrate used by the parent ribozyme in several important respects, including charge and hydrophobicity, and the nucleotide base of these substrates is also different. By selecting for variants of the parent ribozyme that catalyze a reaction with a different transition state, and recognize a different substrate, we hoped to obtain ribozymes whose folds were different from that of the parent ribozyme as well. Selecting for kinase ribozymes appeared reasonable, because it had been previously shown that RNA can readily catalyze this type of transformation, and that kinase ribozymes can be selected using ATP γ S as a substrate²⁰. Finally, phosphorylation and aminoacylation are both of considerable biological importance.

To isolate kinase ribozymes, pool RNA was incubated with GTP γ S, and molecules that became thiophosphorylated during the incubation were purified on *N*-acryloylaminophenylmercuric chloride (APM) polyacrylamide gels^{22,23}, amplified by RT-PCR, and transcribed to generate RNA for the next round of selection. After four rounds of selection, pool activity was detected, and two additional rounds of selection were performed, using progressively shorter incubation times and a lower concentration of GTP γ S. One hundred and twelve clones from rounds 4, 5, and 6 were sequenced, and were grouped into classes based on several criteria. Initially, a phylogenetic tree was used to identify clusters of closely related or identical sequences. Sequences within 4 mutations of one another were assigned to the same class. Of the 52 classes identified in this way, 41 were based on a single unique sequence, while 11 contained between two and 28 similar or identical sequences. At least one member of each class was then tested for catalytic activity. Twenty-nine of 52 classes contained active members. Further examination revealed several sequences that were more than 4 mutations from one another, but that nevertheless shared more mutations in common than would be expected by chance in a data set of this size. For example, the sequences of 6b-12 and 6b-27 differed at 13 positions, yet they had 8 mutations in common. Five other sequence pairs with 6 or more mutations in common were assigned to the same class, producing a total of 23 unique classes (Fig. 1c).

At 1 mM GTP γ S, initial rates of these ribozymes ranged between 8×10^{-6} and $6 \times 10^{-4} \text{ min}^{-1}$, and the extent of the self-thiophosphorylation reaction was typically between 1 and 60 percent in a 24 hour incubation. k_{cat}/K_m values were determined for 12 of these ribozymes, and ranged between 0.1 and $2 \text{ M}^{-1} \text{ min}^{-1}$. Second order rate enhancements

were estimated by comparing ribozyme k_{cat}/K_m values with the non-enzymatic rate of GTP γ S hydrolysis, and ranged between 3×10^5 and 6×10^6 fold.

Ribozymes catalyzing at least three types of reactions were represented among the 23 kinase classes isolated in this selection. Five out of twelve ribozymes characterized (5-10, 5-15, 5-28, 5b-43, and 6-9) transferred a thiophosphate group from GTP γ S to their 5' hydroxyl. These ribozymes could not be radiolabeled using T4 polynucleotide kinase after they reacted with GTP γ S, unless they were first treated with alkaline phosphatase, suggesting that their 5' hydroxyl becomes covalently modified during catalysis. P1 digests of GTP γ S- γ^{35} S labeled ribozymes generated a labeled product that co-migrated with a GMPS marker, indicating that this modification at the 5' hydroxyl was a thiophosphate (Fig. 1d).

Five out of twelve ribozymes characterized used an internal 2' hydroxyl group as a nucleophile. Analysis of reacted ribozymes on APM polyacrylamide sequencing gels after partial base hydrolysis indicated that these ribozymes modify themselves at internal sites, as exemplified by analysis of 5-16 (Fig. 1e). For this ribozyme, fragments from reacted and unreacted RNA co-migrated until U33, indicating that U33 contained a thiol modification. Ribozymes 5-8 (G82), 5-11 (A79), 7-15 (G17), and 7-16 (C62) also modified themselves at internal sites. In each case, the phosphodiester linkage at the modified position was resistant to base hydrolysis, suggesting that the corresponding 2' hydroxyl group became modified during catalysis. Using an optimized version of one of these 2'-kinase ribozymes (E.A.C. and D.P.B., unpublished data), the nature of this modification was further characterized, by incubating the reacted ribozyme with potential leaving groups in an attempt to regenerate the original substrate, presumably through the

reverse reaction. Reacted ribozyme generated GTP γ S in the presence of GDP, but not in the presence of GMP, G, or buffer alone (Fig. 1f). This suggested that GDP was the leaving group in the forward reaction catalyzed by this ribozyme, and therefore that a single thiophosphate group was transferred to the ribozyme during catalysis.

The two remaining ribozymes characterized (5-2 and 7-14) appeared to promote a two-step reaction in which both the 5' hydroxyl group and an internal site become modified during catalysis. Like the 5'-kinase ribozymes described above, these ribozymes could not be radiolabeled using T4 polynucleotide kinase after they reacted with GTP γ S. However, even after removing the 5' end of reacted ribozymes using RNase H and the appropriate DNA oligonucleotide, they still migrated more slowly in APM polyacrylamide gels than unmodified RNA, suggesting that these reacted ribozymes contain a thiol group at some location other than their 5' terminus. The mechanisms of these two ribozymes were not further characterized.

The kinase reactions differ from the reaction catalyzed by the parent ribozyme in several respects. First, the characterized kinase ribozymes used either 5' or internal 2' hydroxyl groups as nucleophiles, whereas the parent ribozyme uses a hydroxyl group at its 3' terminus as a nucleophile¹⁷. Second, these ribozymes break and form bonds between phosphorous and oxygen, while the parent ribozyme breaks and forms bonds between carbon and oxygen. Third, these ribozymes promote a reaction involving a trigonal bipyramidal transition state, while the parent ribozyme promotes one involving a tetrahedral transition state. Thus, ribozymes with new catalytic activities could be found in the sequence neighborhood of an existing parent ribozyme.

Kinase density increases with mutational distance

Next, we set out to determine how readily such ribozymes arise from an existing ribozyme scaffold. Sixty-seven clones from the starting pool used in this experiment were sequenced, and the mutational distance of each sequence from the parent ribozyme was determined. Comparison to an analogous distribution obtained from the 23 classes of kinase ribozymes indicated that kinases were not uniformly distributed in sequence space with respect to the parent (Fig. 2a). Instead, as the mutational distance from the parent ribozyme increased, the probability of finding a kinase ribozyme increased dramatically, at least over the range of mutational distances examined. In the range between 10 and 16 mutations from the parent, the abundance of sequences in the starting pool decreased as the mutational distance from the parent increased (Fig. 2a). This was expected; for a mutagenesis rate of 11% per position and 65 mutagenized positions, the starting pool should contain approximately 60-fold more sequences 10 mutations from the parent than 16 mutations from the parent. Despite this, the number of kinases isolated was approximately constant over this range, implying that the density of kinase ribozymes (ribozyme density = number of ribozymes that are x mutations from the parent / total number of sequences that are x mutational steps from the parent) among sequences 10 mutations from the parent was considerably lower than that among sequences 16 mutations from the parent.

A small difference between the mutational distances of evolved and unevolved sequences from the parent ribozyme might be expected due to random mutation during the reverse transcription and PCR steps of the selection. To measure the magnitude of this effect, the average distance between the parent ribozyme and inactive clones from

rounds 2, 3, and 4 was determined, and from this data calculated an average mutation rate of 0.05 ± 0.01 (best fit \pm standard error) per molecule per PCR cycle during the selection (**Fig. 2b**). This mutation rate was too low to account for the high number of changes observed among the kinase ribozymes isolated in this selection.

In interpreting these distributions, another important consideration is the extent to which the kinase ribozymes analyzed belong to different classes (i.e., have different sequence requirements and folds). If these kinase ribozymes were all members of the same class, their distribution with respect to the parent might reflect an idiosyncratic feature of this particular class. On the other hand, if each of these ribozymes was an example of a different kinase motif, their distribution would reveal more general principals for mapping function within RNA sequence space. The diverse thiophosphorylation sites of these ribozymes suggested that many different motifs were represented among the kinases isolated in this selection. Sequence analysis of these ribozymes also suggested that most or all belong to different classes. For this analysis, we focused on positions at which the same mutation occurred in different kinase ribozymes. The expected number of mutations shared by two ribozymes in different classes (or two independent sequences) derived from the parent depends on the mutational distance between these ribozymes, and can be calculated using the binomial distribution (see methods). Two ribozyme sequences belonging to the same class, on the other hand, are likely to share more mutations in common than would be expected by chance, because ribozymes typically contain nucleotides whose identities are required for catalytic activity. For the 23 classes of kinase ribozymes we analyzed, the number of sequence pairs with different numbers of mutations in common almost perfectly matched

that expected if each ribozyme belonged to a different class (**Fig. 2c**). Note that different isolates of the same kinase motif can only be identified by this approach if they arise in exactly the same part of the parental sequence. Nonetheless, this assumption appears reasonable given the complexity of the kinase motifs we identified (described below), and the low level (11%) at which the parental sequence was mutagenized.

Another important question was whether the non-uniform distribution of kinase ribozymes in the starting pool was related in some idiosyncratic way to the sequence of the parent ribozyme. The available evidence suggests that this was not the case. In a recent study, aptamers to GMP were isolated from variants of flavin aptamers²⁴, and in a second recent report, GTP aptamers were selected from a pool of variants of an aptamer to ATP²⁵. In both of these studies, the aptamers isolated appear to be farther from the starting sequence than expected based on the composition of the starting pools. This suggests that the distribution of kinase ribozymes we observed reflects something general about the way that RNA functions map onto RNA sequence space.

Our results suggest that ribozymes with new catalytic activities are not uniformly distributed in RNA sequence space with respect to existing ribozymes. Instead, the density of such ribozymes increases substantially as the mutational distance from the starting ribozyme increases. Because studies using RNA-folding algorithms have shown that the probability of finding new RNA secondary structures increases dramatically as the mutational distance from a reference sequence increases²⁶, we suspected that the observed distribution of kinase ribozymes reflected a need to escape the fold of the parent ribozyme. In this view, the folds of most or all of the kinase ribozymes we isolated

should differ from that of the parent ribozyme. To test this idea, we set out to characterize the folds of the parent and two kinase ribozymes.

An escape to new ribozyme folds

A combination of *in vitro* selection and comparative sequence analysis^{27,28} was used to characterize the secondary structures of the parent ribozyme and two kinase ribozymes (called 5-16 and 7-16). For each ribozyme, a pool of 10^{14} to 10^{15} variants was synthesized as previously described, except that nucleotides were mutagenized at an average rate of 20% per position. To isolate active variants of the parent, pool RNA was incubated with adenylated phenylalanine in selection buffer, and reacted molecules were selectively derivatized with a biotin group using sulfo-NHS-biotin²⁹. Biotinylated RNA molecules were bound to soluble streptavidin, isolated on a polyacrylamide gel^{30,31}, amplified by RT-PCR, and transcribed to generate RNA for the next round of selection. Active variants of kinase 5-16 and kinase 7-16 were isolated as previously described.

Once activity was detected in each of these pools, molecules were cloned, sequenced, and individually assayed for catalytic activity. Several approaches were then used to derive secondary structure models for each of these ribozymes. First, each sequence alignment was manually searched for paired regions more highly conserved than would be expected given the level at which the pool was mutagenized. Because this type of analysis does not provide direct evidence for pairing, we also searched for co-varying nucleotides in each of the paired regions in the sequence alignment. To better support proposed helices, especially those in which co-variations were not observed, mutant ribozymes were synthesized in which proposed base pairs were either disrupted,

or restored in a compensatory fashion. Finally, to provide additional support for the proposed secondary structures, minimized versions of each ribozyme were synthesized and tested for catalytic activity.

Our secondary structure model of the parent ribozyme (**Fig. 3a-c**) was similar in some respects to that proposed to be the most populated structure in solution¹⁷ but with several important differences. First, P2 and the distal portion of P4 were only weakly conserved among the sequences isolated in this selection (**Fig. 3a**), and could be deleted from the ribozyme with only a 5-fold reduction in catalytic rate (**Supplementary Fig. 1a**). Second, the catalytically active structure of P3 was different from that previously proposed for this part of the sequence. Although not consistent with lead and S1 nuclease probing data¹⁷, this alternative version of P3 was well supported by both comparative sequence analysis (**Fig. 3a**), and site-directed mutagenesis experiments (**Supplementary Fig. 1b**). The initial rate of a 50-nt version of the parent ribozyme in which P2 and P4b were deleted, and P1 and P4a were capped with CUUCGG tetraloops³², was 4-fold slower than that of the full-length ribozyme (**Fig. 3c** and **Supplementary Fig. 1a**). An even smaller 38-nt variant of the parent, in which P1 was also deleted, was 30-fold slower than that of the full-length ribozyme (**Supplementary Fig. 1a**), indicating that the elements most essential for catalysis are located in P3 and P4a.

The secondary structure model of 5-16 (**Fig. 4a-c**) consisted of a bulged hairpin (P2 and P3) folded back on itself in a pseudoknot (P4), flanked by two less conserved helices (P1 and P5). Comparative sequence analysis also revealed 11 highly conserved unpaired nucleotides, all within the P2-P4 domain of the ribozyme, whose identities were likely important for catalytic activity (**Fig. 4a**). P1 and P5 were generally variable in

sequence and length, and they did not always form in exactly the same place in different isolates, unlike helices in the catalytic core of the ribozyme, which were in a constant position (**Fig. 4a**). These helices could be deleted from the ribozyme without any reduction in catalytic rate (**Fig. 4c** and **Supplementary Fig. 2a**). Disrupting pairings in these helices, however, reduced the catalytic rate of this ribozyme up to 10-fold, and activity could be restored with compensatory mutations (**Supplementary Fig. 2b**). These helices appear to prevent nucleotides in flanking regions from interfering with those in the P2-P4 domain of the ribozyme. Deleting them completely is tolerated, but disrupting pairings within them is not (**Supplementary Fig. 2**).

The secondary structure model of kinase 7-16 (**Fig. 5a-c**) was more complex than that of either the parent or 5-16, and consisted of a stem (P1) and three hairpins (P2, P4, and P5), two of which were linked by a loop-loop interaction (P3). Except for the distal portion of P4, each of the proposed helices in this structure were highly conserved. Interspersed between these helical elements were 16 additional invariant nucleotides, 10 of which were adenosines, and several of which were clustered, hinting at possible A-minor interactions in the tertiary structure of this ribozyme³³. In each of these isolates, P4 could be extended (extension only shown for 7-16 in **Fig. 5a**), although individual base pairs were generally not conserved. Consistent with this lack of conservation, the catalytic rate of a minimized version of 7-16 in which P4 was truncated was only 3-fold lower than that of the full-length ribozyme (**Fig. 5c** and **Supplementary Fig. 3a**). Comparative analysis of 7-16 variants was problematic, because only about a third of the isolates had the same fold. Such an outcome has been observed previously^{34,35}, and its occurrence depends on several factors, including the degree to which the pool is

mutagenized, the complexity of the parental ribozyme, and the relative catalytic rates of variants with the parental fold and those with other folds. Because of this complication, site-directed mutagenesis was used to further test the model. The results provided strong support for each of the essential helices in the ribozyme (**Supplementary Fig. 3b**).

The secondary structure model of kinase 5-16 was completely different from that of the parent ribozyme, with no base pairs in common (**Fig. 6**). The secondary structure of kinase 7-16 was mostly different from that of the parent as well, with 23 new base pairs and only 9 retained ones (**Fig. 6**). Furthermore, none of the 24 base pairs in the minimized version of 7-16 (**Fig. 5c**) were present in the minimized version of the parent (**Fig. 3c**). Although the folds of the other 21 classes of kinase ribozymes isolated in this experiment are not known, none of these ribozymes have the potential to form all of the helices in the secondary structure of the parent, and many contain disruptions in each of the parental helices. Thus, the folds of the 23 kinase ribozymes were largely different from that of the parental aminoacylase ribozyme.

DISCUSSION

Our results indicate that ribozymes with new catalytic activities can be found within a short mutational distance of a given ribozyme, but that the probability of finding such ribozymes increases dramatically as the mutational distance from the starting ribozyme increases. This appears to reflect a need to escape the fold of the starting ribozyme, in that the folds of the kinase ribozymes we isolated were distinct from that of the aminoacylase parent ribozyme. These results give a better sense of how readily ribozymes with new biochemical activities can arise from existing ribozymes. On the average, 14 mutational changes were needed to convert a 90-nucleotide aminoacylase ribozyme into a kinase ribozyme. They also suggest that, for a ribozyme with a new catalytic activity to arise, it is likely that a new RNA fold must also arise. This also appears to be true for aptamers^{24,25,36}, except in cases where variants of an existing aptamer are selected to bind a very similar target molecule^{37,38}. In contrast, protein enzymes with very different substrates and activities are sometimes found in the context of the same scaffold, such as the α/β hydrolase fold^{5,6}. Finally, these results provide experimental support for the emerging idea that RNA evolution is not continuous^{39,40}. Instead, a relatively small number of mutational changes in an existing ribozyme can generate new ribozymes with completely different catalytic activities and folds, thereby providing a mechanism for the emergence of new folds and activities from pre-existing ribozymes.

Figure 1 Aminoacylase and kinase ribozymes. (a) Reaction catalyzed by the parent aminoacylase ribozyme. (b) Reaction catalyzed by a 5'-kinase ribozyme. Note that our selection scheme did not require ribozymes to specifically thiophosphorylate their 5' hydroxyl group, and about half of the ribozymes characterized modified an internal 2' hydroxyl group. (c) Sequence alignment of 23 kinase ribozymes isolated in this selection. Boxes indicate positions that differ from the aminoacylase parent ribozyme. Colored bars represent parts of the sequence that form base pairs in the parent. Sequences are shown in order of mutational distance from the parent. (d) Ribozyme 6-9 is a 5' kinase. Ribozyme 6-9 was reacted with $\text{GTP}\gamma\text{S}-\gamma^{35}\text{S}$ and digested with P1 nuclease. The resulting 5' nucleotide monophosphates were analyzed by PEI-cellulose thin layer chromatography. (e) Ribozyme 5-16 is an internal kinase. Reacted 5-16 was radiolabeled at its 5' terminus, partially hydrolyzed with base, and analyzed on an APM polyacrylamide sequencing gel. (f) Ribozyme 5-16 transfers a thiophosphate from $\text{GTP}\gamma\text{S}$ to itself. An optimized version of 5-16 (E.A.C. and D.P.B., unpublished data) was reacted with $\text{GTP}\gamma\text{S}-\gamma^{35}\text{S}$, purified, and incubated in the presence of different leaving groups that might be produced in the forward reaction of this ribozyme with $\text{GTP}\gamma\text{S}$. Reaction aliquots were analyzed by PEI-cellulose thin layer chromatography.

Figure 2 Distribution of kinase ribozymes with respect to the parent ribozyme. (a) The distances between 67 sequences from the starting pool and the parent ribozyme (blue bars; average distance = 7.5) compared to the distances between 23 kinase ribozymes and the parent ribozyme (red bars; average distance = 14). (b) The average mutational distance of inactive sequences from rounds 0, 1, 2, and 3 (closed circles) and kinase

ribozymes (open square) plotted as a function of the number of PCR cycles these sequences were amplified. The slope of this line is 0.05 ± 0.01 (best fit \pm standard error).

(c) Independence of kinase ribozyme classes. The number of mutations in common shared by each possible pair of kinase ribozymes (blue bars) is compared to the distribution expected if these sequences were independent (red bars).

Figure 3 Secondary structure of the aminoacylase parent ribozyme. (a) Sequence alignment of active variants of the parent ribozyme. Colored bars represent paired regions, gray bars indicate less conserved paired regions that are not required for catalytic activity, and brown blocks indicate co-varying nucleotides. Boxes indicate positions that differ from the parental sequence. All variants are at least as active as the parent ribozyme, and are arranged in approximate order of activity in $100 \mu\text{M}$ adenylyated phenylalanine. Note that 12 nucleotides at the 5' end and 13 nucleotides at the 3' end of this pool were not mutagenized. (b) Secondary structure of the parent ribozyme. Thin black dashes represent proposed base pairs, thick black dashes indicate base pairs supported by co-variation, thick violet dashes indicate base pairs supported by site-directed mutagenesis experiments, and thick blue dashes indicate base pairs supported by both co-variation and site-directed mutagenesis. Unpaired nucleotides shown in red were conserved in 29 or more out of 31 isolates. The site of self-aminoacylation is shown in orange. (c) Minimal construct, which reacts 4-fold slower than the full-length parent ribozyme.

Figure 4 Secondary structure of a kinase ribozyme derived from the aminoacylase parent ribozyme. (a) Sequence alignment of active variants of kinase 5-16. Drawing conventions are as in **Fig. 3**. All variants are at least as active as 5-16, and are arranged in approximate order of activity at 100 μ M GTP γ S. Primer binding sites flanking each sequence not shown in the alignment. (b) Secondary structure of 5-16. Single stranded nucleotides shown in red were conserved in at least 21 out of 22 isolates. The site of self-thiophosphorylation is shown in orange. (c) Deletion construct with activity equal to that of the full-length ribozyme.

Figure 5 Secondary structure of a second kinase ribozyme derived from the aminoacylase parent ribozyme. (a) Sequence alignment of active variants of kinase 7-16. Drawing conventions are as in **Fig. 3**. All variants are at least as active as 7-16, and are arranged in approximate order of activity at 100 μ M GTP γ S. Primer binding sites flanking each sequence not shown in the alignment. (b) Secondary structure of 7-16. Single stranded nucleotides shown in red were conserved in all 14 isolates. The site of self-thiophosphorylation is shown in orange. (c) Deletion construct with activity 3-fold slower than that of the full-length ribozyme.

Figure 6 Comparison of the secondary structure of the parent ribozyme and the secondary structures of two kinase ribozymes isolated in this selection. Colored segments represent paired regions in the parental structure, and are mapped onto the structures of 5-16 and 7-16.

Supplementary Figure 1 Testing the secondary structure model of the parent ribozyme by site-directed mutagenesis. **(a)** Identification of the minimized catalytic core of the parent ribozyme by deletion analysis. Initial rates were measured in 100 μ M adenylylated phenylalanine. Time points were taken within the first minute of the reaction, and analyzed using either a streptavidin gel-shift assay or a TLC-based assay. **(b)** Testing proposed base pairs with compensatory mutations. Initial rates were measured as described above. Mutations in P3a and P4a were tested in the context of the full length ribozyme, while mutations in P3a and P4a were tested in the context of the ribozyme's minimized catalytic core. Numbering of base pairs is according to the alignment in **Fig. 3**.

Supplementary Figure 2 Testing the secondary structure model of kinase ribozyme 5-16 by site-directed mutagenesis. **(a)** Identification of the minimized catalytic core of kinase 5-16 by deletion analysis. Initial rates were measured in 1 mM GTP γ S. Time points were taken within the first 60 minutes of the reaction, and analyzed by APM polyacrylamide gel electrophoresis. **(b)** Testing proposed base pairs with compensatory mutations. Initial rates were measured as described. Mutations in P1 and P5 were tested in the context of the full length ribozyme, while mutations in P2 were tested in the context of the ribozyme's minimized catalytic core. Numbering of base pairs according to alignment in **Fig. 4**.

Supplementary Figure 3 Testing the secondary structure model of kinase ribozyme 7-16 by site-directed mutagenesis. **(a)** Identification of the minimized catalytic core of kinase 7-16 by deletion analysis. Initial rates were typically measured at several different

GTP γ S concentrations between 10 μ M and 1 mM. Time points were taken within the first 70 minutes of the reaction, and analyzed by APM polyacrylamide gel electrophoresis. (b) Testing proposed base pairs with compensatory mutations. Initial rates were measured as described. All mutations were tested in the context of the ribozyme's minimized catalytic core. Numbering of base pairs according to alignment in **Fig. 5.**

Figure 1

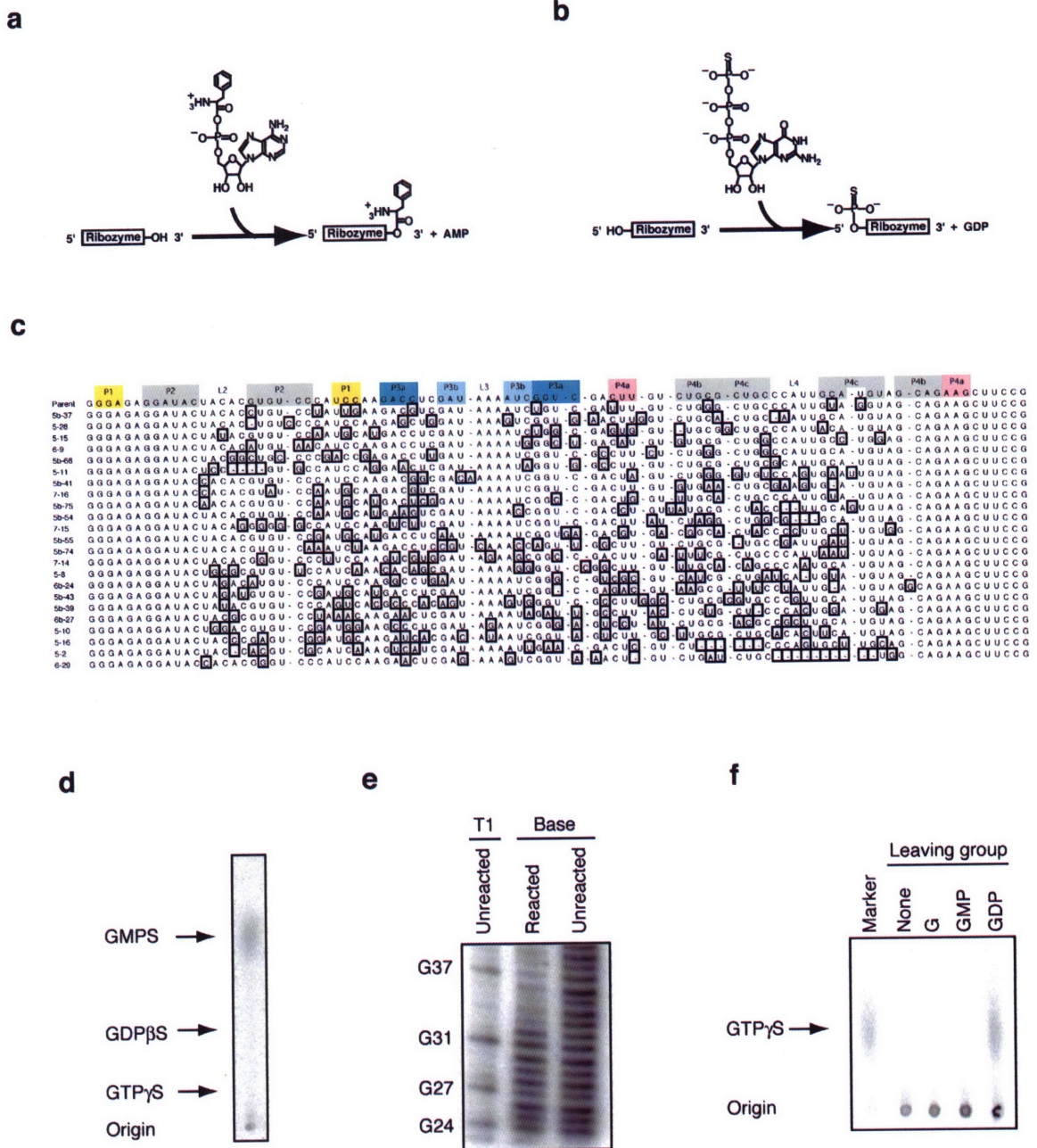


Figure 2

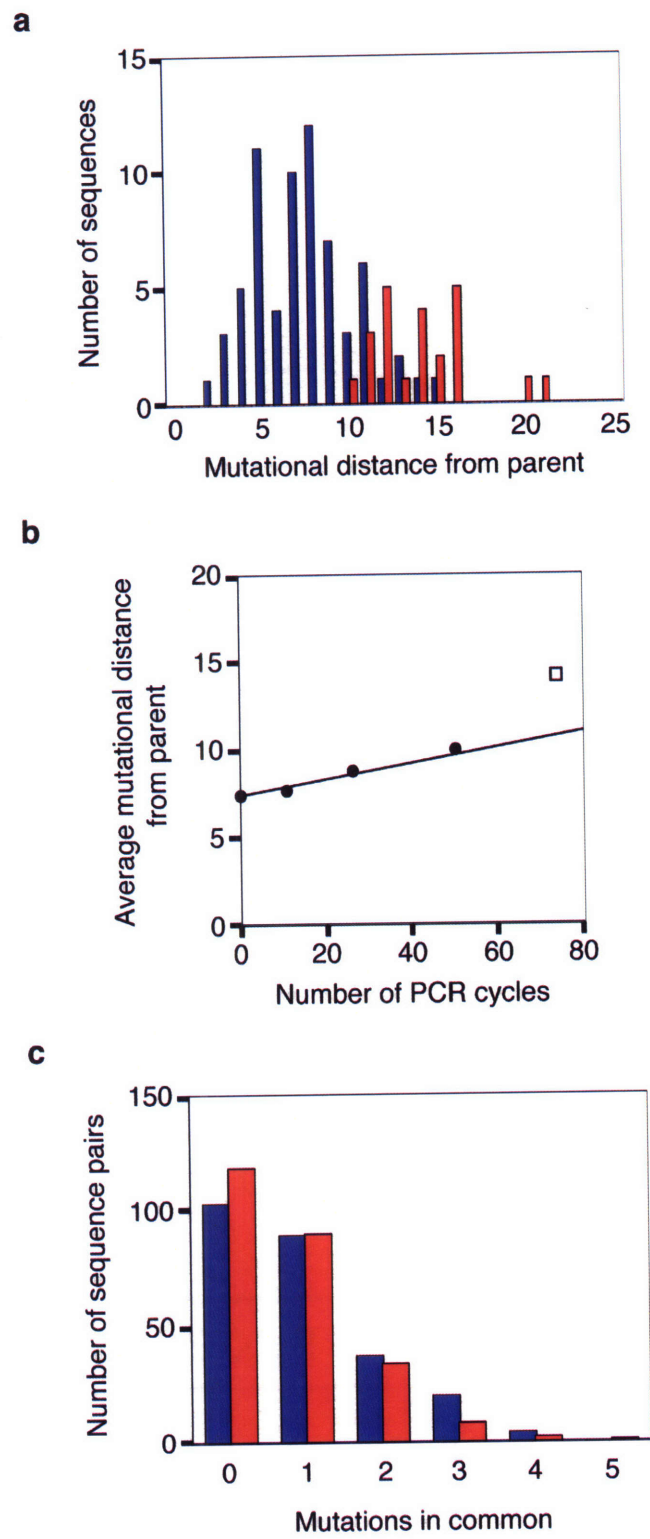
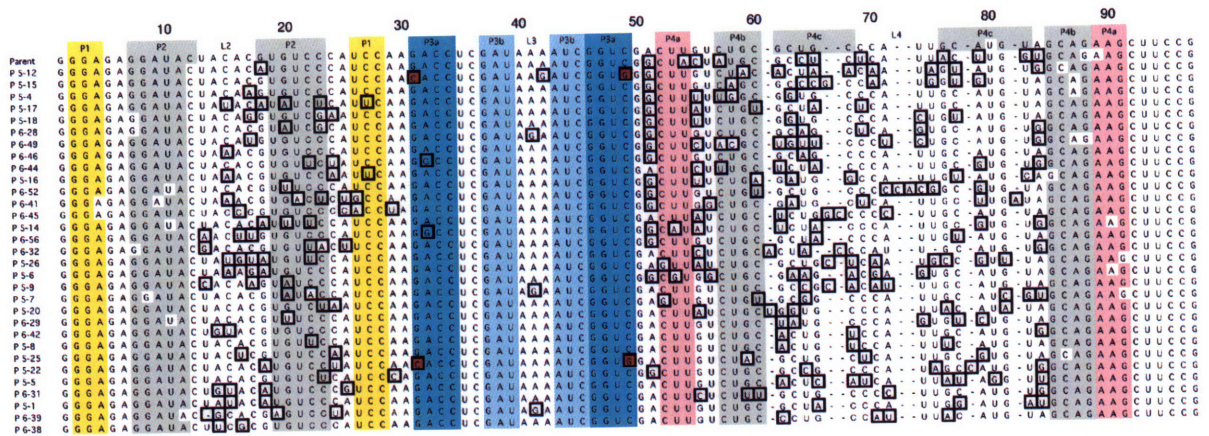
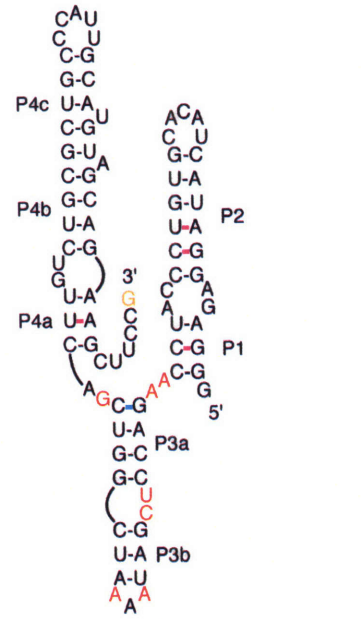


Figure 3

a



b



c

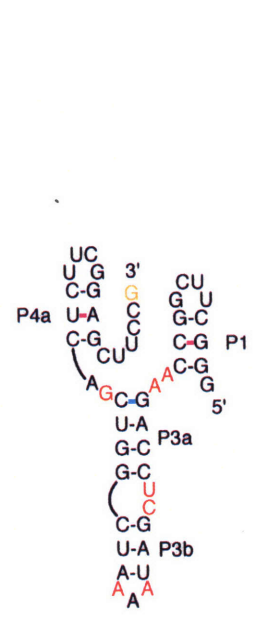
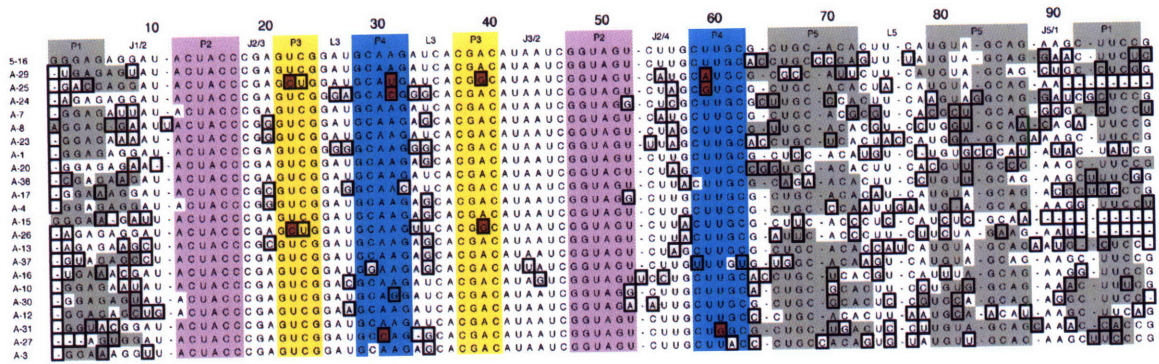
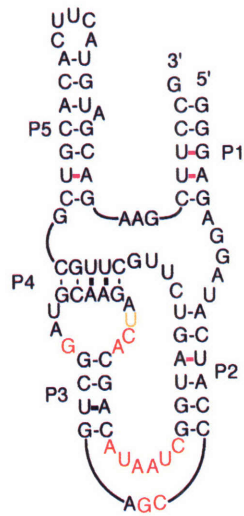


Figure 4

a



b



c

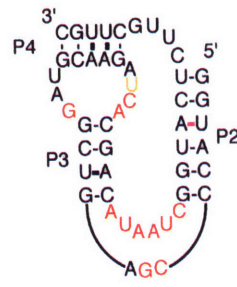


Figure 5

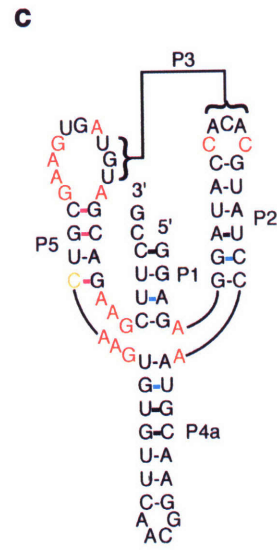
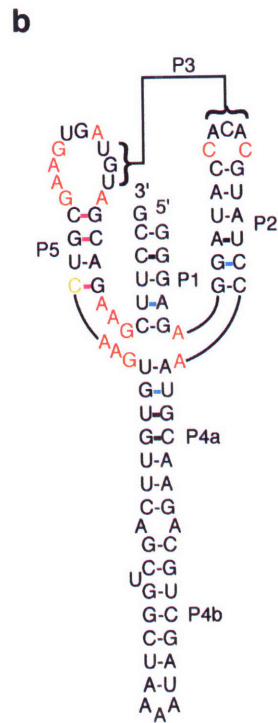
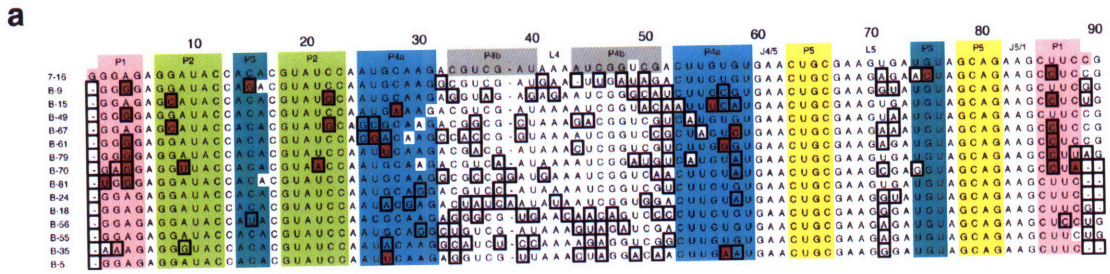
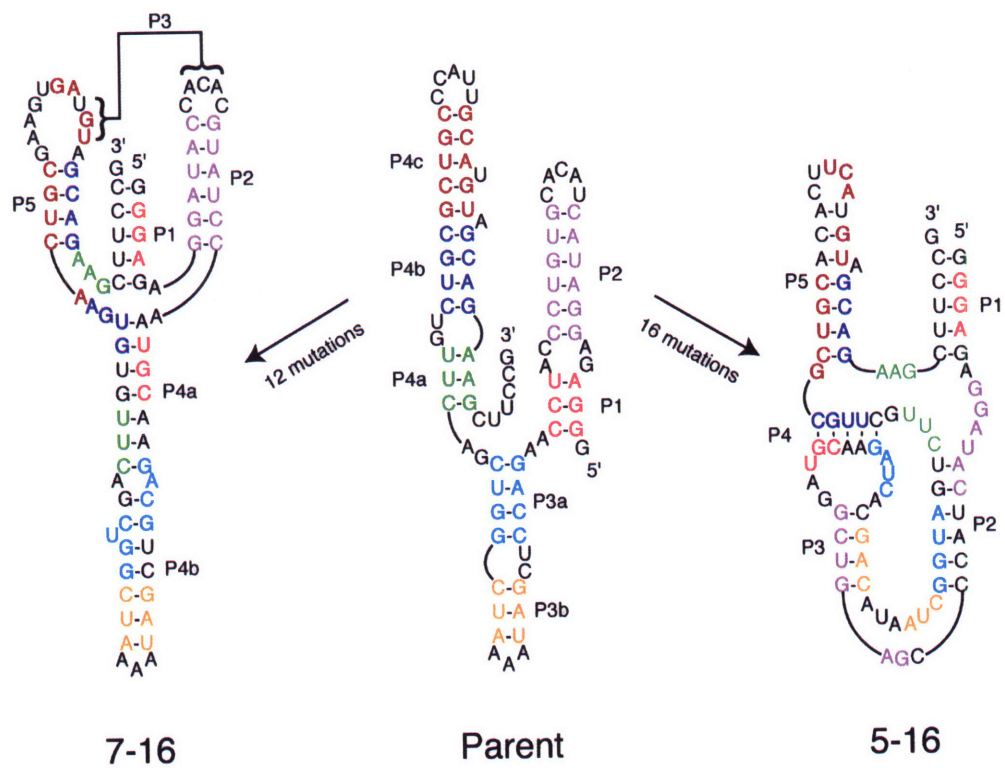

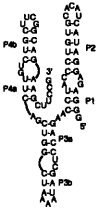
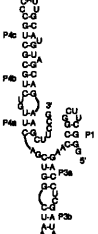
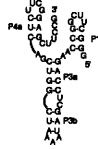
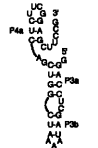


Figure 6



Supplementary Figure 1

a

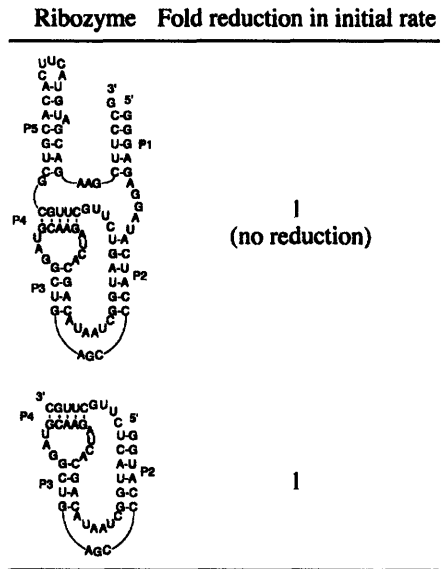
Ribozyme	Fold reduction in initial rate
	1 (no reduction)
	5
	5
	4
	30

b

Helix	Base pair	Sequence	Fold reduction in initial rate
P1	3-27	G-C	1 (no reduction)
		C C	6
		G G	5
		C-G	3
P2	8-22 9-21	G-C	1
		A-U	7
		CC	5
		GG	1
		AA	1
P3	31-49	G-C	> 10
		CC	> 10
		GG	2
		C-G	1
P4	53-90	U-A	40
		UU	> 160
		AA	1
		A-U	1

Supplementary Figure 2

a

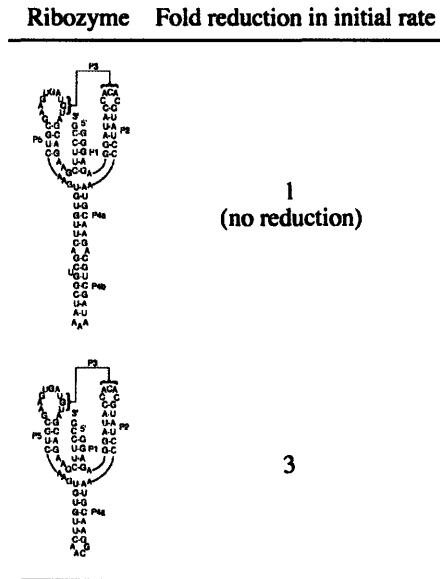


b

Helix	Base pair	Sequence	Fold reduction in initial rate
P1	3-95 4-94	G-U A-U	1 (no reduction)
		GG AG	5
		CU CU	13
		C-G C-G	2
		U-A	1
P2	14-50	UC	> 20
		GA	> 20
		G-C	1
P5	66-86	U-A	1
		UU	1.4
		AA	10
		A-U	1

Supplementary Figure 3

a



b

Helix	Base pair	Sequence	Fold reduction in initial rate
P1	4-86	A-U	1 (no reduction)
		U U	> 50
		AA	> 50
		U-A	4
P2	8-22	G-C	1
		CC	> 50
		GG	7
		C-G	1
P3	15-75	C-G	1
		GG	> 50
		CC	> 50
		G-C	1
P4	26-58	U-G	1
		GG	5
		UC	2
		G-C	1
P5	63-81	C-G	1
		U-G	3
		CA	10
		U-A	1.4
P5	65-79 66-78	G-C	1
		C-G	1
		AC	> 50
		U-G	> 50
		G-U	> 50
		CA	> 50
		U-A	1
		A-U	1

METHODS

Pool synthesis

An Expedite Nucleic Acid Synthesis System was used to synthesize partially randomized DNA oligonucleotides used to construct the four pools used in these experiments. The first of these oligonucleotides encoded a partially degenerate version of the parent ribozyme, and had the sequence **CGGAAGCTTCTGCTACATGCAATGGGCAGCGC**
AGACAAGTCGACCGATTTTATCGAGGTCTTGGATGGGACACGTGTAGTA
TCCTCTCCCTATA (positions shown in bold type mutagenized at a rate of 10%; T7 promoter sequence underlined). The second had the same sequence as the first, but positions were mutagenized at a level of 20% rather than 10%. Following deprotection and gel purification, T7 promoters were added to these oligonucleotides by large scale PCR⁴¹ using the primers AACTAATACGACTCACTATAGGGAGAGG and CGGAAGCTTCTGC. The resulting double-stranded DNA templates were transcribed using T7 RNA polymerase⁴². The third encoded a partially degenerate version of kinase ribozyme 5-16, flanked on each side by a 4-nt spacer and an 18-nt primer-binding site. This oligonucleotide had the sequence **TTACGGGATTCGTCATTCTTTTCGGAAGC**
TTCTGCTACATGAAGTGTGCAGCGCAAGCAAGACTACCGATTATGTCGT
GATCTTGCATCCGACTCGGGTAGTATCCTCTCCCTTTTACTGCAACCGCAT
AGACC (positions shown in bold type mutagenized at a rate of 20%). Following deprotection and gel purification, a T7 promoter was added to this oligonucleotide by mutually primed synthesis using the primer AACTAATACGACTCACTATAGGTCT

ATGCGGTTGCAGT, and the resulting double-stranded DNA template was transcribed using T7 RNA polymerase⁴². This oligonucleotide was amplified by RT-PCR using the primers AACTAATACGACTCACTATAGGTCTATGCGGTTGCAGT and TTACGGGATTCGTCATTC. The fourth encoded a partially degenerate version of kinase ribozyme 7-16, flanked on each side by a 4-nt spacer and an 18-nt primer-binding site. This oligonucleotide had the sequence CAGCTAGCTTGCATAACGTTTTTCGGA **AGCTTCTGCTACATCACTTCGCAGTTCACACAAGTCGACCGATTTTATCG** **ACGTCTTGCATTGGATACGTGTGGTATCCTCTCCCTTTTTCTTACGCGATG** **CCGACC** (positions shown in bold type mutagenized at a rate of 20%). Following deprotection and gel purification, a T7 promoter was added to this oligonucleotide by mutually primed synthesis using the primer AACTAATACGACTCACTATAGGTCGG CATCGCGTAAGA, and the resulting double-stranded DNA template was transcribed using T7 RNA polymerase⁴². This oligonucleotide was amplified by RT-PCR using the primers AACTAATACGACTCACTATAGGTCGGCATCGCGTAAGA and CAGCTAGCTTGCATAACG.

Isolation of kinase ribozymes by *in vitro* selection

To dephosphorylate pool RNA, it was incubated in the presence of calf alkaline phosphatase (Roche) at 50°C for 15 minutes (10 μ M RNA and 0.2 units/ μ L phosphatase). After extracting with phenol and chloroform, RNA was purified by polyacrylamide gel electrophoresis, eluted, and precipitated with ethanol.

Pool RNA, including a trace amount of body-labeled pool RNA, was heated at 65°C for 5 minutes in water, and cooled at room temperature for 5 minutes. The reaction

was initiated by adding an equal volume of a mix containing selection buffer and GTP γ S (Sigma). Final concentrations were 1 μ M RNA, either 1 mM (rounds 1–4) or 100 μ M (rounds 5–6) GTP γ S, 10 mM MgCl₂, 5 mM CaCl₂, 200 mM KCl, and 100 mM HEPES buffer, pH 7.25, in a volume of up to 15 mL. After incubating at room temperature for either 24 hours (rounds 1–4), 1 hour (round 5), or 20 minutes (round 6), the reaction was split into 2 tubes, and a 90-nt radiolabeled RNA marker with a 5' thiophosphate was added to one of the tubes. Following ethanol precipitation, the contents of each tube were resuspended in gel loading buffer (8 M urea, 25 mM EDTA, xylene cyanol, and bromophenol blue).

RNA was then heated at 80°C for 5 minutes, and loaded on a 6% polyacrylamide, 20–40 μ M APM stacking gel. Electrophoresis was at 22 watts for 45–90 minutes. RNA was visualized by autoradiography, and molecules that co-migrated with the marker oligonucleotide were excised from the gel with a sterile razor blade. RNA was eluted from the gel overnight, on a rotator, at 4°C, in 20 mM DTT and 300 mM NaCl. In some rounds of selection, two consecutive gel purifications were performed.

Following ethanol precipitation in the presence of carrier (either tRNA or reverse transcription primer), RNA was resuspended in water, heated at 70°C for 3 minutes, and cooled at room temperature for 3 minutes. After adding buffer, and heating at 48°C for 2 minutes, reverse transcriptase was added to the reaction (1900 units SuperScript II RNase H⁻ reverse transcriptase from Invitrogen and 2.5 μ M reverse transcription primer in a volume of 100 μ L). After a 1 hour incubation at 48°C, RNA was hydrolyzed at 90°C for 15 minutes in the presence of 20 mM Tris base, 4 mM EDTA, and 0.3 M KOH. The pH of the solution was then adjusted to 8.5 by adding HCl. DNA was amplified by PCR

without further purification in a volume at least 15 times greater than that of the reverse transcription. At various times, aliquots were removed and analyzed by agarose gel electrophoresis. Once a PCR product could be detected, DNA was precipitated with ethanol and transcribed using T7 RNA polymerase.

In subsequent selections to characterize 5-16 and 7-16, pool RNA was not dephosphorylated, because both the phosphorylated and dephosphorylated forms of these ribozymes were active. In addition, to prevent potential pairing between primer binding sites and the core of each ribozyme, blocking oligonucleotides (each at a final concentration of 1 μ M) complementary to the primer binding sites of these pools were present during incubation with GTP γ S. The oligonucleotides ACTGCAACCGCATAGACC and TTACGGGATTCGTCATTC were used to block the primer binding sites in the 5-16 pool, and the oligonucleotides TCTTACGCGATGCCGACC and CAGCTAGCTTG CATAACG were used to block the primer binding sites in the 7-16 pool.

Isolation of aminoacylase ribozymes by *in vitro* selection

Adenylated phenylalanine was synthesized as described¹⁶. Following methanol elution from a Waters Sep-Pak C18 cartridge, aliquots (typically 25 μ L) were concentrated to dryness in a Savant speed vacuum using a low drying rate. Evaporation typically took 45–90 minutes, and analysis by reversed-phase HPLC¹⁶ indicated that preparations were typically 75% pure (contaminated with phenylalanine and 5' AMP that presumably formed during purification). Dry aliquots were stored at –80°C, and typically used within 2–3 weeks.

Following transcription and gel purification, pool RNA, including a trace amount of body-labeled pool RNA, was heated at 65°C for 5 minutes in water, and cooled at room temperature for 5 minutes. The reaction was initiated by first adding selection buffer, and then adding a solution of adenylated phenylalanine. Because adenylated phenylalanine is extremely susceptible to hydrolysis¹⁸, aliquots were resuspended in water immediately before use. Final conditions were 1 μM RNA, either 1 mM (rounds 1–4) or 100 μM (round 5) phenylalanine-AMP, 10 mM MgCl₂, 5 mM CaCl₂, 200 mM KCl, and 100 mM HEPES buffer, pH 7.25, in a volume of up to 5 mL. Note that this buffer differs from that used in the original selection to isolate this ribozyme in several ways: it contains 200 mM KCl rather than 20 mM KOAc, it does not contain NaCl, and it does not contain trace amounts of CuSO₄, FeCl₃, MnCl₂, and ZnCl₂¹⁶. After incubating at room temperature for either 15 minutes (round 1), or 1–3 minutes (rounds 2–5), the reaction was split into 2 tubes, and a radiolabeled positive control (a 10-nt RNA oligonucleotide from Dharmacon Research containing a primary amine at its 3' terminus) was added to one of the tubes.

Following ethanol precipitation in the presence of 300 mM sodium acetate (pH 5), aminoacylated RNA was biotinylated (280 mM HEPES buffer, pH 7.6, and 50 mM sulfo-NHS-biotin from Pierce in a volume of 100 μL). After a 30 minute incubation at room temperature, RNA was ethanol precipitated in the presence of 300 mM sodium acetate (pH 5), and the contents of each tube were resuspended in 90 μL gel-loading buffer (8 M urea, 25 mM EDTA, xylene cyanol, and bromophenol blue). 30 μL of a 50 μM streptavidin solution was then added to each tube.

RNA was loaded without heating on a 6%, 0.75 mm thick polyacrylamide gel, and electrophoresis was at 22 watts for 35 minutes. Note that because aminoacylated RNA is considerably more stable following derivatization with sulfo-NHS-biotin, it was not necessary to perform electrophoresis at low pH, as is typical when working with aminoacylated RNA. RNA was visualized by autoradiography, and molecules that co-migrated with the positive control were excised from the gel with a sterile razor blade. RNA was eluted from the gel overnight, on a rotator, at 4°C, in 50 mM Tris-HCl, pH 8.5 and 5 mM EDTA. Reverse transcription, PCR, and transcription were as described for the kinase selection.

Ribozyme activity assays

In a typical assay, a trace amount of body labeled RNA was heated and cooled, and selection buffer and substrate was added to initiate the reaction. Kinase reactions were monitored on APM polyacrylamide gels. Aminoacylase reactions were analyzed using either a streptavidin gel shift assay or a TLC assay similar to that previously described¹⁷. Initial rates were estimated from the earliest extent of reaction that could be measured using the formula $R = 1 - e^{-kt}$, where R represents the fraction of ribozyme reacted, t represents time, and k represents the ribozyme rate. For aminoacylase reactions, time points were typically taken within the first minute of the reaction, so that hydrolysis of adenylated phenylalanine was not significant. k_{cat}/K_m values were determined by plotting initial rate as a function of substrate concentration and determining the slope of the linear part of this curve. Full-length ribozymes were not separated from n+1 transcripts, which

were presumably inactive. Inclusion of these inactive transcripts is estimated to have decreased the observed reaction rates by 2- to 3-fold.

Characterization of kinase ribozyme activities

To determine whether the 5' hydroxyl group of a reacted ribozyme was modified, it was incubated at 37°C for 30 minutes in the presence of 10 units of T4 polynucleotide kinase (New England BioLabs) and ATP- γ ³²P and analyzed by polyacrylamide gel electrophoresis. To further characterize the ribozymes that could not be 5' labeled with T4 polynucleotide kinase, they were reacted with GTP γ S- γ ³⁵S, digested with P1 nuclease, and analyzed by PEI-cellulose thin layer chromatography as previously described²⁰. RNase H digestions were performed at 37°C for 2.5 hours in the presence of body-labeled reacted ribozyme, a 13 nucleotide DNA oligonucleotide complementary to the 5' end of these ribozymes (25 μ M final concentration), reaction buffer, and 5 units RNase H (New England BioLabs). To map the reaction sites of the ribozymes that modify themselves at internal sites, reacted RNA was 5' labeled, and following partial base hydrolysis in 100 mM sodium citrate (80°C for 75 seconds), aliquots were neutralized with HCl and analyzed on either polyacrylamide or APM polyacrylamide sequencing gels. To observe the reverse reaction of 5-16, an optimized version of this ribozyme (described in Chapter 2 of this thesis) was labeled with GTP γ S- γ ³⁵S and purified by APM polyacrylamide gel electrophoresis. Reacted ribozyme was incubated in the presence of selection buffer and 1 mM leaving group, and time points were analyzed by TLC as described²⁰. For the experiment shown in **Fig. 1f**, a 92-hour time point was analyzed.

Calculations

Standard formulas were used to calculate the number of unique sequences in our starting pool¹⁹, the number of sequences at specified mutational distances from the parent ribozyme¹⁹, and the probability of a base pair being conserved in a certain number of isolates from a degenerate pool⁴³. The probability that two kinase ribozymes in different classes derived from the parent will contain the same mutation at a particular position by chance (P) is approximately equal to the probability that the ribozymes will contain the same point mutation at this position ($P_m=M^2/3$, where M is the frequency of point mutations in the kinase ribozymes isolated in this selection), plus the probability that the ribozymes will both contain a deletion at this position ($P_d=D^2$, where D is the frequency of deletions in the kinase ribozymes isolated in this selection). Because the frequency of insertions in these kinase ribozymes was low (F=0.007), they make a negligible contribution to P, and therefore were not considered in this calculation. For the 23 kinase ribozymes analyzed, M=0.18 and D=0.024, so that P=0.011. For 65 mutagenized positions, then, the probability that two unrelated kinase ribozymes will contain x mutations in common can be calculated using the binomial distribution.

ACKNOWLEDGMENTS

We thank M. Lawrence for comments on this manuscript, and members of the lab for helpful discussions. This work was supported by a grant from the National Institutes of Health.

REFERENCES

1. Kendrew, J.C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666 (1958).
2. Patthy, L. *Protein Evolution*, (Blackwell Science, Oxford, 1999).
3. Branden, C. & Tooze, J. *Introduction to Protein Structure*, (Garland Publishing, New York, New York, 1999).
4. Babbitt, P.C. & Gerlt, J.A. Understanding enzyme superfamilies. Chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **272**, 30591-30594 (1997).
5. Ollis, D.L. *et al.* The alpha/beta hydrolase fold. *Protein Eng.* **5**, 197-211 (1992).
6. Holmquist, M. Alpha/beta-hydrolase fold enzymes: structures, functions and mechanisms. *Curr. Protein Pept. Sci.* **1**, 209-235 (2000).
7. Ohno, S. *Evolution By Gene Duplication*, (Springer-Verlag, New York, New York, 1970).
8. Newcomb, R.D. *et al.* A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. *Proc. Natl. Acad. Sci. USA* **94**, 7464-7468 (1997).
9. O'Brien, P.J. & Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91-R105 (1999).
10. Keefe, A.D. & Szostak, J.W. Functional proteins from a random-sequence library. *Nature* **410**, 715-718 (2001).
11. Raffler, N.A., Schneider-Mergener, J. & Famulok, M. A novel class of small functional peptides that bind and inhibit human alpha-thrombin isolated by mRNA display. *Chem. Biol.* **10**, 69-79 (2003).
12. Skerra, A. Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* **13**, 167-187 (2000).
13. Jaeger, L., Wright, M.C. & Joyce, G.F. A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc. Natl. Acad. Sci. USA* **96**, 14712-14717 (1999).
14. Yoshioka, W., Ikawa, Y., Jaeger, L., Shiraishi, H. & Inoue, T. Generation of a catalytic module on a self-folding RNA. *RNA* **10**, 1900-1906 (2004).
15. Grishin, N.V. Fold change in evolution of protein structures. *J. Struct. Biol.* **134**, 167-185 (2001).
16. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643-647 (1995).
17. Illangasekare, M. & Yarus, M. Specific, rapid synthesis of Phe-RNA by RNA. *Proc. Natl. Acad. Sci. USA* **96**, 5470-5475 (1999).
18. Lacey, J.C., Jr., Senaratne, N. & Mullins, D.W., Jr. Hydrolytic properties of phenylalanyl- and N-acetylphenylalanyl adenylate anhydrides. *Orig. Life Evol. Biosph.* **15**, 45-54 (1984).
19. Knight, R. & Yarus, M. Analyzing partially randomized nucleic acid pools: straight dope on doping. *Nucleic Acids Res.* **31**, e30 (2003).
20. Lorsch, J.R. & Szostak, J.W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31-36 (1994).

21. Li, Y. & Breaker, R.R. Phosphorylating DNA with DNA. *Proc. Natl. Acad. Sci. USA* **96**, 2746-2751 (1999).
22. Igloi, G.L. Interaction of tRNAs and of phosphorothioate-substituted nucleic acids with an organomercurial. Probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry* **27**, 3842-3849 (1988).
23. Unrau, P.J. & Bartel, D.P. RNA-catalysed nucleotide synthesis. *Nature* **395**, 260-263 (1998).
24. Held, D.M., Greathouse, S.T., Agrawal, A. & Burke, D.H. Evolutionary landscapes for the acquisition of new ligand recognition by RNA aptamers. *J. Mol. Evol.* **57**, 299-308 (2003).
25. Huang, Z. & Szostak, J.W. Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer. *RNA* **9**, 1456-1463 (2003).
26. Schuster, P., Fontana, W., Stadler, P.F. & Hofacker, I.L. From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B. Biol. Sci.* **255**, 279-284 (1994).
27. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
28. Eklund, E.H. & Bartel, D.P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231-3238 (1995).
29. Putz, J. *et al.* Rapid selection of aminoacyl-tRNAs based on biotinylation of alpha-NH₂ group of charged amino acids. *Nucleic Acids Res.* **25**, 1862-1863 (1997).
30. Pagratis, N.C. Rapid preparation of single stranded DNA from PCR products by streptavidin induced electrophoretic mobility shift. *Nucleic Acids Res.* **24**, 3645-3646 (1996).
31. Baskerville, S. & Bartel, D.P. A ribozyme that ligates RNA to protein. *Proc. Natl. Acad. Sci. USA* **99**, 9154-9159 (2002).
32. Tuerk, C. *et al.* CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci. USA* **85**, 1364-1368 (1988).
33. Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B. & Steitz, T.A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. USA* **98**, 4899-4903 (2001).
34. Burke, D.H. *et al.* RNA aptamers to the peptidyl transferase inhibitor chloramphenicol. *Chem. Biol.* **4**, 833-843 (1997).
35. Tuschl, T., Sharp, P.A. & Bartel, D.P. A ribozyme selected from variants of U6 snRNA promotes 2',5'-branch formation. *RNA* **7**, 29-43 (2001).
36. Mannironi, C., Scerch, C., Fruscoloni, P. & Tocchini-Valentini, G.P. Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif. *RNA* **6**, 520-527 (2000).
37. Famulok, M. Molecular recognition of amino acids by RNA-aptamers: an L-citrulline binding RNA motif and its evolution into an L-arginine binder. *J. Am. Chem. Soc.* **116**, 1698-1706 (1994).
38. Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. & Famulok, M. Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science* **272**, 1343-1347 (1996).

39. Fontana, W. & Schuster, P. Continuity in evolution: on the nature of transitions. *Science* **280**, 1451-1455 (1998).
40. Schultes, E.A. & Bartel, D.P. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**, 448-452 (2000).
41. Bartel, D.P. & Szostak, J.W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411-1418 (1993).
42. Milligan, J.F., Groebe, D.R., Witherell, G.W. & Uhlenbeck, O.C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783-8798 (1987).
43. Ekland, E.H., Szostak, J.W. & Bartel, D.P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364-370 (1995).

CHAPTER 2

Modularity of a catalytic RNA

ABSTRACT

Because proteins are often built of smaller modular units, they can typically be optimized using *in vitro* recombination methods such as synthetic shuffling. Here we investigated whether a previously described kinase ribozyme was modular enough to be optimized using such an approach. Point mutations from previously isolated sequence variants of this ribozyme were shuffled in more than 10^{14} different combinations, and active variants isolated by *in vitro* selection. The rate of the most efficient ribozyme identified was 30-fold faster than that of the most efficient ribozyme used to build the pool, with a second-order rate enhancement approaching 10^{10} -fold. Further analysis revealed two groups of mutations, derived from two different ribozymes used to build the synthetically shuffled pool, that each increased the rate of the ribozyme by approximately 30-fold. The effects of these mutations were independent of one another, and when combined produced a ribozyme with a rate 600-fold faster than that of the initial isolate.

INTRODUCTION

An important goal in the field of molecular evolution is to better understand the structure and properties of adaptive fitness landscapes¹. Theoretical models suggest that many properties of such landscapes, such as the total number of fitness peaks and the average number of steps to a local fitness peak, depend primarily on two factors: N , the number of modular building blocks in a macromolecule, and K , the extent to which these building blocks interact¹. These models also provide insight into the best way to search adaptive landscapes for molecules with optimized properties: they suggest that if a macromolecule is built of smaller modular units, it can be more efficiently optimized by recombination than by random mutagenesis alone^{1,2}.

It is well established that proteins can be built of modular units. These units often correspond to independently folding structural domains, and in some cases they also represent functional units of the protein. Furthermore, mutational effects of amino acid substitutions in proteins are often independent of one another³. For example, when combined the effects of the subtilisin mutations Q103R and N218S are multiplicative with respect to k_{cat}/K_m and additive with respect to the free energy of transition-state stabilization⁴. These types of observations suggested that recombination might be an effective way to search sequence space for proteins with optimized properties, and this was confirmed by Stemmer, who showed that a method of *in vitro* recombination called DNA shuffling could be used to optimize the activity of the protein enzyme β -lactamase toward the antibiotic cefotaxime^{5,6}. A number of subsequent studies have demonstrated the general utility of this approach⁷.

Several lines of evidence suggest that ribozymes can also be built of modular building blocks. For example, the P4-P6 domain of the Group I intron can correctly fold when separated from the rest of the ribozyme⁸⁻¹⁰, and secondary structure elements that can fold and function by themselves are often identified in the context of larger

ribozymes by comparative sequence analysis^{11,12}. Some evidence also suggests that such domains can often be broken down into even smaller building blocks. For example, in many cases the thermodynamic stability of an RNA helix can be determined by assuming that the contribution of a particular base pair depends only on the identity of the flanking base pairs in the helix^{13,14}. In other studies secondary structure elements in ribozymes such as helices have been treated as independent units^{15,16}.

If a typical ribozyme is built of smaller modular building blocks, recombination between these elements could generate variants with improved catalytic properties. With this idea in mind we set out to identify the types of RNA motifs (analogous to protein schemas¹⁷) that could be recombined among ribozyme variants with the same fold, and the extent to which this process could generate ribozymes with improved catalytic properties. To address these questions, point mutations from sequence variants of a previously described kinase ribozyme were recombined *in vitro* in more than 10¹⁴ different arrangements using a method called synthetic shuffling¹⁸, and the most active variants isolated by *in vitro* selection¹². These variants were up to 30-fold more efficient than any of the ribozymes used to build the pool, suggesting that this ribozyme is built of smaller modular units. The structures of at least part of two of these units, which we refer to as modules, were identified using a combination of comparative sequence analysis and site-directed mutagenesis: one corresponded to secondary structure elements in a helix, while the other involved interactions between opposite sides of a loop. Further analysis revealed two blocks of mutations, derived from two different ribozyme variants used to build the pool, that each increased the rate of the ribozyme by approximately 30-fold. By combining these mutational changes into a single molecule a ribozyme with a rate more than 600-fold higher than that of the initial isolate was generated. These results suggest that even a ribozyme as small as the one characterized here, with a 50-nucleotide minimized catalytic core, can be built of smaller modular units. They also indicate that ribozymes with improved catalytic properties can be generated by synthetic shuffling and

re-selection.

RESULTS

In vitro recombination by synthetic shuffling

Our questions about ribozyme modularity were addressed in the context of a previously described kinase ribozyme called 5-16¹². This ribozyme thiophosphorylates itself at an internal 2' hydroxyl group using GTP γ S as a substrate (**Fig. 1a**). Its secondary structure consists of an asymmetrically bulged hairpin folded back on itself in a pseudoknot, flanked by two less conserved helices that are not required for catalytic activity (**Fig. 1b**). In addition to the prototype sequence, 22 variants of this ribozyme were previously isolated by random mutagenesis and re-selection¹², some of which had catalytic rates up to 40-fold faster than that of the parental kinase.

Analysis of these variants had revealed several pairs of co-varying positions that corresponded to base pairs in the secondary structure of the ribozyme, but had provided little information about other types of correlations¹². Furthermore, the mutations responsible for the increased activity of these variants could not be readily identified by comparative sequence analysis, and the extent to which combining them into a single molecule would generate an even faster ribozyme was not clear. To address each of these issues, an oligonucleotide template encoding this ribozyme was synthesized by synthetic shuffling¹⁸. Degenerate positions in the template corresponded to positions that were variable in a sequence alignment of kinase ribozyme variants. For example, at position 11 in the minimized catalytic core of the ribozyme, at which most ribozyme variants contained a U but several contained a C, the pool was synthesized to contain 50% U and 50% C. A pool synthesized in such a way will contain the point mutations present in the

sequence alignment in a large number of possible combinations, facilitating identification of correlations that might not be readily detected in ribozymes isolated from a randomly mutagenized pool. Furthermore, to the extent that this ribozyme is built of smaller modular units, combining beneficial mutations from different ribozyme variants into single molecules was expected to produce ribozymes with improved catalytic properties.

Isolation of faster kinase ribozymes

To isolate kinase ribozymes from our recombined pool, RNA was incubated with GTP γ S^{12,19}, and molecules that became thiophosphorylated during the incubation were isolated on *N*-acryloylaminophenylmercuric chloride (APM) polyacrylamide gels^{20,21}. These molecules were then amplified by RT-PCR and transcribed to generate RNA for the next round of selection. After three rounds, the pool catalyzed self-thiophosphorylation as efficiently as the fastest kinase ribozymes used in its design, and after 2-3 more rounds of selection clones were sequenced and assayed individually for kinase activity.

The most efficient kinase ribozyme isolated in the selection was 30-fold faster than that of any of the ribozymes used to build it, and 1300-fold faster than that of the initial isolate (Fig. 2), with a k_{cat}/K_m of $8.8 \times 10^2 \text{ M}^{-1} \text{ min}^{-1}$. The rate of this ribozyme was even faster in an optimized buffer, with a k_{cat} of 0.4 min^{-1} , a K_m for GTP γ S of $20 \mu\text{M}$, and a k_{cat}/K_m of $10^4 \text{ M}^{-1} \text{ min}^{-1}$. With respect to the non-enzymatic rate of GTP γ S hydrolysis, the second-order rate enhancement of this ribozyme approached 10^{10} -fold (Fig. 2b), which compares favorably with that of previously isolated kinases made of DNA²² or RNA¹⁹. In contrast, the second-order rate enhancement of the protein enzyme

T4 polynucleotide kinase is approximately 10^{15} -fold²³, and approaches the diffusion controlled limit of the reaction²⁴.

Identification of ribozyme modules

The previous results suggested that this ribozyme is built of smaller modular units. We next set out to determine the structure of some of these modules. Sequences of 67 active ribozymes were aligned with the parental sequences (**Supplementary Fig. 1**), and correlated pairs of nucleotides were identified by mutual information analysis^{25,26} using the program MatrixPlot²⁷. Because this approach required positions being compared to be homologous, only portions of the sequence that could be readily aligned were analyzed, corresponding to 14 recombined positions (and 91 possible pairs), all within the 50-nucleotide minimized catalytic core of the ribozyme. Mutual information values from the ribozyme alignment were compared to values from control alignments (in which the order of nucleotides in each column of the ribozyme alignment was randomized) to determine a cutoff for significance of mutual information values. This analysis suggested a high degree of coupling between different positions in the ribozyme (**Supplementary Table 1 and Supplementary Fig. 2**): for example, using an alignment containing 67 recombined kinase ribozymes and the 7 parents used to construct the pool, 22 out of 91 possible pairs of positions had mutual information values ≥ 0.094 , whereas values ≥ 0.094 only occurred approximately once in every 100 pairs from control alignments.

The potential significance of many of these correlations was further investigated using an independent method. For each pair of positions tested, the effect of a point mutation on the rate of the ribozyme was determined for each nucleotide in the pair, and

compared to the double-mutant effect²⁸⁻³⁰. The departure from independence, D , was defined as the product of the two single-mutant effects divided by the double-mutant effect, and was expressed such that D was always ≥ 1 . For independent positions, the double-mutant effect should equal the product of the two single-mutant effects³, and D should equal 1. Mutational effects were examined in the context of an optimized, minimal version of the ribozyme thought to be representative of the ribozymes isolated in the selection. In addition, D values for most pairs were determined in a second mutational background. This background differed from the initial background at a single position that was correlated with at least one of the positions in the pair being tested. For example, because positions 15 and 22 were both correlated with position 23 according to mutual information analysis, the D value for the 15-22 pair was determined in both the 23 A and 23 G backgrounds.

Of the 17 pairs of positions tested, 7 were strongly coupled in at least one sequence background, with D values ranging from 10 to 10^5 (**Table 1** and **Supplementary Fig. 2**). These values were comparable to those previously determined for base-paired positions in minimal versions of several different ribozymes, which ranged from 10 to more than 6400 (reference 12). D values for these pairs were strongly dependent on mutational background, and in several cases changed by more than 100-fold when examined in a different background (**Table 1**). The strongly coupled pairs of nucleotides identified in this analysis corresponded to a base pair (the 20-47 pair) and two maximally connected groups of nucleotides we refer to as modules: the 11-12-28 module (made up of the 11-12, 11-28, and 12-28 pairs), and the 15-22-23 module (made up of the 15-22, 15-23, and 22-23 pairs) (**Fig. 3**). While the 20-47 pair and the 11-12-28 module

could be understood in terms of our previously proposed secondary structure model of the ribozyme, the 15-22-23 module provided evidence for a previously undetected interaction between opposite sides of L3. If these nucleotides physically interact in some way, nucleotide 14 (an invariant G) probably also pairs with nucleotide 24 (an invariant C); without the 15-22-23 interaction, this would be an isolated base pair. The resulting structure, which contains a bulged A flanked by G-C base pairs, is similar in some respects to that of the G-binding site of the Group I intron³¹⁻³³. Since the thiophosphorylation site of this ribozyme is the 2' hydroxyl of nucleotide 23, the location of this putative GTP γ S-binding site is intriguing.

The remaining 10 pairs of positions tested linked parts of the ribozyme that were distant in its primary sequence and secondary structure. These pairs appeared to be either weakly coupled or independent, with D values ranging from 1 to 4.6 (**Table 2** and **Supplementary Fig. 2**). In comparison, D values for 4 negative controls (pairs that were not strongly correlated based on mutual information analysis) ranged from 1.2 to 2 (E.A.C. and D.P.B., data not shown). Some of these correlations may represent interactions that occur in some sequence backgrounds but not in others. Such an interpretation could explain why these correlations were detected by comparative sequence analysis but not well supported by site-directed mutagenesis: mutual information values measured the correlation between positions in the context of many different ribozyme sequences, while D values were typically measured in only two sequence backgrounds. Alternatively, we cannot rule out the possibility that some of these correlations are unrelated to ribozyme function, but instead are linked to other properties that increase the overall fitness of a ribozyme during the selection. Such

properties include the ability of a ribozyme template to be transcribed by T7 RNA polymerase and amplified by PCR, as well as the ability of a ribozyme to act as a template for reverse transcriptase.

Combining advantageous mutations can produce faster ribozymes

We next set out to identify the mutations responsible for the increased activity of these ribozymes, and to quantitatively determine the extent to which combining them would produce faster ribozymes. At 8 out of 14 variable positions in the minimized core of the ribozyme, mutational changes were identified that were more common than expected based on the composition of the starting pool (**Supplementary Fig. 1**). In most cases the advantageous effects of these changes could be observed by mutating single positions, base pairs, or modules either individually or in combination. The advantageous mutations that could be rationalized by site-directed mutagenesis came from two different parental ribozymes used to build the pool. In the context of a minimized version of the initial isolate of the ribozyme, the 11-12-28 U-C-A to C-U-G change (derived from the A-25 ribozyme used to build the pool) increased the rate of the ribozyme 32-fold (**Fig. 4, Table 3**). These mutations were tightly coupled (**Table 1**), and changing either the 11-28 base pair without changing position 12, or changing position 12 without changing the 11-28 base pair, was deleterious (E.A.C. and D.P.B., data not shown). The 16-20-47 U-A-U to A-C-G change (derived from the A-24 ribozyme used to build the pool) increased the rate of the ribozyme 23-fold (**Fig. 4, Table 3**). These mutations were weakly coupled: in the context of a minimized version of the initial isolate of the ribozyme, changing position 16 from U to A without mutating the 20-47 base pair was slightly deleterious,

and changing the 20-47 base pair from A-U to C-G without mutating position 16 increased the rate of the ribozyme 8.3-fold, but combining these changes increased the rate of the ribozyme 23-fold (**Table 3**). Although these experiments indicated that 16 and 20-47 were weakly coupled, 16-20-47 was not defined as a module for two reasons: the D value of the 16 and 20-47 pair ($D = 4.6$) (**Table 2**) was smaller than typical D values for nucleotide pairs in modules ($D = 10$ to 1000) (**Table 1**), and position 16 was also weakly coupled to positions 11 ($D = 3.4$), 22 ($D = 4$), and 41 ($D = 3$) (**Table 2**).

The mutational effects of the 11-12-28 U-C-A to C-U-G and 16-20-47 U-A-U to A-C-G changes were independent of one another: the rate of a ribozyme in which these changes were combined was more than 600-fold faster than that of the initial isolate of this ribozyme (**Fig. 4, Table 3**). This indicates that combining beneficial mutations derived from different ribozyme variants into a single molecule can generate ribozymes with even faster rates.

Several other mutations that were predicted to increase the rate of the ribozyme could not be rationalized by site-directed mutagenesis. In 29 out of 67 ribozymes from the recombined pool, the 15-22-23 module had changed from A-A-U to G-G-G (**Supplementary Fig. 1**), yet this change was deleterious in several different sequence backgrounds, as well as in the context of the full-length ribozyme (E.A.C. and D.P.B., data not shown). It is possible that this mutational change only increases the rate of the ribozyme in the presence of specific mutations in P1 and/or P5. Since this portion of the sequence could not be reliably aligned, such mutations could not be identified. Alternatively, as previously discussed, we cannot rule the possibility that the 15-22-23 A-

A-U to G-G-G change increases the overall fitness of a ribozyme variant in which it occurs but does not increase its catalytic rate.

Analysis of the sequence alignment of kinase ribozymes isolated in this selection also revealed several mutations likely to be deleterious based on their frequencies (**Supplementary Fig. 1**), although the effects of these mutations were not tested by site-directed mutagenesis. Taken together, these results suggest that the ribozymes used to build the recombined pool (which were previously isolated from a randomly mutagenized pool¹²) contained both beneficial and deleterious mutations, and that homologous recombination was a useful optimization strategy for at least two reasons: it combined beneficial mutations from different ribozyme variants into single molecules, and it removed deleterious mutations from these ribozymes.

DISCUSSION

The sequence neighborhood of a ribozyme is often enriched for variants with improved catalytic activities. For this reason, ribozymes can typically be optimized using a combination of random mutagenesis and re-selection^{11,34}. Additional methods to optimize ribozymes have not been extensively explored. In one approach, diversity was produced by fusing an RNA ligase ribozyme with weak polymerase activity to a random sequence domain. A more efficient and general polymerase was successfully isolated from the resulting pool³⁵. In another approach, a technique called nonhomologous random recombination was used to generate a pool of topological variants of a DNA aptamer to streptavidin.³⁶ The affinity of one of these variants towards streptavidin was 46-fold higher than that of the aptamer used to build the pool. Nonhomologous random recombination was also used to generate a pool of topological variants of a nucleotide synthase ribozyme³⁷. Although the ribozyme variants isolated from this pool were not faster than the ribozyme used to build it, some were considerably smaller. It is also known that homologous recombination between different aptamers domains can generate molecules with the binding properties of both parental aptamers³⁸. In this study we found that recombination of point mutations within a single domain was an effective way to further improve a kinase ribozyme that had been previously optimized by random mutagenesis and re-selection: the rate of the most efficient kinase ribozyme isolated in this selection was 30-fold faster than that of the most efficient kinase used to build the recombined pool, and 1300-fold faster than that of the initial isolate (**Fig. 2**).

Pools of recombined sequences are most commonly generated using the technique of DNA shuffling, in which homologous DNA sequences are randomly fragmented using

DNase I and reassembled by mutual primed annealing and PCR amplification⁵⁻⁷. In this study we instead used a technique called synthetic shuffling¹⁸ to generate our pool of recombined kinase ribozymes. A partially degenerate oligonucleotide template encoding this ribozyme was synthesized in which variable positions corresponded to variable positions in an alignment of ribozyme sequence variants. We used this method rather than DNA shuffling to construct our pool for several reasons. First, even nucleotides that are close in the primary sequence of the ribozyme can be readily recombined by synthetic shuffling, whereas the recombination frequency between two positions in a pool generated by DNA shuffling is expected to be proportional to the distance between them. This consideration appeared to be especially important considering the small size of the ribozyme used in these experiments. Second, because synthetic shuffling does not rely on DNA annealing, even blocks of sequence linked by highly variable joining regions can be efficiently recombined. Third, the method is extremely straightforward from a technical perspective. A major disadvantage of this approach relative to DNA shuffling, however, is that ribozymes cannot be recombined unless their sequences are first known.

In addition to generating ribozymes with improved catalytic rates, synthetic shuffling was also expected to facilitate the detection of correlated positions that could not be readily identified by analyzing ribozyme variants isolated from a randomly mutagenized pool. To see this, consider a hypothetical ribozyme in which nucleotides 1 and 2 are coupled. Assume that the nucleotide sequence of the initial isolate of this ribozyme at these positions is AA, that the mutational change CC is neutral, and that all other mutational changes at these positions inactivate the ribozyme. In a pool in which this ribozyme is randomly mutagenized at a rate of 20% per position, $(0.8)^2 \times 100 = 64\%$

of the molecules in the starting pool will have the sequence AA at positions 1 and 2, while only $(0.2/3)^2 \times 100 = 0.44\%$ will have the sequence CC at these positions (see also reference ¹¹ for a discussion of these issues). If the probability that a ribozyme variant is isolated from the pool depends only on its catalytic rate and its abundance in the starting pool, the CC co-variation will occur approximately once in every 140 active ribozymes isolated from the pool. On the other hand, in a pool generated by synthetic shuffling based on the sequences AA and CC, $(0.5)^2 \times 100 = 25\%$ of pool molecules will contain the sequence AA and 25% will contain the sequence CC. The CC co-variation will therefore occur once in every two active ribozymes isolated from this pool, making the correlation much easier to detect. Note that this effect is even more pronounced for larger blocks of correlated positions: given the assumptions above, a neutral AAA to CCC mutation will occur about once in every 1700 active ribozymes isolated from a randomly mutagenized pool, but once in every two active ribozymes isolated from a synthetically shuffled pool. Note also that meaningful sequence diversity cannot be encoded into a synthetically shuffled pool unless it is first observed in a ribozyme variant isolated in some other way (i.e. from a randomly mutagenized pool).

All of the strongly coupled pairs of nucleotides identified in our analysis could be rationalized in terms of the secondary structure of this ribozyme, although not all corresponded to Watson-Crick base pairs. Of the 7 pairs of positions identified with D values ≥ 10 , four corresponded to interactions within helices and three corresponded to interactions between opposite sides of a loop (**Table 1 and Fig. 3b**). Pairs of positions in a ribozyme could be coupled for many other reasons: for example, because they form tertiary interactions in the catalytically active fold of the ribozyme, because they interact

at some point in its folding pathway, or because they prevent interactions from occurring that either destabilize the catalytically active fold of the ribozyme or interfere with its folding. Our evidence with regard to the importance of these types of interactions was somewhat contradictory. On the one hand, mutual information analysis suggested a high degree of coupling between different positions in the ribozyme, even among pairs that were distant in its secondary structure (**Supplementary Table 1** and **Supplementary Fig. 2**). On the other hand, the higher order pairs tested by site-directed mutagenesis were either independent or weakly coupled, with D values ranging from 1 to 4.6 (**Table 2** and **Supplementary Fig. 2**). The effects of these pairs may be strongly dependent on the mutational background in which they are tested, which is consistent with our finding that D values of strongly coupled positions can be extremely sensitive to mutational background (**Table 1**). It is also consistent with several other types of observations. For instance, pairs of functional groups in RNA (and DNA) hairpins sometimes only interact in the presence of a functional group at a third site^{39,40}. Also, comparison of crystal structures of homologous tRNA molecules from different species indicates that RNA tertiary interactions such as base triples sometimes only form in certain sequence backgrounds⁴¹. In addition, specific mutations that increase the thermostability of a ribozyme in one sequence context can be destabilizing when tested in a different sequence background⁴². Alternatively, as previously discussed, some of these correlations may reflect interactions that increase the overall fitness of a ribozyme during the selection but do not affect its catalytic rate.

In conclusion, our results indicate that even a ribozyme as small as the kinase analyzed here is built of smaller modular units. The most strongly coupled pairs of positions identified in this analysis can be rationalized in terms of the secondary structure

of the ribozyme, although not all correspond to Watson-Crick base pairs. We also obtained evidence for a weakly coupled network of interactions that in many cases linked pairs of positions distant in the primary sequence and secondary structure of the ribozyme. Our results also show that synthetic shuffling of point mutations within a single ribozyme domain can generate molecules with improved catalytic properties.

Figure 1 Catalytic activity and secondary structure of kinase ribozyme 5-16. (a) The ribozyme thiophosphorylates itself at an internal 2' hydroxyl group using GTP γ S as a substrate. (b) Secondary structure and minimized catalytic core of the ribozyme.

Figure 2 Faster ribozymes generated by synthetic shuffling. (a) APM polyacrylamide gel showing time course of 5-16 (the prototype ribozyme), A-25 (one of the most efficient ribozymes used to build the recombined pool), and Rec 7-5 (the most efficient ribozyme isolated from the recombined pool). Reaction conditions were 0.2 μ M ribozyme, 0.2 μ M each blocking oligonucleotide (DNA oligonucleotides complementary to the primer binding sites of the ribozyme) when necessary, 10 μ M GTP γ S, and ribozyme selection buffer (10 mM MgCl₂, 5 mM CaCl₂, 200 mM KCl, 100 mM HEPES, pH 7.2). The reaction of Rec 7-5 in an optimized buffer (30 mM CaCl₂, 200 mM KCl, 100 mM HEPES, pH 7.2) is also shown. (b) Graph showing k_{cat}/K_m values for 5-16 (0.7 M⁻¹ min⁻¹), A-25 (30 M⁻¹ min⁻¹), and Rec 7-5 (880 M⁻¹ min⁻¹). In optimized buffer, Michaelis-Menten parameters for Rec 7-5 were $k_{\text{cat}} = 0.4 \text{ min}^{-1}$, $K_m = 30 \mu\text{M}$, and $k_{\text{cat}}/K_m = 1.3 \times 10^4 \text{ M}^{-1} \text{ min}^{-1}$. For comparison, the nonenzymatic hydrolysis rate of GTP γ S in ribozyme selection buffer is also shown.

Figure 3 Identification of ribozyme modules. (a) APM polyacrylamide gel showing the effects of the 15 A to G, 22 A to G, and 15-22 A-A to G-G mutations. This pair of positions was strongly coupled, with a D value of 170 in the mutational background shown here. Reaction conditions were as described in the legend to **Fig. 2**, except that the GTP γ S concentration was at 30 μ M, and the ribozyme concentration was at 1 μ M. (b)

Ribozyme modules (as defined in the text) mapped onto the secondary structure of the ribozyme. Nucleotides of the same color are strongly coupled with one another ($D \geq 10$ in at least one mutational background). Positions that were variable in the synthetically shuffled pool are numbered.

Figure 4 Mutations that increase the rate of the ribozyme. APM gel showing the effects of the 11-12-28 U-C-A to C-U-G and 16-20-47 U-A-U to A-C-G mutations either individually or in combination. The 11-12-28 mutation was derived from the A-25 ribozyme, and the 16-20-47 mutation was derived from the A-24 ribozyme. Reaction conditions were as described in the legend to **Fig. 2**.

Table 1 D values for strongly correlated pairs of positions ($D \geq 10$ in at least one mutational background). Sequences and relative rates of mutants used to calculate D values are given in **Supplementary Table 2**.

Table 2 D values for weakly correlated or independent pairs of positions. Sequences and relative rates of mutants used to calculate D values are given in **Supplementary Table 2**.

Table 3 Mutations that increase the rate of the ribozyme. Reaction conditions were as described in the legend to **Fig. 3a**.

Supplementary Figure 1 Sequence alignment of synthetically shuffled kinase ribozymes. The seven ribozymes shown at the top of the alignment were used to build

the shuffled pool, and the 67 ribozymes below were isolated by *in vitro* selection.

Colored bars indicate paired regions in the secondary structure of the ribozyme, and numbers indicate shuffled positions. Note that only the minimized catalytic core of each ribozyme is shown in the alignment. Based on their frequencies, changes shown in red were predicted to be up mutations, and changes shown in pink were predicted to be down mutations. Changes shown in blue were not encoded in the pool, but likely occurred during the RT-PCR step of the selection.

Supplementary Figure 2 Network of correlated pairs of positions in the ribozyme. **a**, Correlated positions among 14 nucleotides in the minimized catalytic core of the ribozyme. Lines indicate nucleotide pairs with mutual information values ≥ 0.094 . Red lines indicate pairs with D values greater than 10 in at least one sequence background; orange lines indicate pairs with D values between 2 and 5 in at least one sequence background; thick black lines indicate pairs that were not tested by site-directed mutagenesis; thin black lines indicate pairs with D values less than 2. **b**, Secondary structure of the minimized catalytic core of the ribozyme. Positions that were variable in the synthetically shuffled pool are numbered.

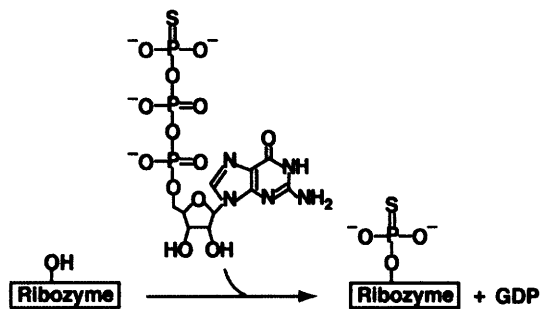
Supplementary Table 1 Pairs of positions with mutual information values ≥ 0.094 .

Supplementary Table 2 Relative rates of ribozyme mutants characterized in these experiments. Reactions were performed in ribozyme selection buffer (10 mM MgCl₂, 5 mM CaCl₂, 200 mM KCl, 100 mM HEPES, pH 7.2) at 1 μ M ribozyme and typically 5 to

30 μM GTP γ S. $k_{\text{cat}}/K_{\text{m}}$ values were determined from at least three independent points over at least a 6-fold range of GTP γ S concentrations in the linear part of the substrate titration curve for each ribozyme. Due to the slow rates of some mutants, rates were not normalized to the fraction of reacted ribozyme. All mutations were generated in the 69.20 mutational background (5' GGUACCCGAGCUGGAAGCACGAUCACGGCAUA AUCGGUACUCUUGCGUGC 3').

Figure 1

a



b

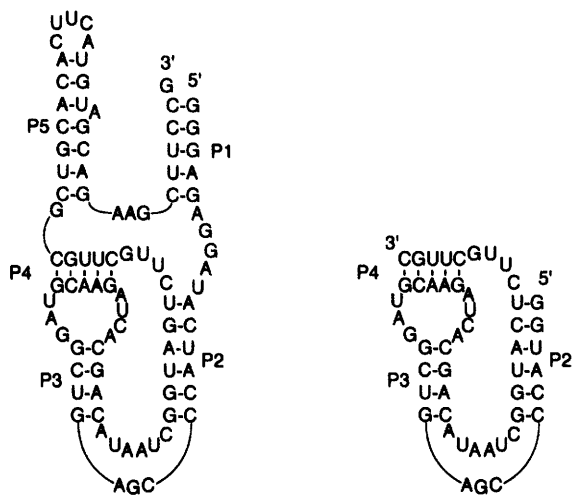
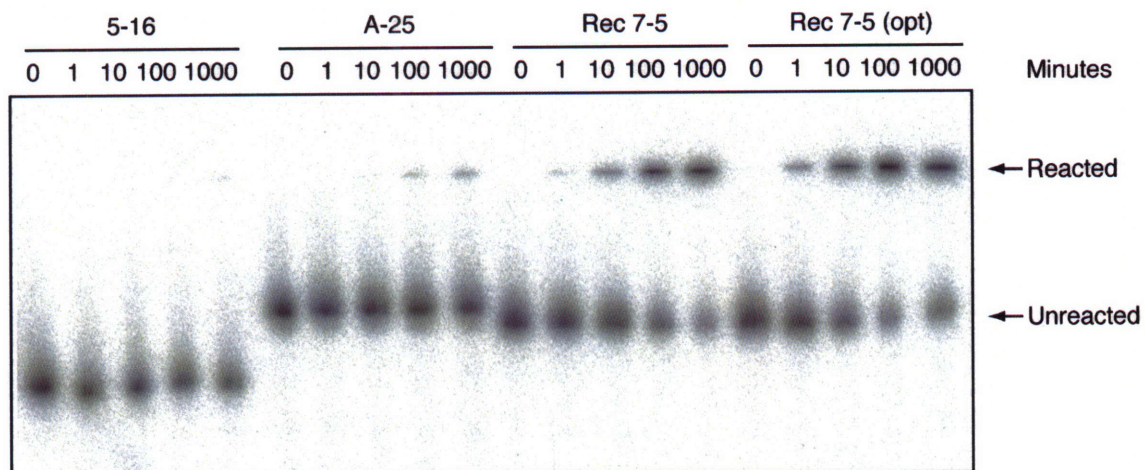


Figure 2

a



b

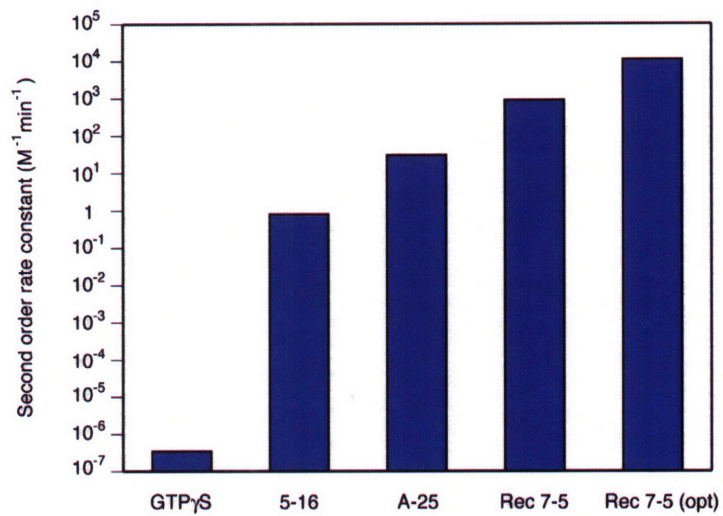
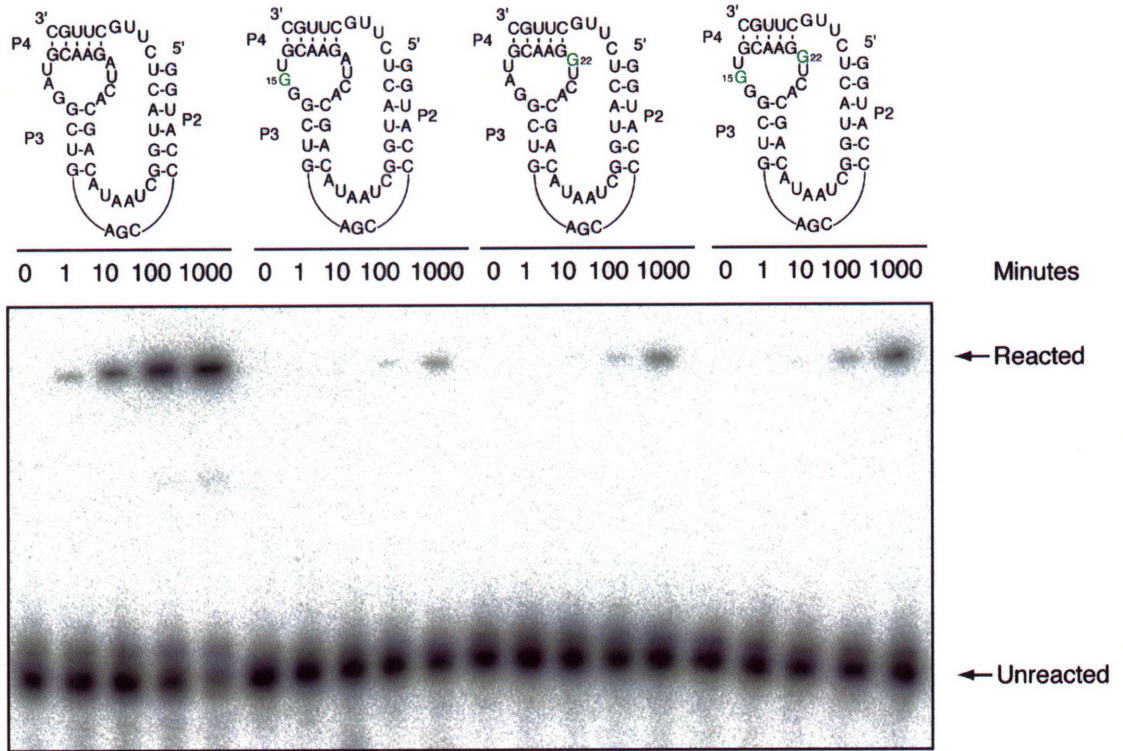


Figure 3

a



b

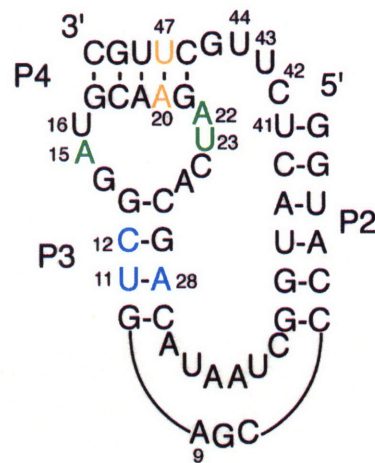


Figure 4

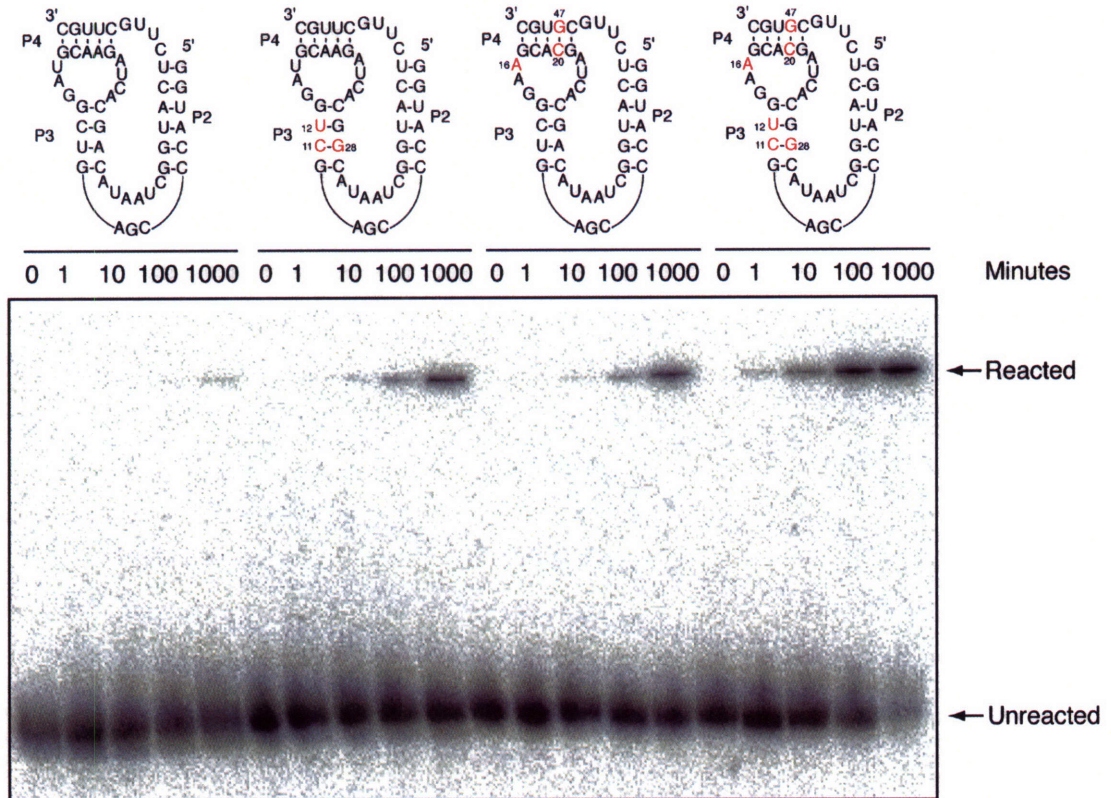


Table 1

Module or pair	Positions mutated	Background	D
11-28-12	11 and 28	69.20 (12 = U)	110,000
		69.61 (12 = C)	500
	11-28 and 12	69.20	90
15-22-23	15 and 22	69.20 (23 = U)	170
		69.64 (23 = G)	12
	15 and 23	69.20 (22 = A)	800
		69.64 (22 = G)	60
	22 and 23	69.20 (15 = A)	1
		69.64 (15 = G)	14
20-47	20 and 47	69.20	1,100

Table 2

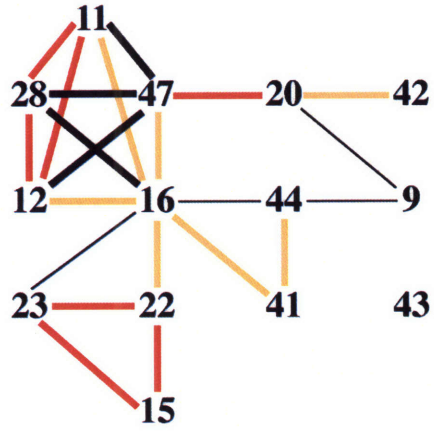
Positions mutated	Background	D
16 and 41	69.20 (44 = U)	1.4
	69.52 (44 = A)	3
16 and 44	69.20 (41 = U)	1.4
	69.52 (41 = G)	1.6
41 and 44	69.20 (16 = A)	2
	69.52 (16 = U)	1.1
11 and 16	69.20 (12 = U)	3.4
	69.62 (12 = C)	1.1
12 and 16	69.20 (11 = C)	1.7
	69.62 (11 = U)	2.2
16 and 22	69.20 (23 = U)	4
	69.63 (23 = G)	2.7
16 and 23	69.20 (22 = A)	1.1
	69.63 (22 = G)	1.7
9 and 20	69.20	1.2
9 and 44	69.20	1.5
20 and 42	69.20	2.4
16 and 20-47	69.20	4.6

Table 3

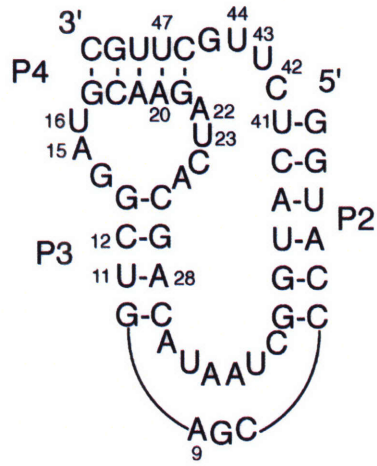
Positions mutated	Fold increase in rate
None	-
11-12-28	32
16	0.75
20-47	8.3
11-12-28 and 16	20
11-12-28 and 20-47	220
16 and 20-47	23
11-12-28, 16 and 20-47	630

Supplementary Figure 2

a



b



Supplementary Table 1

Nucleotide pair	Mutual information value
20 and 47	0.947
22 and 23	0.424
11 and 28	0.370
11 and 12	0.336
15 and 22	0.318
12 and 28	0.308
15 and 23	0.236
16 and 44	0.186
11 and 47	0.184
16 and 41	0.168
9 and 20	0.148
12 and 16	0.126
16 and 28	0.120
9 and 44	0.113
12 and 47	0.111
16 and 47	0.105
16 and 22	0.104
41 and 44	0.101
16 and 23	0.100
11 and 16	0.097
20 and 42	0.094
28 and 47	0.094

Supplementary Table 2

Ribozyme	Mutation(s)	Fold decrease in rate
69.20	None	-
69.32	9 A to G	0.72
69.21	11 C to U	180
69.22	12 U to C	38
69.47	15 A to G	650
69.24	16 A to U	2.5
69.31	20 C to U	8
69.45	22 A to G	13
69.55	23 U to G	80
69.26	28 G to A	20,000
69.42	41 U to G	13
69.37	42 C to U	2.7
69.34	43 U to A	2
69.39	44 U to A	2.1
69.53	47 G to A	360
69.33	9 A to G, 20 C to U	4.7
69.35	9 G to A, 43 U to A	2.4
69.40	9 A to G, 44 U to A	1.1
69.23	11 C to U, 12 U to C	130
69.25	11 C to U, 16 A to U	1,500
69.27	11 C to U, 28 G to A	33
69.28	12 U to C, 16 A to U	160
69.29	12 U to C, 28 G to A	2,200
69.36	12 U to C, 43 U to A	90
69.60	15 A to G, 16 A to U	1,300
69.48	15 A to G, 22 A to G	51
69.58	15 A to G, 23 U to G	66
69.46	16 A to U, 22 A to G	8.1
69.59	16 A to U, 23 U to G	220
69.43	16 A to U, 41 U to G	22
69.41	16 A to U, 44 U to A	7.9
69.38	20 C to U, 42 C to U	8
69.54	20 C to U, 47 G to A	2.6
69.56	22 A to G, 23 U to G	1,100
69.44	41 U to G, 44 U to A	57
69.51	43 U to A, 44 U to A	8.4
69.65	9 A to G, 20 C to U, 42 C to U	3.7
69.66	9 A to G, 20 C to U, 44 U to A	7.8
69.62	11 C to U, 12 U to C, 16 A to U	500
69.61	11 C to U, 12 U to C, 28 G to A	14
69.64	15 A to G, 22 A to G, 23 U to G	73
69.63	16 A to U, 22 A to G, 23 U to G	1,100
69.52	16 A to U, 41 U to G, 44 U to A	62

METHODS

Pool synthesis

Two separate pools were synthesized for these experiments. Pool A incorporated mutations from the minimized catalytic core (P2-P4) of each of 22 previously described kinase variants¹², as well as mutations from P1 and P5 of 5-16, A-29, and A-25, and had the sequence 5' **GGAUGCCUGGUAAAGKRMGAGKAUACUACCCGVGKKGGRN** **GSMNSDKCACGRCAUWRUCGGUAGDHHHRYDKRYRCKSCMYWSUUYAUGU** **AGCAGVWRCYUCBSAAGUCAAUAGCCUAGGG** 3' (R = A or G; Y = C or U; M = A or C; K = G or U; S = G or C; W = A or U; H = A, U or C; B = G, U or C; V = G, A or C; D = G, A or U; N = G, A, C or U; primer binding sites shown in bold, and connected to the pool by AA linkers). Pool B incorporated mutations from kinase variants 5-16, A-29, A-25, A-24, A-7, A-8, and A-1¹², and had the sequence 5' **GGAUGCCUGGUAAAG** **DRMGAKDAUACUACCCGRGYYGGRDGCAHGRKCACGRCAUAAUCGGUAGK** **YWHGCDUGCRYKSCHYDSKWYAHGBWGCAGVDDCYUCBBAAGUCAAUAG** **CCUAGGG** 3'. The two pools were mixed in equal amounts, and kinases were isolated by *in vitro* selection. Most of the kinases isolated were from pool B, and only these ribozymes were used for comparative sequence analysis, although the fastest kinase isolated (Rec 7-5) was derived from pool A, suggesting that mutations that increase the catalytic rate of the ribozyme were concentrated in the minimized catalytic core rather than P1 or P5.

Isolation of kinase ribozymes by *in vitro* selection

In vitro selection was as described in Chapter 1 of this thesis. As was the case in the re-selection of 5-16 following random mutagenesis (Chapter 1), pool RNA was not dephosphorylated before each round of selection. In round 1 the pool was incubated for 5 minutes at 100 μM GTP γ S, in rounds 2-3 the pool was incubated for 1 minute at 100 μM GTP γ S, in rounds 4-5 the pool was incubated for 6 seconds at 100 μM GTP γ S, and in rounds 6-7 the pool was incubated for 6 seconds at 10 μM GTP γ S.

Comparative sequence analysis

An alignment of 74 sequences consisting of 14 variable positions in the minimized catalytic core of the ribozyme (comprising the P2-P4 helices) was analyzed using the program MatrixPlot²⁷. Randomized alignments were generated by shuffling each column in the ribozyme alignment using the program Sequence Shuffle Tool (available at the Arizona Research Laboratories Division of Biotechnology website).

Kinetic analysis

Ribozyme rates were measured as described in Chapter 1 of this thesis. Reactions were performed in ribozyme selection buffer (10 mM MgCl₂, 5 mM CaCl₂, 200 mM KCl, 100 mM HEPES, pH 7.2) at 0.2-1 μM ribozyme and typically 5 to 30 μM GTP γ S. k_{cat}/K_m values were determined from at least three independent points over at least a 6-fold range of GTP γ S concentrations in the linear part of the substrate titration curve for each ribozyme. Due to the slow rates of some mutants, rates were not normalized to the fraction of reacted ribozyme.

ACKNOWLEDGEMENTS

We thank U. Muller and M. Lawrence for comments on this manuscript, and members of the lab for helpful discussions. This work was supported by a grant from the National Institutes of Health.

REFERENCES

1. Kauffman, S.A. *The Origins of Order*, (Oxford University Press, Oxford, 1993).
2. Holland, J.H. Genetic algorithms. *Sci. Am.* **267**, 66-72 (1992).
3. Wells, J.A. Additivity of mutational effects in proteins. *Biochemistry* **29**, 8509-8517 (1990).
4. Chen, K.Q. *et al.* Enzyme engineering for nonaqueous solvents. II. Additive effects of mutations on the stability and activity of subtilisin E in polar organic media. *Biotechnol. Prog.* **7**, 125-129 (1991).
5. Stemmer, W.P. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389-391 (1994).
6. Stemmer, W.P. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10747-10751 (1994).
7. Cramer, A., Raillard, S.A., Bermudez, E. & Stemmer, W.P. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288-291 (1998).
8. Murphy, F.L. & Cech, T.R. An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry* **32**, 5291-5300 (1993).
9. Cate, J.H. *et al.* Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-1685 (1996).
10. Doherty, E.A. & Doudna, J.A. The P4-P6 domain directs higher order folding of the *Tetrahymena* ribozyme core. *Biochemistry* **36**, 3159-3169 (1997).
11. Eklund, E.H. & Bartel, D.P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231-3238 (1995).
12. Curtis, E.A. & Bartel, D.P. New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.* **12**, 994-1000 (2005).
13. Tinoco, I., Jr. *et al.* Improved estimation of secondary structure in ribonucleic acids. *Nat. New Biol.* **246**, 40-41 (1973).
14. Freier, S.M. *et al.* Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. USA* **83**, 9373-9377 (1986).
15. Sabeti, P.C., Unrau, P.J. & Bartel, D.P. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem. Biol.* **4**, 767-774 (1997).
16. Knight, R. & Yarus, M. Finding specific RNA motifs: function in a zeptomole world? *RNA* **9**, 218-230 (2003).
17. Voigt, C.A., Martinez, C., Wang, Z.G., Mayo, S.L. & Arnold, F.H. Protein building blocks preserved by recombination. *Nat. Struct. Biol.* **9**, 553-558 (2002).
18. Ness, J.E. *et al.* Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* **20**, 1251-1255 (2002).
19. Lorsch, J.R. & Szostak, J.W. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31-36 (1994).
20. Igloi, G.L. Interaction of tRNAs and of phosphorothioate-substituted nucleic acids with an organomercurial. Probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry* **27**, 3842-3849 (1988).

21. Unrau, P.J. & Bartel, D.P. RNA-catalysed nucleotide synthesis. *Nature* **395**, 260-263 (1998).
22. Li, Y. & Breaker, R.R. Phosphorylating DNA with DNA. *Proc. Natl. Acad. Sci. USA* **96**, 2746-2751 (1999).
23. Lillehaug, J.R. & Kleppe, K. Kinetics and specificity of T4 polynucleotide kinase. *Biochemistry* **14**, 1221-1225 (1975).
24. Abeles, R.H., Frey, P.A. & Jencks, W.P. *Biochemistry*, (Jones and Bartlett, Boston, MA, 1992).
25. Chiu, D.K. & Kolodziejczak, T. Inferring consensus structure from nucleic acid sequences. *Comput. Appl. Biosci.* **7**, 347-352 (1991).
26. Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J. & Stormo, G.D. Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res.* **20**, 5785-5795 (1992).
27. Gorodkin, J., Staerfeldt, H.H., Lund, O. & Brunak, S. MatrixPlot: visualizing sequence constraints. *Bioinformatics* **15**, 769-770 (1999).
28. Carter, P.J., Winter, G., Wilkinson, A.J. & Fersht, A.R. The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell* **38**, 835-840 (1984).
29. Mildvan, A.S., Weber, D.J. & Kuliopulos, A. Quantitative interpretations of double mutations of enzymes. *Arch. Biochem. Biophys.* **294**, 327-340 (1992).
30. Horovitz, A. Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold Des.* **1**, R121-126 (1996).
31. Michel, F., Hanna, M., Green, R., Bartel, D.P. & Szostak, J.W. The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* **342**, 391-395 (1989).
32. Yarus, M., Illangesekare, M. & Christian, E. Selection of small molecules by the *Tetrahymena* catalytic center. *Nucleic Acids Res.* **19**, 1297-1304 (1991).
33. Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J. & Strobel, S.A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45-50 (2004).
34. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
35. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319-1325 (2001).
36. Bittker, J.A., Le, B.V. & Liu, D.R. Nucleic acid evolution and minimization by nonhomologous random recombination. *Nat. Biotechnol.* **20**, 1024-1029 (2002).
37. Wang, Q.S. & Unrau, P.J. Ribozyme motif structure mapped using random recombination and selection. *RNA* **11**, 404-411 (2005).
38. Burke, D.H. & Willis, J.H. Recombination, RNA evolution, and bifunctional RNA molecules isolated through chimeric SELEX. *RNA* **4**, 1165-1175 (1998).
39. Moody, E.M. & Bevilacqua, P.C. Folding of a stable DNA motif involves a highly cooperative network of interactions. *J. Am. Chem. Soc.* **125**, 16285-16293 (2003).

40. Moody, E.M., Feerrar, J.C. & Bevilacqua, P.C. Evidence that folding of an RNA tetraloop hairpin is less cooperative than its DNA counterpart. *Biochemistry* **43**, 7992-7998 (2004).
41. Gautheret, D., Damberger, S.H. & Gutell, R.R. Identification of base-triples in RNA using comparative sequence analysis. *J. Mol. Biol.* **248**, 27-43 (1995).
42. Guo, F., Gooding, A.R. & Cech, T.R. Comparison of crystal structure interactions and thermodynamics for stabilizing mutations in the Tetrahymena ribozyme. *RNA* **12**, 387-395 (2006).

FUTURE DIRECTIONS

Here I briefly discuss two experiments related to this thesis that I believe are especially important. A major conclusion of Chapter 1 is that ribozymes with new catalytic activities and folds can be found in the sequence neighborhood of existing ribozymes: kinase ribozymes were found as close as 10 mutations away from the aminoacylase parent ribozyme¹. A question these experiments did not address is whether these kinases can be accessed by neutral mutation of the parent ribozyme. Previous work has shown that the neutral networks of the HDV and Class III ligase ribozymes can closely approach one another, and several sequences (called intersection sequences) were designed that could adopt the folds of both ribozymes, and catalyze the respective reactions at low but detectable levels². Investigating whether a neutral path between the parent and kinase ribozyme 5-16 exists appears especially promising for several reasons. First, the parent and kinase 5-16 only differ at 16 out of 90 positions, and it turns out that a relatively small number of positions need to be changed to design a sequence that can adopt both secondary structures. This task was considerably more difficult for the HDV and Class III ligase ribozymes because the prototype sequences shared only approximately 25% sequence identity. Second, several mutations have been identified that increase the catalytic rate of 5-16 (described in Chapter 2), and these mutations could potentially be used to improve the rate of an intersection sequence that catalyzed aminoacylation efficiently but thiophosphorylation poorly. Third, unlike the reactions catalyzed by the HDV-Class III ligase intersection sequence, which either remove or add sequence to the ribozyme, the self-modification reactions catalyzed by the aminoacylase parent and kinase 5-16 result in relatively minor changes to the ribozyme (either aminoacylation of the 3' terminus or thiophosphorylation of an internal 2'-hydroxyl group). It is

conceivable that under appropriate conditions a single intersection sequence molecule could become aminoacylated and thiophosphorylated, and such a molecule might be a useful tool for studying RNA folding.

A major conclusion of Chapter 2 is that synthetic shuffling³ is an effective way to search sequence space for ribozymes with optimized properties. I believe that this method could be used to optimize the polymerase ribozyme isolated in this laboratory⁴ for several reasons. First, synthetic shuffling is only useful when a ribozyme is modular, and the polymerase is known to consist of at least two domains (the Class I ligase core and the accessory domain). Second, sequence variants that could be used to construct a synthetically pool are already available for both the Class I ligase⁵ and the accessory domain⁴. Third, a long-term goal of the polymerase project is to isolate a ribozyme that can catalyze its own replication. Synthetic shuffling could potentially generate a more efficient polymerase without making the ribozyme longer and thus more difficult to copy.

REFERENCES

1. Curtis, E.A. & Bartel, D.P. New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.* **12**, 994-1000 (2005).
2. Schultes, E.A. & Bartel, D.P. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**, 448-452 (2000).
3. Ness, J.E. *et al.* Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nat. Biotechnol.* **20**, 1251-1255 (2002).
4. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319-1325 (2001).
5. Eklund, E.H. & Bartel, D.P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231-3238 (1995).

APPENDIX

The hammerhead cleavage reaction in monovalent cations

ABSTRACT

Recently, Murray et al.¹ found that the hammerhead ribozyme does not require divalent metal ions for activity if incubated in high (≥ 1 M) concentrations of monovalent ions. We further characterized the hammerhead cleavage reaction in the absence of divalent metal. The hammerhead is active in a wide range of monovalent ions, and the rate enhancement in 4 M Li⁺ is only 20-fold less than that in 10 mM Mg²⁺. Among the Group I monovalent metals, rate correlates in a log-linear manner with ionic radius. The pH dependence of the reaction is similar in 10 mM Mg²⁺, 4 M Li⁺, and 4 M Na⁺. The exchange-inert metal complex Co(NH₃)₆³⁺ also supports substantial hammerhead activity. These results suggest that a metal ion does not act as a base in the reaction, and that the effects of different metal ions on hammerhead cleavage rates primarily reflect structural contributions to catalysis.

INTRODUCTION

Originally identified in the genomes of certain plant viroids and virusoids, the hammerhead ribozyme (Fig. 1) is a small catalytic RNA that cleaves itself at a specific phosphodiester linkage to generate 5' hydroxyl and 2',3'-cyclic phosphate termini.^{2,3} It has been extensively studied in an attempt to better understand RNA catalysis,^{4,6} but despite its small size, the mechanism by which the hammerhead accelerates the cleavage of RNA has proven difficult to elucidate.

Until recently, it was thought that the hammerhead ribozyme required divalent cations for activity, and based on a correlation between hammerhead cleavage rate and metal pK_a , it was suggested that either a solvated metal hydroxide⁷, or metal ion directly coordinated to the 2'-OH at the site of cleavage⁸, acts as a base in the reaction. However, Murray et al.¹ found that high (≥ 1 M) concentrations of monovalent ions can substitute for divalent ions in the hammerhead cleavage mechanism. Because monovalent ions have little effect on the acidity of water molecules to which they are bound, and because the hammerhead cleavage rate in 4 M Li^+ has been reported to be only 30-fold slower than that in 10 mM Mg^{2+} ¹, this is not consistent with the hypothesis that a metal ion acts as a base in the reaction.

To better understand the reaction in the absence of divalent metal, we determined rate enhancements and pH dependence of the cleavage rate in the presence of various monovalent ions, and also examined hammerhead activity in the exchange-inert metal complex $\text{Co}(\text{NH}_3)_6^{3+}$. Our results do not support the idea that a metal ion acts as a base in the reaction, and suggest that the primary role of metal ions in the reaction may be structural rather than catalytic.

RESULTS & DISCUSSION

The rate enhancement in Li^+ approaches that in Mg^{2+}

In comparing the catalytic proficiency of the hammerhead in different metal ions, it is informative to compare not only the ribozyme-catalyzed rates, but also the intrinsic abilities of different metals to cleave RNA. The stability of RNA depends on both the identity and concentration of metal ions present, and rates of background RNA cleavage can differ by more than 1000-fold in different ions^{9,10}. If the non-enzymatic reaction was much faster in 4 M Li^+ than in 10 mM Mg^{2+} , then the idea that divalent metals are crucial for a significant proportion of hammerhead catalysis might be retained. On the other hand, an identical rate enhancement in 4 M Li^+ and in 10 mM Mg^{2+} would support the idea that Li^+ can fully replace Mg^{2+} in hammerhead catalysis and would not be consistent with metal acting as a base in the reaction. With these considerations in mind, we examined the ability of different ions to cleave RNA.

Background rates of RNA cleavage were determined in 4 M Li^+ , Na^+ , K^+ , Rb^+ , Cs^+ , and NH_4^+ , as well as in 10 mM Mg^{2+} . At the hammerhead cleavage site, non-enzymatic rates varied by as much as 20-fold in different monovalent ions (**Table 1**). In 4 M K^+ , rates were comparable to those predicted by Li and Breaker¹⁰, and rates in 10 mM Mg^{2+} were similar to those observed previously for unconstrained RNA linkages¹¹. Rates in K^+ , Rb^+ , and Cs^+ approached the rate in buffer alone (**Table 1**). This result is consistent with metal hydroxide being one of the active species in the non-enzymatic cleavage reaction, since the pK_a 's of these ions¹² appear to differ little from the pK_a of water (15.74¹³). The background rate in Li^+ is 3-fold faster than in Mg^{2+} , further supporting this idea. Although the pK_a of Mg^{2+} (11.4) is lower than that of Li^+ (13.8), the concentration of metal hydroxide in 4 M Li^+ is about 2-fold greater than in 10 mM Mg^{2+}

(calculated using the Henderson-Hasselbach equation). This idea is also consistent with the observation that RNA is unstable in the presence of Pb^{2+} ($\text{pK}_a = 7.2$), Eu^{3+} ($\text{pK}_a = 8.5$), and Zn^{2+} ($\text{pK}_a = 9.6$), all of which have low pK_a 's¹⁴. However, other factors must also be important, because the background rate of RNA cleavage is 40-fold faster in 4 M NH_4^+ than in buffer alone (Table 1). Furthermore, Mg^{2+} has a greater affinity for RNA than does Li^+ ⁹, which would be expected to increase the effective concentration of the Mg^{2+} metal hydroxide, yet the rates correlate with bulk metal hydroxide concentration.

The hammerhead was active in all monovalent ions tested, and rate enhancements ranged from 50,000-fold in Li^+ to 130-fold in Cs^+ (Table 1). The rate enhancement in 4 M Li^+ is only 20-fold less than that in 10 mM Mg^{2+} (Table 1). This comparable rate enhancement demonstrates that the hammerhead is catalytically proficient in monovalent ions alone. The reason for the 20-fold greater rate enhancement in Mg^{2+} is unclear. If the 20-fold difference is due to the pK_a difference between Mg^{2+} and Li^+ , it suggests that the hammerhead is much less sensitive to metal ion pK_a than was previously thought. Hydration number could also be important, as Li^+ ions tend to coordinate fewer water molecules than Mg^{2+} ions do¹⁵. Another possibility is that the hammerhead structure is more efficiently stabilized by the higher charge density of Mg^{2+} .

Correlation between ionic radius and cleavage rate

Among the Group I metals we noted a log-linear relationship between ionic radius and cleavage rate (Fig. 2). We suggest that this relationship reflects the stabilizing effects of different ions on the catalytically active conformation of the hammerhead. The idea that cations could stabilize the catalytically active conformation of the hammerhead relative to that of the ground state is

consistent with both crystallographic¹⁶ and biochemical data^{17,18}, which suggest that the catalytically active and ground state structures are significantly different. Smaller monovalent ions are more efficient than larger ones at stabilizing tRNA, RNA pseudoknots, and RNA dimers, and in some cases the relationship between ionic radius and thermal stability is linear¹⁹⁻²³. In the case of yeast tRNA-Phe, the first unfolding transition occurs 25 °C lower in Cs⁺ than in Li⁺^{19,21}, indicating that the magnitude of this effect can be significant. Also, NH₄⁺ stabilizes tRNA and RNA pseudoknots more efficiently than would be expected based on its size^{21,23}, and we find that the hammerhead cleavage rate in NH₄⁺ (ionic radius = 1.61 Å) is 80-fold faster than that in Rb⁺ (ionic radius = 1.66 Å). Furthermore, differences in the stability of various RNA structures in different monovalent ions are typically much greater than differences in different divalent ions²³. This could explain why hammerhead cleavage rates vary 8000-fold among the Group I monovalent metals (Table 1), but only 130-fold among the Group II divalent metals²⁴.

A possible clue to the mechanism of structural stabilization comes from analysis of a defined metal-binding site in the hammerhead. The affinity of metal ion binding to the G₅ site in the hammerhead correlates in a log-linear manner with the ionic potential [(charge)²/ionic radius] of different monovalent and divalent ions²⁵. We observe a similar correlation between catalytic rate and ionic potential in monovalent ions, suggesting a link between a metal ion's affinity to RNA and its ability to support the transition-state conformation of the hammerhead. Such a correlation between affinity and activity has also been suggested for the reaction in divalent ions²⁶. For the G₅ site, the slope of the line relating affinity and ionic potential is 0.6²⁵. For cleavage in Group I monovalent ions, the slope of an analogous graph (not shown) is seven times steeper, suggesting involvement of numerous additional metal-RNA interactions in the transition state conformation compared to those of the ground state. The structural basis for these

additional interactions could be a conformational change similar to that proposed by Peracchi et al.¹⁷.

Hammerhead activity in $\text{Co}(\text{NH}_3)_6^{3+}$

To further explore the role played by metal ions in the hammerhead cleavage reaction, we tested ribozyme activity in the presence of $\text{Co}(\text{NH}_3)_6^{3+}$. The ammine ligands in this complex are exchange-inert over the time scale of the reaction, so that rates measured under these conditions reflect catalysis in the absence of inner-sphere coordination between metal ions and ribozyme²⁷⁻³⁰. It was previously reported that in 500 μM Mn^{2+} the hammerhead is inhibited by $\text{Co}(\text{NH}_3)_6^{3+}$ ³¹ and that it is inactive in $\text{Co}(\text{NH}_3)_6^{3+}$ alone²⁸. We also found that the hammerhead is inactive in low concentrations of $\text{Co}(\text{NH}_3)_6^{3+}$ (≤ 10 mM), but between 10 mM and 100 mM the cleavage rate increases dramatically. In 100 mM $\text{Co}(\text{NH}_3)_6^{3+}$ the hammerhead cleavage rate is 0.0071 min^{-1} , and the rate enhancement is 7,000-fold. The hammerhead does not appear to be saturated under these conditions, but the solubility of $\text{Co}(\text{NH}_3)_6^{3+}$ prevented us from using higher concentrations. Rates did not change in the presence of 25 mM EDTA, indicating that activity is not due to divalent metal contamination. These results show that millimolar concentrations of cations can support significant levels of hammerhead activity in the absence of inner-sphere coordination.

Hammerhead activity in $\text{Co}(\text{NH}_3)_6^{3+}$ is intriguing, because two important metal-binding ligands believed to involve inner-sphere coordination have been identified by phosphorothioate experiments^{17,18,24,32,33}. One possible explanation for this activity is that divalent metal ions and $\text{Co}(\text{NH}_3)_6^{3+}$ make outer-sphere rather than inner-sphere contacts at these sites. This effects and apparent thiophilic metal rescue of these effects have been reported at sites of outer-sphere contacts³⁴. Another possibility is that divalent metal ions bind at these sites by inner-sphere

coordination, but $\text{Co}(\text{NH}_3)_6^{3+}$ makes outer-sphere contacts instead. For example, although Tb^{3+} appears to make three inner-sphere contacts at the G_5 binding site in the hammerhead, it can be displaced by $\text{Co}(\text{NH}_3)_6^{3+}$ ²⁵. A third possibility, supported by recent phosphorothioate experiments³⁵, is that the 310-fold difference in rate and the 170-fold difference in rate enhancement between the reactions in Mg^{2+} and $\text{Co}(\text{NH}_3)_6^{3+}$ are due to the inability of $\text{Co}(\text{NH}_3)_6^{3+}$ to productively bind at these sites. If this is true, then there must also be a compensatory effect of high concentrations of $\text{Co}(\text{NH}_3)_6^{3+}$, because thio effects at several positions in the hammerhead are much greater than 310-fold. For example, the thio effect at the pro- R_p oxygen at the scissile phosphate is 2000- to 80,000 fold^{32,33}.

The pH dependence of the cleavage rate in monovalent ions

The finding that the rate enhancement in Li^+ approaches that in Mg^{2+} , and that the hammerhead retains significant activity in $\text{Co}(\text{NH}_3)_6^{3+}$, weakens the case for a metal ion acting as a base in the reaction. Furthermore, the relationship between ionic radius and cleavage rate suggests that the differences among ions reflect their differential abilities in stabilizing the catalytically active conformation of the hammerhead. This contrasts to the picture emerging for the hepatitis delta virus (HDV) ribozyme, a different ribozyme that also carries out site-specific cleavage leaving 5' hydroxyl and 2',3'-cyclic phosphate termini³⁶. For the HDV ribozyme a divalent metal ion is thought to act as a base in the cleavage reaction: removal of the divalent metal reduces the rate significantly (5,000-fold at pH 7), and exposes the underlying acid catalysis, as indicated by an inversion of the pH dependence³⁷. With this in mind, we investigated the pH dependence of HH16.1 in monovalent and divalent metals.

In 10 mM Mg^{2+} , the hammerhead cleavage rate increases approximately 10-fold with each one unit increase in pH^7 (Fig. 3). We observed a similar pH dependence in 4 M Li^+ and in 4 M Na^+ (Fig. 4). For both monovalent and divalent cations, it appears there is a single deprotonation prior to the rate-limiting step, most likely deprotonation of the 2'-OH at the site of cleavage^{7,8,38}. Throughout this pH range, deprotonation of the 2'-OH might be only about sevenfold greater in Mg^{2+} than in Li^+ .

The pH-rate profiles can be added to the list of similarities between the reaction in monovalent ions compared with that in divalent ions, which includes requirement for a 2'-OH at the site of cleavage¹, formation of a 2',3'-cyclic phosphate product¹, requirement for G_5 in the conserved core of the ribozyme¹, requirement for numerous other residues and functional groups in the hammerhead core³⁵, activity in a wide range of ions (Table 1), and similar rate enhancements (Table 1). These similarities suggest that monovalent and divalent ions play essentially the same roles in hammerhead catalysis. Regarding the nature of the base in the hammerhead reaction, this suggests two possibilities: that both monovalent and divalent ions can act as bases in the reaction, or that neither can act as bases in the reaction. The latter possibility appears more likely for several reasons. First, the pK_a of Li^+ (13.8) is considerably higher than that of Mg^{2+} (11.4), and is also higher than the 2'-OH of ribose (13.1¹⁰), yet the rates and rate enhancements are similar in Li^+ and Mg^{2+} . The hammerhead also retains significant activity in $\text{Co}(\text{NH}_3)_6^{3+}$, arguing against the role of a solvated metal hydroxide, or a metal directly coordinated to the 2'-OH at the site of cleavage, as a base in the reaction. Finally, the relationship between ionic radius and cleavage rate suggests that a primary role played by metal ions in the hammerhead cleavage reaction may be structural.

What is the base in the hammerhead cleavage reaction?

If a metal ion does not act as a base in the hammerhead cleavage reaction, what does? One possibility is that a nucleotide base in the ribozyme core functions in this manner. Work by Ferre-D'Amare et al.³⁹, Perrotta et al.⁴⁰, and Nakano et al.³⁷ suggests that such a role is possible for bases in the HDV ribozymes. Furthermore, recent experiments indicate that a conserved adenosine in domain V of 23S rRNA might act as a general acid-base catalyst in the peptidyl transferase reaction of the ribosome⁴¹⁻⁴³. We tested the hypothesis that G₅ acts as a base in the hammerhead cleavage reaction but our results were inconclusive. Another possibility is that the pK_a of the 2'-OH at the site of cleavage is perturbed by its environment, defined by the three dimensional structure of the hammerhead. Several examples of perturbed pK_a values in RNA structures have been reported^{40,43-47}. A third possibility is that the hammerhead does not accelerate deprotonation of the 2'-OH at all, and instead relies on solution levels of OH⁻. But regardless, since the rate of a well behaved hammerhead (10 min⁻¹ at pH 8.5) is at least 450-fold faster than the rate of uncatalyzed RNA cleavage when the attacking 2'-hydroxyl is fully deprotonated (0.022 min⁻¹), the hammerhead must employ additional catalytic strategies¹⁰.

In summary, our results suggest that monovalent and divalent ions play essentially the same roles in the hammerhead cleavage reaction. Inner-sphere coordination is not required for a substantial fraction of the hammerhead rate enhancement. Furthermore, a solvated metal hydroxide, or a metal ion directly coordinated to the 2'-OH at the site of cleavage, does not appear to accelerate deprotonation at the site of hammerhead cleavage. This does not rule out a catalytic role for a metal ion in the reaction. For example, a fully hydrated metal ion could still provide electrostatic neutralization of developing negative charge in the transition state⁴⁸. An interesting alternative to a metal ion acting as a base in the reaction is that a core nucleotide plays

a direct role in the cleavage chemistry, but further experiments will be needed to explore this hypothesis.

Figure 1 Secondary structure of hammerhead HH16.1. Conserved nucleotides are numbered, and the cleavage site is indicated by an arrow.

Figure 2 Relationship between the ionic radius of Group I monovalent metals and hammerhead-catalyzed (squares) and uncatalyzed (circles) RNA cleavage rates. Ionic radii are from Cotton et al.⁴⁹. All rates were measured at pH 8.0.

Figure 3 The pH dependence of the hammerhead cleavage rate in 10 mM Mg²⁺ (triangles), 4 M Li⁺ (circles), and 4 M Na⁺ (squares). Slopes of the lines are 1.1 in Mg²⁺, 0.90 in Li⁺, and 0.97 in Na⁺.

Figure 4 Effect of urea on the hammerhead cleavage rate in 10 mM Mg²⁺ (circles) and in 4 M Li⁺ (squares). All rates were measured at pH 8.0.

Table 1 Cleavage rates in monovalent and divalent ions at pH 8.0.

Figure 1

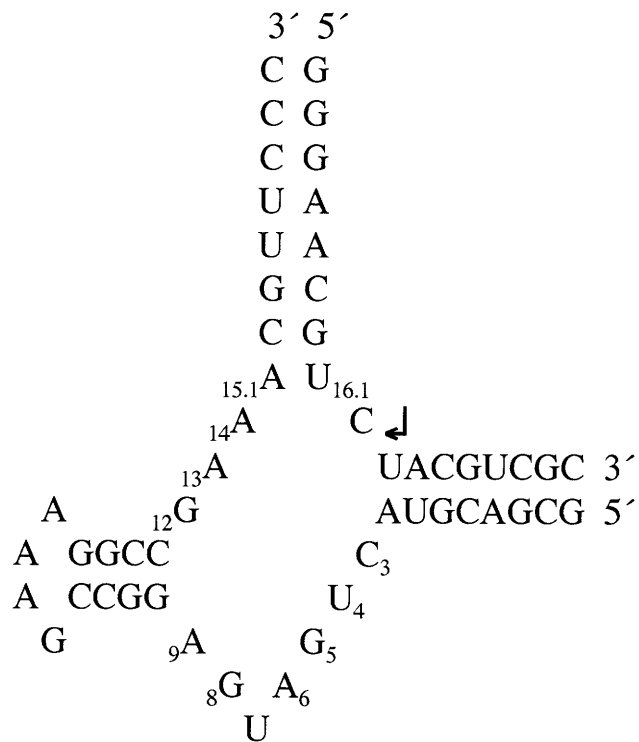


Figure 2

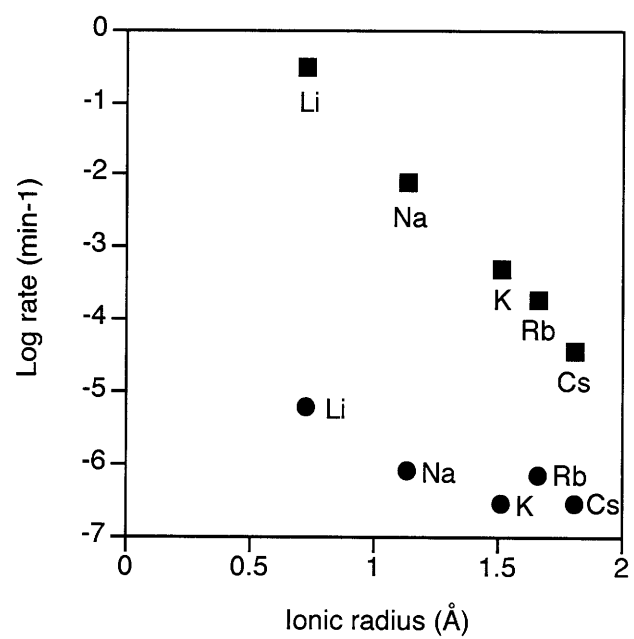


Figure 3

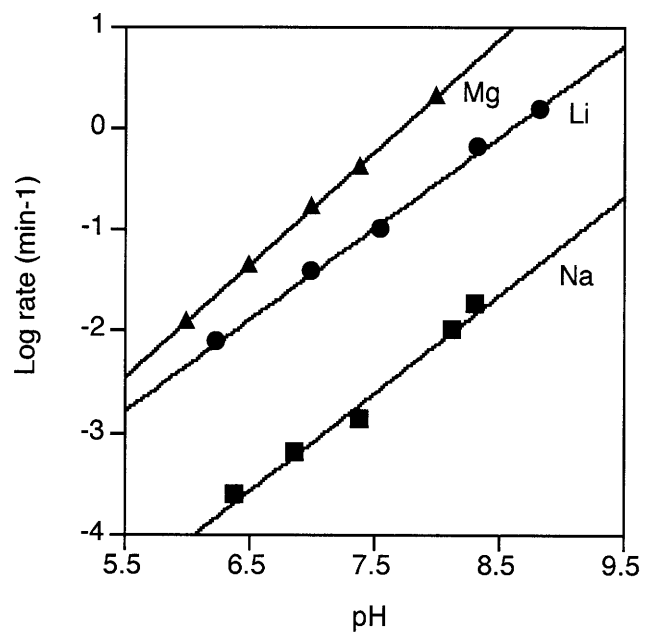


Figure 4

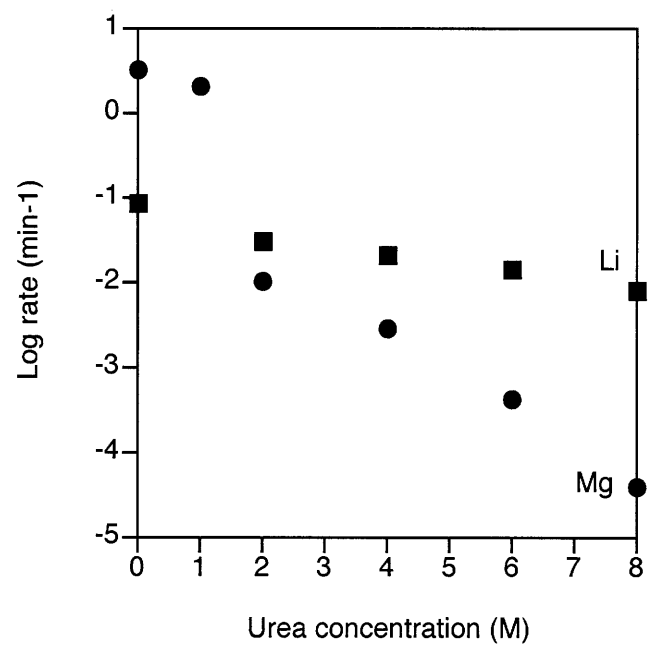


Table 1

Cation	Uncatalyzed rate (min ⁻¹)	Catalyzed rate (min ⁻¹)	Rate enhancement
10 mM MgCl ₂	1.9×10^{-6}	2.2	1.2×10^6
4 M LiCl	6×10^{-6}	2.9×10^{-1}	5×10^4
4 M NaCl	8×10^{-7}	7.5×10^{-3}	9×10^3
4 M KCl	2.7×10^{-7}	4.7×10^{-4}	1.7×10^3
4 M RbCl	7×10^{-7}	1.8×10^{-4}	2.6×10^2
4 M CsCl	2.8×10^{-7}	3.7×10^{-5}	1.3×10^2
4 M NH ₄ Cl	5×10^{-6}	1.4×10^{-2}	2.8×10^3
None	1.3×10^{-7}	not determined	not determined

METHODS

The initial report describing hammerhead activity in high monovalent salt used the hammerhead HH16.1 construct¹. Since we wished to build on these findings, HH16.1⁵⁰ (**Fig. 1**) was used for all experiments. RNA was transcribed from gel-purified DNA templates using T7 RNA polymerase. Following alkaline phosphatase treatment, substrate RNA was 5'-radiolabeled using T4 polynucleotide kinase and γ -³²P ATP. RNA was purified on denaturing polyacrylamide gels. RNA concentrations were determined by optical density⁵¹.

Because 4 M approaches the solubility limit of some of the salts used in this study, reactions were initiated using a modified protocol. Enzyme and substrate strand were combined in 4 μ l, incubated at 95°C for 2 minutes and cooled to 25°C over 5 minutes. Then, samples were dried using a SpeedVac Concentrator and reactions initiated with a solution containing 4 M monovalent salt, 50 mM buffer, and 25 mM EDTA. Rates in divalent metal were measured in 10 mM Mg²⁺ and 50 mM buffer. Final concentrations of enzyme and substrate strands were 0.65-7.3 μ M and <0.088 μ M, respectively. Rates did not change over the ten-fold range in ribozyme concentration used, confirming that ribozyme was saturating under these conditions. For each time point, a 1 μ l aliquot was removed and quenched in 20 μ l of a stop solution containing 8 M urea, 25 mM EDTA, and placed on dry ice. Product and substrate were separated on 20% denaturing polyacrylamide gels, and quantitated on a Fujix Phosphorimager using MacBAS and Image Reader software. To calculate rates, fraction reacted was plotted against time, and, depending on the extent of the reaction, was fitted to either a line or to equation (1),

$$\text{Fraction reacted} = F(1 - e^{-kt}) \quad (1)$$

where F is the maximum fraction reacted (typically 0.9), k is the observed rate constant, and t is time. All rates were measured at least twice, and independent determinations differed by less than 15 percent for catalyzed rates, and less than 2-fold for non-enzymatic cleavage rates.

To confirm that accurate rates could be determined using our modified protocol, some rates were also determined using a standard annealing protocol⁶. The two protocols yielded indistinguishable values, and these values were similar to those previously reported^{1,50}. One difference between our modified protocol and the standard protocol is that, when using the modified protocol, a small amount of cleavage (usually about 5%) is typically observed during the evaporation. Such cleavage has been observed by others⁵², and, as mentioned above, we confirmed that it had no effect on observed rates. We also note that reactions in 4 M salt were not effectively terminated when 10 volumes of stop solution were used to quench the reaction. A primary role of the stop solution was to reduce the concentration of monovalent salt by dilution, because urea does not effectively denature the hammerhead in the presence of high concentrations of monovalent salt. For example, in 4 M Li^+ the reaction rate is reduced only about 10-fold in 8 M urea, while in 10 mM Mg^{2+} the reaction rate is reduced about 200-fold in 2 M urea, and about 60,000-fold in 8 M urea (Fig. 4). Consequently, we stopped reactions by diluting in 20 volumes of stop solution and freezing in dry ice. Background rates of RNA cleavage were measured in the same way as ribozyme-catalyzed rates, but in the absence of enzyme-strand RNA. A ladder of cleavage products was observed when time points were run on denaturing polyacrylamide gels, and rate constants were calculated for the hammerhead cleavage site as well as 8 neighboring phosphodiester linkages. For each metal, cleavage rates at different linkages systematically varied up to 12-fold, but usually no more than 5-fold. The background rates at the hammerhead cleavage site were representative of rates at neighboring linkages.

The buffers MES (pH 6–6.5), BES (pH 6.3–7.5), MOPS (pH 7), and Tris (pH 7.5–8.8) were used to determine rates at different pH's. Experiments using different buffers at the same pH indicated that changing buffers did not affect rate. Because high ionic strength can affect buffers, pH's were adjusted using HCl or the appropriate metal hydroxide after adding monovalent salt. Values were determined on a Beckman ϕ 200 pH Meter with a Futura Refillable Micro Calomel Combination pH Electrode, and were consistent with those determined by indicator dyes on pH paper.

ACKNOWLEDGMENTS

We thank M. Been as well as M. Lawrence and other members of the lab for helpful discussions.

This work was supported by a grant from the N.I.H.

REFERENCES

1. Murray, J.B., Seyhan, A.A., Walter, N.G., Burke, J.M. & Scott, W.G. The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem. Biol.* **5**, 587-595 (1998).
2. Hutchins, C.J., Rathjen, P.D., Forster, A.C. & Symons, R.H. Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res.* **14**, 3627-3640 (1986).
3. Forster, A.C. & Symons, R.H. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell* **49**, 211-220 (1987).
4. Thomson, J.B., Tuschl, T. & Eckstein, F. The hammerhead ribozyme. in *Nucleic acids and molecular biology: Catalytic RNA* (eds. Eckstein, F. & Lilley, D.M.J.) 173-196 (Springer-Verlag, Berlin, 1996).
5. McKay, D.B. Structure and function of the hammerhead ribozyme: an unfinished story. *RNA* **2**, 395-403 (1996).
6. Stage-Zimmermann, T.K. & Uhlenbeck, O.C. Hammerhead ribozyme kinetics. *RNA* **4**, 875-889 (1998).
7. Dahm, S.C., Derrick, W.B. & Uhlenbeck, O.C. Evidence for the role of solvated metal hydroxide in the hammerhead cleavage mechanism. *Biochemistry* **32**, 13040-13045 (1993).
8. Sawata, S., Komiyama, M. & Taira, K. Kinetic evidence based on solvent isotope effects for the nonexistence of a proton-transfer process in reactions catalyzed by a hammerhead ribozyme: Implications to the double-metal-ion mechanism of catalysis. *J. Am. Chem. Soc.* **117**, 2357-2358 (1995).
9. Kazakov, S.A. Nucleic acid binding and catalysis by metal ions. in *Bioorganic chemistry: Nucleic Acids* (ed. Hecht, S.M.) 244-287 (Oxford University Press, New York, 1996).
10. Li, Y. & Breaker, R.R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. *J. Am. Chem. Soc.* **121**, 5364-5372 (1999).
11. Soukup, G.A. & Breaker, R.R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308-1325 (1999).
12. Burgess, J. *Ions in solution: basic principles of chemical interactions*, (John Wiley & Sons, New York, 1988).
13. Jencks, W.P. *Catalysis in chemistry and enzymology*, (McGraw Hill, New York, 1969).
14. Ciesiolka, J., Michalowski, D., Wrzesinski, J., Krajewski, J. & Krzyzosiak, W.J. Patterns of cleavages induced by lead ions in defined RNA secondary structure motifs. *J. Mol. Biol.* **275**, 211-220 (1998).
15. Feig, A.L. & Uhlenbeck, O.C. The role of metal ions in RNA biochemistry. in *The RNA world* (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 287-319 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
16. Murray, J.B., Szoke, H., Szoke, A. & Scott, W.G. Capture and visualization of a catalytic RNA enzyme-product complex using crystal lattice trapping and X-ray holographic reconstruction. *Mol. Cell* **5**, 279-287 (2000).
17. Peracchi, A., Beigelman, L., Scott, E.C., Uhlenbeck, O.C. & Herschlag, D. Involvement of a specific metal ion in the transition of the hammerhead ribozyme to its catalytic conformation. *J. Biol. Chem.* **272**, 26822-26826 (1997).

18. Wang, S., Karbstein, K., Peracchi, A., Beigelman, L. & Herschlag, D. Identification of the hammerhead ribozyme metal ion binding site responsible for rescue of the deleterious effect of a cleavage site phosphorothioate. *Biochemistry* **38**, 14363-14378 (1999).
19. Urbanke, C., Romer, R. & Maass, G. Tertiary structure of tRNA^{Phe} (yeast): kinetics and electrostatic repulsion. *Eur. J. Biochem.* **55**, 439-444 (1975).
20. Labuda, D. & Augustyniak, J. Dependence of tRNA structure in solution upon ionic condition of the solvent. Fluorescence studies of monovalent cation binding to tRNA^{Phe} from barley embryos. *Eur. J. Biochem.* **79**, 303-307 (1977).
21. Heerschap, A., Walters, J.A. & Hibers, C.W. Interactions of some naturally occurring cations with phenylalanine and initiator tRNA from yeast as reflected by their thermal stability. *Biophys. Chem.* **22**, 205-217 (1985).
22. Torrent, C., Bordet, T. & Darlix, J.L. Analytical study of rat retrotransposon VL30 RNA dimerization in vitro and packaging in murine leukemia virus. *J. Mol. Biol.* **240**, 434-444 (1994).
23. Gluick, T.C., Wills, N.M., Gesteland, R.F. & Draper, D.E. Folding of an mRNA pseudoknot required for stop codon readthrough: effects of mono- and divalent ions on stability. *Biochemistry* **36**, 16173-16186 (1997).
24. Dahm, S.C. & Uhlenbeck, O.C. Role of divalent metal ions in the hammerhead RNA cleavage reaction. *Biochemistry* **30**, 9464-9469 (1991).
25. Feig, A.L., Panek, M., Horrocks, W.D., Jr. & Uhlenbeck, O.C. Probing the binding of Tb(III) and Eu(III) to the hammerhead ribozyme using luminescence spectroscopy. *Chem. Biol.* **6**, 801-810 (1999).
26. Hunsicker, L.M. & DeRose, V.J. Activities and relative affinities of divalent metals in unmodified and phosphorothioate-substituted hammerhead ribozymes. *J. Inorg. Biochem.* **80**, 271-281 (2000).
27. Hampel, A. & Cowan, J.A. A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. *Chem. Biol.* **4**, 513-517 (1997).
28. Nesbitt, S., Hegg, L.A. & Fedor, M.J. An unusual pH-independent and metal-ion-independent mechanism for hairpin ribozyme catalysis. *Chem. Biol.* **4**, 619-630 (1997).
29. Young, K.J., Gill, F. & Grasby, J.A. Metal ions play a passive role in the hairpin ribozyme catalysed reaction. *Nucleic Acids Res.* **25**, 3760-3766 (1997).
30. Suga, H., Cowan, J.A. & Szostak, J.W. Unusual metal ion catalysis in an acyl-transferase ribozyme. *Biochemistry* **37**, 10118-10125 (1998).
31. Horton, T.E. & DeRose, V.J. Cobalt hexamine inhibition of the hammerhead ribozyme. *Biochemistry* **39**, 11408-11416 (2000).
32. Scott, E.C. & Uhlenbeck, O.C. A re-investigation of the thio effect at the hammerhead cleavage site. *Nucleic Acids Res.* **27**, 479-484 (1999).
33. Derrick, W.B., Greef, C.H., Caruthers, M.H. & Uhlenbeck, O.C. Hammerhead cleavage of the phosphorodithioate linkage. *Biochemistry* **39**, 4947-4954 (2000).
34. Basu, S. & Strobel, S.A. Thiophilic metal ion rescue of phosphorothioate interference within the *Tetrahymena* ribozyme P4-P6 domain. *RNA* **5**, 1399-1407 (1999).
35. O'Rear, J.L. *et al.* Comparison of the hammerhead cleavage reactions stimulated by monovalent and divalent cations. *RNA* **7**, 537-545 (2001).
36. Been, M.D. & Wickham, G.S. Self-cleaving ribozymes of hepatitis delta virus RNA. *Eur. J. Biochem.* **247**, 741-753 (1997).

37. Nakano, S., Chadalavada, D.M. & Bevilacqua, P.C. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**, 1493-1497 (2000).
38. Kuimelis, R.G. & McLaughlin, L.W. Hammerhead ribozyme-mediated cleavage of a substrate analogue containing an internucleotide bridging 5'-phosphorothioate: Implications for the cleavage mechanism and the catalytic role of the metal cofactor. *J. Am. Chem. Soc.* **117**, 11019-11020 (1995).
39. Ferre-D'Amare, A.R., Zhou, K. & Doudna, J.A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567-574 (1998).
40. Perrotta, A.T., Shih, I. & Been, M.D. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science* **286**, 123-126 (1999).
41. Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905-920 (2000).
42. Nissen, P., Hansen, J., Ban, N., Moore, P.B. & Steitz, T.A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-930 (2000).
43. Muth, G.W., Ortoleva-Donnelly, L. & Strobel, S.A. A single adenosine with a neutral pKa in the ribosomal peptidyl transferase center. *Science* **289**, 947-950 (2000).
44. Legault, P. & Pardi, A. In-situ probing of adenine protonation in RNA by C-13 NMR. *J. Am. Chem. Soc.* **116**, 8390-8391 (1994).
45. Connell, G.J. & Yarus, M. RNAs with dual specificity and dual RNAs with similar specificity. *Science* **264**, 1137-1141 (1994).
46. Legault, P. & Pardi, A. Unusual dynamics and pKa shift at the active site of a lead-dependent ribozyme. *J. Am. Chem. Soc.* **119**, 6621-6628 (1997).
47. Narlikar, G.J. & Herschlag, D. Mechanistic aspects of enzymatic catalysis: lessons from comparison of RNA and protein enzymes. *Annu. Rev. Biochem.* **66**, 19-59 (1997).
48. Cowan, J.A. Metal Activation of Enzymes in Nucleic Acid Biochemistry. *Chem. Rev.* **98**, 1067-1088 (1998).
49. Cotton, F.A., Wilkinson, G., Murillo, C.A. & Bochmann, M. *Advanced inorganic chemistry*, (John Wiley & Sons, New York, 1999).
50. Clouet-d'Orval, B. & Uhlenbeck, O.C. Hammerhead ribozymes with a faster cleavage rate. *Biochemistry* **36**, 9087-9092 (1997).
51. Dawson, R.M.C., Elliott, D.C., Elliott, W.H. & Jones, K.M. *Data for biochemical research*, (Clarendon Press, Oxford, 1986).
52. Seyhan, A.A. & Burke, J.M. Mg²⁺-independent hairpin ribozyme catalysis in hydrated RNA films. *RNA* **6**, 189-198 (2000).