RNA (2001), *7*:395–404. Cambridge University Press. Printed in the USA.

The effect of cytidine on the structure and function of an RNA ligase ribozyme

JEFF ROGERS and GERALD F. JOYCE

Departments of Chemistry and Molecular Biology and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, California 92037, USA

ABSTRACT

A cytidine-free ribozyme with RNA ligase activity was obtained by in vitro evolution, starting from a pool of random-sequence RNAs that contained only guanosine, adenosine, and uridine. This ribozyme contains 74 nt and catalyzes formation of a 3′,5′-phosphodiester linkage with a catalytic rate of 0.016 min^{-1}. The RNA adopts a simple secondary structure based on a three-way junction motif, with ligation occurring at the end of a stem region located several nucleotides away from the junction. Cytidine was introduced to the cytidine-free ribozyme in a combinatorial fashion and additional rounds of in vitro evolution were carried out to allow the molecule to adapt to this added component. The resulting cytidine-containing ribozyme formed a 3′,5′ linkage with a catalytic rate of 0.32 min^{-1}. The improved rate of the cytidine-containing ribozyme was the result of 12 mutations, including seven added cytidines, that remodeled the internal bulge loops located adjacent to the three-way junction and stabilized the peripheral stem regions.

Keywords: in vitro evolution; ribozyme; RNA ligase

INTRODUCTION

There are several known ribozymes that catalyze the template-directed ligation of oligonucleotide substrates. These include naturally occurring ribozymes, such as the hammerhead and hairpin (Hegg & Fedor, 1995; Hertel & Uhlenbeck, 1995), and ribozymes that were obtained through in vitro evolution (Bartel & Szostak, 1993; Hager & Szostak, 1997; Jaeger et al., 1999; Landweber & Pokrovskaya, 1999; Robertson & Ellington, 2000). Most in vitro evolution experiments have focused on RNA molecules that catalyze attack of either the 2′- or 3′-hydroxyl of an oligonucleotide substrate on the 5′-triphosphate of the ribozyme, forming a 2′,5′- or 3′,5′-phosphodiester linkage, respectively. The most proficient and best studied of the in vitro evolved ligase ribozymes is the class I motif (Ekland et al., 1995). It forms a 3′,5′-phosphodiester linkage with a catalytic rate of ~100 min^{-1}. The class I ligase has been adapted to catalyze the template-directed polymerization of nucleoside triphosphates (Ekland & Bartel, 1996) and has been made to undergo continuous in vitro evolution (Wright & Joyce, 1997).

Reprint requests to: Gerald F. Joyce, Departments of Chemistry and Molecular Biology and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA; e-mail: gjoyce@scripps.edu.

The product of the first continuous in vitro evolution experiment was the E100 ligase. This molecule contains 29 mutations relative to the class I ligase and is able to ligate a chimeric DNA/RNA substrate that contains the sequence of the T7 RNA polymerase promoter element (Wright & Joyce, 1997). The E100 ligase was used as a starting point to develop a ligase ribozyme that completely lacks cytidine (Rogers & Joyce, 1999). Cytidine was chosen for elimination because of its tendency to undergo spontaneous deamination to uridine, which has led some to suggest that cytidine was not present in the first genetic material (Levy & Miller, 1998). In developing the cytidine-free ligase, the rate of cytidine deamination was greatly accelerated by treatment with sodium bisulfite and CTP was withheld from the in vitro transcription mixture. After 24 successive rounds of in vitro selective amplification, a cytidine-free ligase was obtained that formed a 2′,5′-phosphodiester linkage with a catalytic rate of 0.01 min^{-1}. This is about 10^5-fold faster than the uncatalyzed rate of reaction (Rohatgi et al., 1996a), but about 10^3-fold slower than the rate of the E100 ligase. The cytidine-free ligase adopted a completely different secondary structure compared to the class I ligase and formed a 2′,5′ rather than a 3′,5′ linkage (Rogers & Joyce, 1999).

The reduced catalytic rate and altered regiospecificity of the cytidine-free ligase might be a reflection of

limited evolutionary opportunities that were available starting from the E100 ligase. To evaluate this possibility, a second in vitro evolution experiment was carried out starting from a large pool of random-sequence, cytidine-free RNAs. This resulted in the isolation of additional cytidine-free ligases, including the "R3" motif, which catalyzes formation of a 3′,5′-phosphodiester linkage with a catalytic rate of approximately 0.02 min^{-1}. This rate is still much slower than that of the class I ligase, suggesting that the lack of cytidine is a general and substantial impediment to RNA-based catalytic function. On the other hand, RNA-catalyzed ligation to form a 3′,5′-phosphodiester linkage can be achieved in the absence of cytidine.

The goal of the present study was to assess the contribution of cytidine to RNA structure and function, employing the R3 ligase as a model system. Chemical probing, 3′-terminal deletion analysis, and site-directed mutagenesis studies were used to define the secondary structure of this ribozyme. Cytidine then was introduced to the molecule in a combinatorial fashion and additional rounds of in vitro evolution were carried out to determine how this added component might be exploited. The resulting cytidine-containing ribozyme exhibited a 20-fold improvement in catalytic rate compared to the R3 ligase, while retaining the same overall secondary structure and the same regiospecificity of reaction. The improved rate of the cytidine-containing compared to cytidine-free ligase was attributable to stabilization and subtle refinement of the molecule's secondary structure.

RESULTS

In vitro evolution and screening

In vitro evolution was carried out starting with a pool of 10^{14} different RNA molecules, each containing two regions of 64 nt that were randomly chosen from G, A, and U. The two randomized regions were separated by the fixed sequence 5′-UAUAGUGAG-3′, which was complementary to the 3′ end of the oligonucleotide substrate. All of the RNAs in the pool contained two G residues at their 5′ end to facilitate transcription by T7 RNA polymerase, and the sequence 5′-AGGUUG-AGAAGGGUUAGG-3′ at their 3′ end to serve as a primer binding site for cDNA synthesis and PCR amplification. The other primer binding site for PCR amplification was provided by the oligonucleotide substrate, so that only those RNA molecules that had undergone RNA-catalyzed ligation would contain both primer binding sites and be eligible for amplification. The substrate was a 34mer DNA/RNA molecule that had a single riboadenosine at its 3′ end and contained the sequence of the T7 RNA polymerase promoter element. This allowed the products of RNA-catalyzed ligation and subsequent selective amplification to be transcribed directly to yield "progeny" molecules that were used to begin the next round of selective amplification.

RNA-catalyzed RNA ligation was carried out in the presence of ~1 μM ribozyme, 5 μM substrate, 25 mM $MgCl_2$, 50 mM KCl, and 2 mM spermidine at pH 8.5 and 23 °C. The ligated products were isolated by denaturing polyacrylamide gel electrophoresis, eluted from the gel, reverse transcribed, PCR amplified, then forward transcribed. The time for the reaction was decreased progressively over successive rounds, beginning with 18 h in the first round, then 3.5 h in the second round, 30 min in the third round, 30 s in the fourth round, and 10 s in the fifth round. Following the fifth round, individual molecules were cloned from the population, sequenced, and tested for their catalytic activity.

A screen was developed to determine whether individual ribozymes were capable of forming either a 2′,5′- or 3′,5′-phosphodiester linkage. Unlabeled ribozyme was allowed to react with [5′-^{32}P]-labeled substrate and the ligated product was isolated by polyacrylamide gel electrophoresis. The product then was digested with RNase T2, which cleaves 3′,5′- but not 2′,5′-phosphodiester linkages of RNA and is unable to cleave DNA. Because the substrate contained only a single RNA residue at its 3′ end, the labeled product of RNase T2 digestion would be the same length as the substrate if a 3′,5′ linkage had formed and 1 nt longer if a 2′,5′ linkage had formed. Of seven clones that were surveyed, one formed a 3′,5′ linkage and six formed a 2′,5′ linkage.

Starting with the ribozyme that was capable of forming a 3′,5′-phosphodiester linkage, a pool of cytidine-free variants was prepared by chemical synthesis, introducing random mutations at a frequency of 8% per nucleotide position. Five additional rounds of in vitro evolution were carried out, with reaction times of 1 min in the first round and 15 s in the second through fifth rounds. Individual molecules again were cloned from the population, sequenced, and tested for catalytic activity. Of nine clones that were surveyed, all formed a 3′,5′-phosphodiester linkage. The most active clone, designated R3, contained three mutations relative to the molecule that had been randomized following the first five rounds of in vitro evolution.

Deletion analysis was used to define the shortest 3′-truncated form of the R3 ribozyme that retained full catalytic activity. Unlabeled ribozyme was subjected to partial alkaline hydrolysis, then allowed to react with [5′-^{32}P]-labeled substrate. The reaction products were separated by high-resolution polyacrylamide gel electrophoresis in comparison to size markers. This defined the minimum length of the fully active ribozyme as 74 nt (Fig. 1). The 74mer ribozyme had a twofold faster catalytic rate compared to the full-length ribozyme. Formal kinetic analyses employed a 12mer oligonucleotide substrate that contained four ribonucleotides at its 3′ end and had a

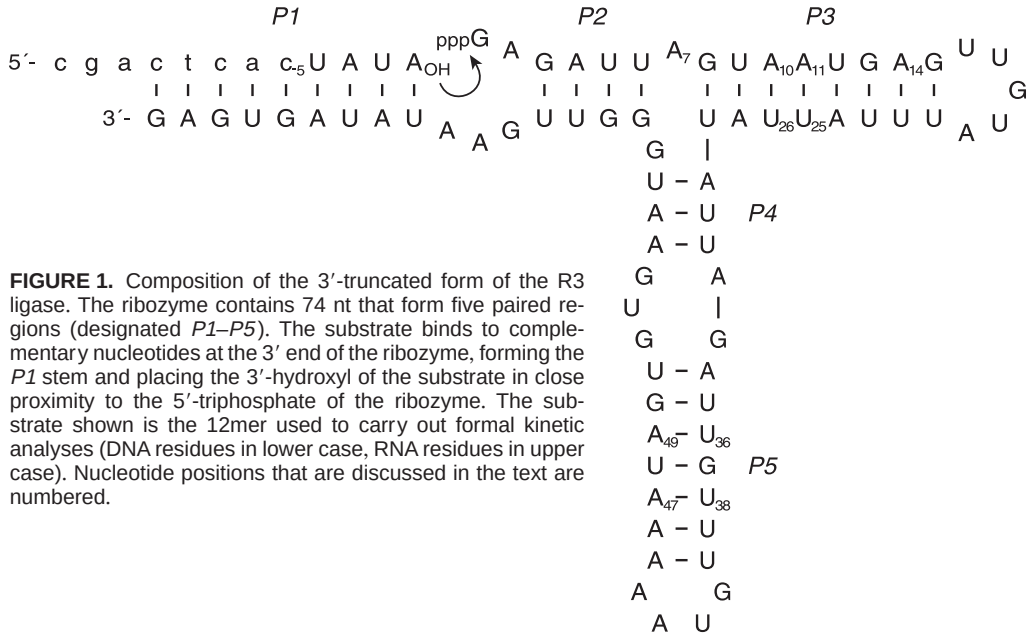

FIGURE 1. Composition of the 3′-truncated form of the R3 ligase. The ribozyme contains 74 nt that form five paired regions (designated *P1–P5*). The substrate binds to complementary nucleotides at the 3′ end of the ribozyme, forming the *P1* stem and placing the 3′-hydroxyl of the substrate in close proximity to the 5′-triphosphate of the ribozyme. The substrate shown is the 12mer used to carry out formal kinetic analyses (DNA residues in lower case, RNA residues in upper case). Nucleotide positions that are discussed in the text are numbered.

sequence corresponding to the 3′ portion of the substrate that was used during in vitro evolution. The 74mer ribozyme ligated the 12mer substrate with a k_{cat} of 0.016 min^{-1} and apparent K_m of 0.2 μM, measured under single-turnover conditions (Fig. 2A).

Two methods were used to confirm the 3′,5′-regiospecificity of the 3′-truncated version of the R3 ligase. First, as before, the ribozyme was allowed to react with a [5′-^{32}P]-labeled substrate that contained a single ribonucleotide at its 3′ end. The ligation product was isolated by polyacrylamide gel electrophoresis, cleaved with RNase T2, and compared to authentic materials that were obtained from ligase ribozymes that are known to form either a 2′,5′- or 3′,5′-phosphodiester linkage (Fig. 3A). The second method involved reaction of the ribozyme with an all-RNA substrate, followed by digestion of the purified ligated product with the 8–17 DNA enzyme (Santoro & Joyce, 1997). The DNA enzyme cleaves 3′,5′- but not 2′,5′-phosphodiester linkages, and when directed to cleave the ligation junction, gave rise to a [5′-^{32}P]-labeled product of the appropriate length (Fig. 3B). As a control, the DNA enzyme also was directed to cleave synthetic oligonucleotide substrates of the same sequence that contained either a 2′,5′ or 3′,5′ linkage at the cleavage site, demonstrating no detectable cleavage of the former substrate and efficient cleavage of the latter.

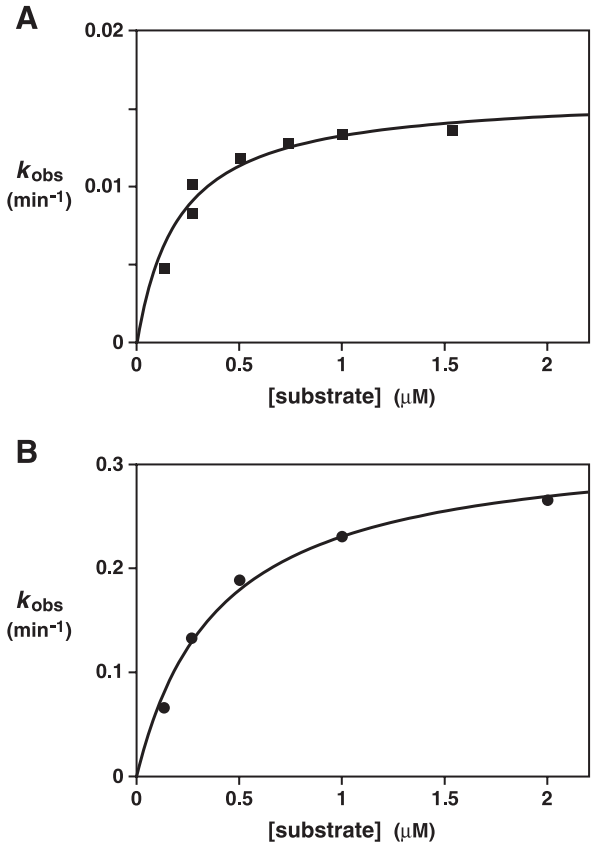

FIGURE 2. Catalytic activity of the evolved ligase ribozymes. **A**: The cytidine-free R3 ribozyme. **B**: The cytidine-containing R3C ribozyme. Values for k_{obs} were obtained in the presence of 10 nM ribozyme and various concentrations of substrate, under standard reaction conditions (see Materials and Methods). Data were fit to a curve based on the Michaelis–Menten equation: $v = k_{cat}$ [substrate]/(K_m + [substrate]).

Secondary structure analysis

The 74mer form of the R3 ligase ribozyme was chemically modified by either dimethyl sulfate or kethoxal,

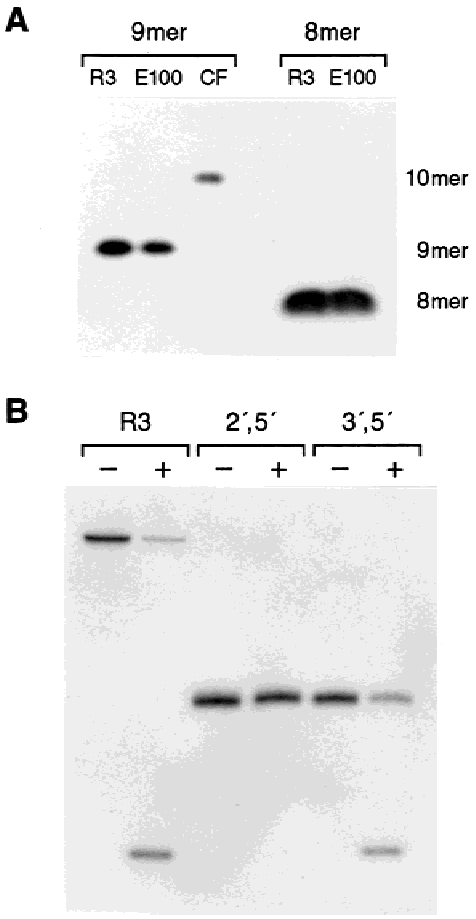

FIGURE 3. Analysis of the regiospecificity of ligation. **A**: Reaction of the ribozyme with a [5′-^{32}P]-labeled 8mer or 9mer substrate having a single ribonucleotide at its 3′ end, followed by digestion of the ligated product with RNase T2. The ribozyme was either the R3 ligase, the E100 ligase which forms a 3′,5′ linkage, or the E100-derived cytidine free ligase (CF) which forms a 2′,5′ linkage (Rogers & Joyce, 1999). **B**: Reaction of the R3 ribozyme with a [5′-^{32}P]-labeled all-RNA substrate, followed by digestion of the 89-nt ligated product with a DNA enzyme that is capable of cleaving 3′,5′ but not 2′,5′ linkages. The DNA enzyme also was directed to cleave synthetic 25mer RNAs that had the same sequence as the nucleotides surrounding the ligation junction, but with either a 2′,5′ or 3′,5′ linkage at the junction position. RNAs were incubated in either the absence (−) or presence (+) of the DNA enzyme. The products of either RNase T2 or DNA enzyme digestion were separated by electrophoresis in a denaturing polyacrylamide gel, an autoradiograms of which is shown.

followed by primer extension analysis using reverse transcriptase (Stern et al., 1988). Dimethyl sulfate methylates N1 of adenine, whereas kethoxal modifies the N1 and N2 positions of guanine, in both cases blocking primer extension during reverse transcription. Chemical probing was carried out in either the presence or absence of the oligonucleotide substrate under the conditions of the RNA-catalyzed reaction. Three adenosine residues, A10, A11, and A14, were found to be strongly protected from modification in the presence but not the absence of substrate. Surprisingly, these residues were located near the 5′ end of the ribozyme, rather than in the 3′-terminal portion that is complementary to the substrate. On the other hand, when a [5′-^{32}P]-labeled substrate was hybridized to the ribozyme and extended by reverse transcriptase, it generated a single product that was consistent with binding of the substrate at the expected position near the 3′ end of the ribozyme. This suggests that the protected adenosines near the 5′ end of the ribozyme underwent a conformational change upon binding of the substrate at the 3′ end.

To clarify the issue of ribozyme–substrate interactions, a unimolecular construct was prepared in which the substrate was covalently attached to the 3′ end of the ribozyme via a stable 5′-GAAA-3′ tetraloop (Antao & Tinoco, 1992). This ensured hybridization of the substrate at the expected position and allowed a self-ligation reaction to occur, giving rise to a circular product. The circularization reaction proceeded with a rate of 0.04 min^{-1} in the absence of any exogenous substrate. A small amount of ribozyme dimers and circularized ribozyme dimers also were formed. This result supports the notion that the substrate binds at the complementary position near the 3′ end of the ribozyme and that substrate binding induces a conformational change that causes residues A10, A11, and A14 to become engaged in hydrogen-bonding interactions.

If the interaction between ribozyme and substrate involves standard Watson–Crick pairing, then it should be possible to change the identity of those pairs without significantly decreasing catalytic activity. A special concern is that the substrate that was employed during in vitro evolution contained several cytidine residues, which were necessary to encode the T7 RNA polymerase promoter element. One or more of these cytidines might be required for activity, which would undermine the argument that cytidine is not essential for RNA catalysis. Replacing the cytidines of the 12mer substrate by uridines, forming G•U wobble pairs rather than G•C pairs between the ribozyme and substrate, eliminated catalytic activity. Activity could not be restored by replacing the corresponding guanosines of the ribozyme by adenosines, thus forming A•U pairs. After testing several combinations of ribozyme and substrate, it was determined that a uridine residue at position −5 of the substrate was highly detrimental to catalysis. Activity was maintained with a guanosine, adenosine, or cytosine at this position, so long as there was a complementary residue at the corresponding position of the ribozyme. Accordingly, adenosine was chosen for the −5 position of the substrate and uridine for the opposing position of the ribozyme. The other two G•C pairs were replaced by A•U pairs, resulting in an enzyme-substrate complex that was completely devoid of cytidine, yet retained full activity.

The cytidine-free reaction system involved a 12mer all-RNA substrate having the sequence 5′-UUAAU-AAAUAUA-3′ and a 77mer ribozyme with the com-

plementary sequence 5′-UAUAUUUAUUAA-3′ at its 3′ end. The ligation reaction proceeded with a k_{cat} of 0.013 min^{-1} and apparent K_m of 6.2 μM, reaching a maximum extent of about 85%. The product of the reaction contained a 3′,5′-phosphodiester linkage at the ligation junction (Fig. 3B). The 30-fold higher K_m of the cytidine-free substrate compared to the cytidine-containing substrate may be due to the weaker base pairing of the former. In addition, tertiary contacts may have developed between the ribozyme and cytidine-containing substrate during in vitro evolution, which were not possible with the cytidine-free substrate. The −5 position of the substrate, as discussed above, is one candidate for where such a tertiary contact might occur.

Introduction of cytidine

The R3 ribozyme had never been exposed to cytidine during its evolutionary history. Cytidine was introduced at random locations throughout the molecule by preparing an ensemble of synthetic DNA templates that encoded cytidine at a frequency of 1% per nucleotide position. These templates were further randomized by hypermutagenic PCR (Vartanian et al., 1996), then transcribed to generate a pool of cytidine-containing RNAs. Six rounds of in vitro evolution were carried out, with reaction times of 5 min in the first round, 30 s in the second and third rounds, and 10 s in the fourth through sixth rounds. Individual molecules were cloned from the population and sequenced after the third and sixth rounds. Following the third round, all of the clones that were examined had a unique sequence and contained at least two cytidine residues. Following the sixth round, all of the clones had the same sequence, as depicted in Figure 4. This ribozyme is referred to as the R3C ligase. It contains 73 nt, one less than the parent R3 ligase due to deletion of residue A7. The R3C ligase contains 11 other mutations relative to the R3 ligase: six U→C changes, one A→C change, two A→G changes, one U→A change, and one G→U change.

The reaction of the R3C ligase with the standard 12mer oligonucleotide substrate proceeded with a k_{cat} of 0.32 min^{-1} and apparent K_m of 0.4 μM (Fig. 2B). This corresponds to a 20-fold increase in catalytic rate and twofold increase in K_m compared to the behavior of the R3 ribozyme. The cytidine-containing R3C ligase, like its cytidine-free parent, formed a 3′,5′-phosphodiester linkage, based on analysis with a DNA enzyme that was directed to cleave the ligation junction.

The R3C ribozyme was chemically modified by either dimethyl sulfate or kethoxal, in either the presence or absence of the oligonucleotide substrate, followed by primer extension analysis using reverse transcriptase. Dimethyl sulfate methylates both N3 of cytidine and N1 of adenine, providing structural information regarding the seven cytidine residues that appeared in the R3C ribozyme. Six of these seven cytidines were protected from modification, suggesting that they were involved in Watson–Crick pairing. This is consistent with the proposed secondary structure model (Fig. 4), in which each of the protected cytidines is part of a G•C pair within a stem region. The single unprotected cytidine is located in an internal bulge loop adjacent to the three-way junction that is formed by the P2, P3, and P4 stems.

Further support for the proposed secondary structure model comes from examination of the cloned se-

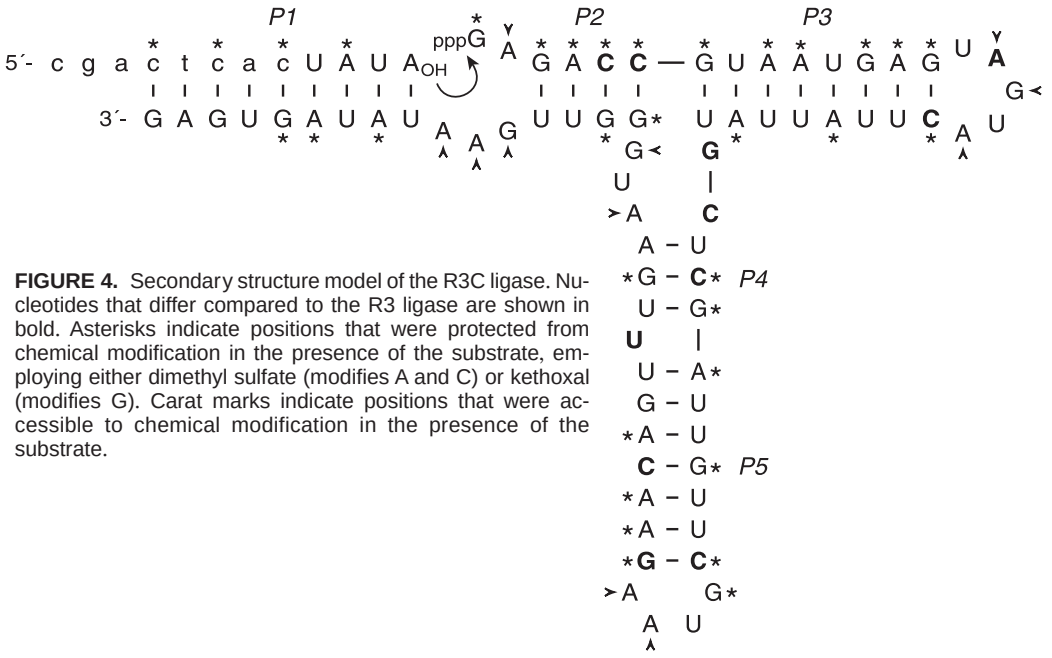

FIGURE 4. Secondary structure model of the R3C ligase. Nucleotides that differ compared to the R3 ligase are shown in bold. Asterisks indicate positions that were protected from chemical modification in the presence of the substrate, employing either dimethyl sulfate (modifies A and C) or kethoxal (modifies G). Carat marks indicate positions that were accessible to chemical modification in the presence of the substrate.

quences that were isolated following the third round of in vitro evolution in the presence of cytidine. Three sets of compensatory mutations were found, involving replacement of the U12•A24 pair within the P3 stem and the U31•A55 and U38•A47 pairs within the P5 stem by C•G pairs. Site-directed mutagenesis studies were carried out to test the effect of similar changes in the cytidine-free R3 ribozyme. Replacement of the A10•U26 and A11•U25 pairs within P3 by U•A pairs reduced the catalytic rate by about fivefold, and replacement of the U36•A49 and U38•A47 pairs within P5 by A•U pairs reduced the catalytic rate by less than twofold.

Comparison of the R3 and R3C ribozymes indicates that the two molecules have the same global secondary structure, but differ in the location of internal bulge loops within the P4–P5 domain (compare Figs. 1 and 4). The R3 ligase contains a single unpaired guanosine between the 3′ portion of P4 and the three-way junction, whereas the R3C ligase contains the sequence 5′-AUG-3′ at this location and the sequence 5′-GC-3′ on the opposing strand. Changing the unpaired C30 residue of the R3C ligase to U, which extended the P4 stem toward the junction (as in the R3 ligase), reduced the catalytic rate by 16-fold. Changing C30 to A reduced activity by sixfold.

The R3 ligase contains the sequence 5′-GUG-3′ between the 3′ portions of P5 and P4 and the sequence 5′-AG-3′ on the opposing strand, whereas the R3C ligase contains a single unpaired uridine (U52) between P5 and P4. When the unpaired uridine of R3C was deleted, the catalytic rate was reduced by 64-fold and the maximum extent of reaction was reduced by 15-fold. This suggests that when P4 and P5 are made to form a continuous helix, the majority of the molecules became trapped in an inactive conformation. As noted previously, the R3C ligase contains one less nucleotide than the R3 ligase due to deletion of residue A7. When this residue was restored to the R3C ligase, the catalytic rate decreased by more than 100-fold. In summary, detailed differences in the secondary structure of the R3 and R3C ligases are important for the enhanced activity of the cytidine-containing ribozyme.

Cytidine appears to play two roles in the R3C ligase. First, it enhances the stability of the P2, P3, and P5 stems due to replacement of one A•U pair and four G•U pairs by G•C pairs. The P5 stem is especially well stabilized, so much so that the U36•A49, G37•C48, and U38•A47 pairs all could be deleted with only a twofold decrease in catalytic rate. Second, cytidine enables reorganization of the internal bulge loops adjacent to the three-way junction, forming a structure that is more reactive, but apparently difficult to achieve in the absence of cytidine. The C32•G54 pair within the short P4 stem is essential for maintaining the integrity of the bulge loops that surround this stem. When C32 was mutated to A, the catalytic rate of the R3C ribozyme decreased by more than 300-fold.

Efforts were made to determine which of the seven added cytidines and accompanying mutations were most responsible for the improved activity of the R3C ligase compared to the R3 ligase. These mutation were divided into two classes: the stem-stabilizing mutations U5→C, U6→C, U21→C, U40→C, A45→G, and U48→C; and the bulge-remodeling mutations A7→deleted, A29→G, U30→C, A32→C, and G52→U (Fig. 5). The remaining U17→A mutation, within the P3 hairpin loop,

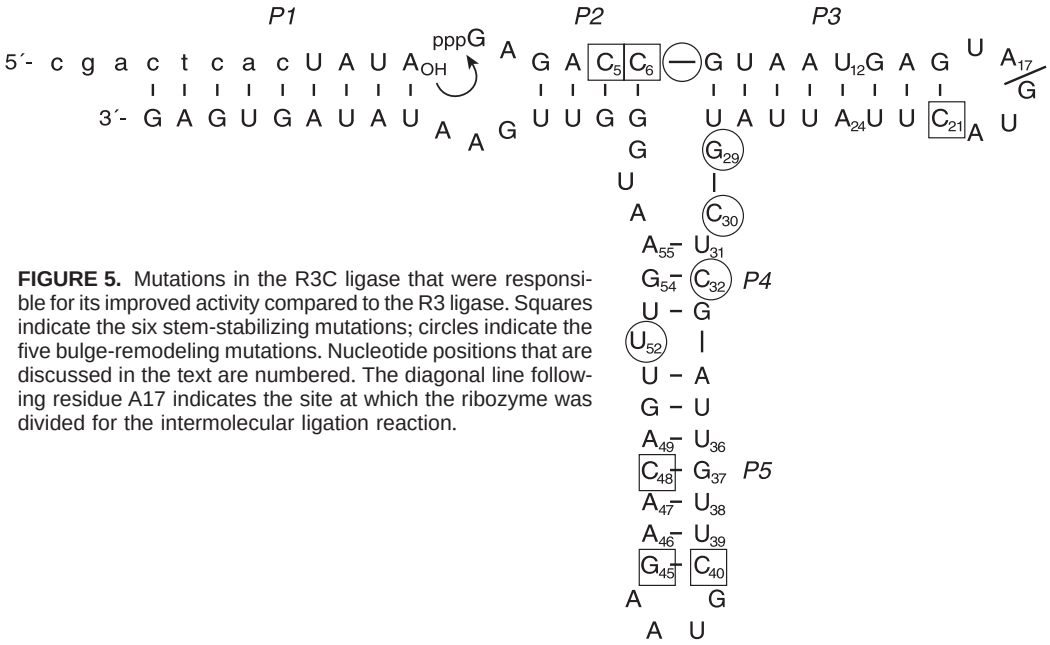

FIGURE 5. Mutations in the R3C ligase that were responsible for its improved activity compared to the R3 ligase. Squares indicate the six stem-stabilizing mutations; circles indicate the five bulge-remodeling mutations. Nucleotide positions that are discussed in the text are numbered. The diagonal line following residue A17 indicates the site at which the ribozyme was divided for the intermolecular ligation reaction.

was found to have no significant effect on catalytic activity. Starting with the R3 ligase and adding only the six stem-stabilizing mutations, the rate of reaction was unchanged, with k_{cat} remaining 0.02 min^{-1}. On the other hand, starting with the R3 ligase and adding only the five bulge-remodeling mutations, the catalytic rate decreased to 0.0008 min^{-1}. Thus the beneficial effects of the bulge-remodeling mutations can only be realized in the context of one or more of the stem-stabilizing mutations. Adding just two of these stem-stabilizing mutations, U5→C and U6→C, to a molecule that also contained the five bulge-remodeling mutations resulted in a 50-fold improvement in catalytic rate to 0.04 min^{-1}. This is still eightfold slower than the rate of the R3C ligase, a difference that must be attributed to one or more of the four more peripheral stem-stabilizing mutations.

The length and sequence of the P3 and P5 stems can be changed without significantly affecting catalytic activity. This suggests that the hairpin loops at the end of these stems are unlikely to play a role in catalysis. If the stems were to be made sufficiently stable, then it should be possible to delete the loops and assemble the ribozyme from separate strands of RNA. Introducing a separation within the P3 hairpin loop would be especially interesting because it would result in a trimolecular complex involving two short oligonucleotide substrates and a separate ligase ribozyme. To test this reaction format, the P3 stem of the R3C ligase was extended by 4 bp and the P3 hairpin loop was deleted. This resulted in three separate molecules: the standard 12mer substrate ending in a riboadenylate; an 18mer substrate bearing a 5′-triphosphate and having the same sequence as the first 16 nt of the R3C ligase with two added cytidine residues at its 3′ end; and a 58mer ribozyme having the same sequence as the last 57 nt of the R3C ligase with an added guanosine residue at its 5′ end (Fig. 5). Under saturating, single-turnover conditions, the 58mer ribozyme catalyzed intermolecular ligation with a catalytic rate of approximately 0.2 min^{-1} (Fig. 6). In the presence of saturating concentrations of the 12mer substrate, the apparent K_m of the ribozyme for the 18mer substrate was 0.1 μM.

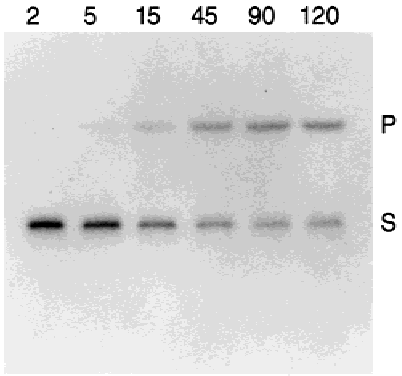

FIGURE 6. Time course of the intermolecular ligation reaction catalyzed by the R3C ribozyme. The 58mer ribozyme was allowed to react with an unlabeled 12mer substrate and labeled 18mer substrate (S), under standard conditions (see Materials and Methods), giving rise to a unique 30mer product (P). The ribozyme and 12mer substrate were present at saturating concentrations. The reaction was sampled at 2, 5, 15, 45, 90, and 120 min. The products were separated by electrophoresis in a denaturing polyacrylamide gel, an autoradiogram of which is shown.

DISCUSSION

There now are three known ribozymes that catalyze attack of an oligonucleotide 3′-hydroxyl on an oligonucleotide 5′-triphosphate to generate a 3′,5′-phosphodiester linkage: the class I ligase (Bartel & Szostak, 1993), the class hc ligase (Jaeger et al., 1999), and the R3 ligase. Of the three, the cytidine-free R3 ligase has the slowest catalytic rate, with a k_{cat} of 0.016 min^{-1}. Following the introduction of cytidine, the catalytic rate of the resulting R3C ligase was 0.32 min^{-1}. This is comparable to the rate of the class hc ligase, but still about 300-fold slower than that of the class I ligase. On the other hand, the R3/R3C ligase is the simplest of the three motifs, containing 74 nt when operating in the bimolecular reaction format and only 57 nt when operating with two separate oligonucleotide substrates. The class I and hc ligases contain 119 and 303 nt, respectively.

The uncatalyzed template-directed ligation of an oligonucleotide 2′,3′-hydroxyl and oligonucleotide 5′-triphosphate favors formation of a 3′,5′ compared to 2′,5′ linkage by about 70-fold (Rohatgi et al., 1996b). It appears to be much easier, however, to obtain ligase ribozymes that form a 2′,5′-phosphodiester linkage, in part because the 2′-hydroxyl is more nucleophilic than the 3′-hydroxyl (Lohrmann & Orgel, 1978). Ribozymes that form a 2′,5′ linkage do not require Watson–Crick pairing of the nucleotides that surround the ligation junction. In contrast, both the class I and class hc ligases position the ligation junction within a base-paired region. Attempts to develop cytidine-free ligases, therefore, might be expected to lead almost invariably to molecules that form a 2′,5′ linkage. A guanosine 5′-triphosphate is needed at the 5′ end of the ribozyme for efficient in vitro transcription, but that guanosine is unable to form a G•C pair. Indeed, six of the seven clones that were isolated following the first five rounds of in vitro evolution for a cytidine-free ligase formed a 2′,5′ linkage. The R3 clone, however, formed a 3′,5′ linkage, as did all of its descendants that were examined.

The ligation junction of the R3 ligase occurs within a purine-rich bulge that is flanked by base-paired regions (Fig. 1). The oligonucleotide substrate is bound by Watson–Crick pairing, but the 5′ end of the ribozyme, beginning with pppGpA, lies opposite the sequence 5′-GAA-3′. Chemical modification studies demonstrated that the guanosine 5′-triphosphate is protected at the

N1 and/or N2 positions, whereas all of the other nucleotides of the purine-rich bulge are unprotected. This suggests that the guanosine 5′-triphosphate is engaged in a nonstandard pair, perhaps involving hydrogen bonding of its N1 and O6 positions with the N7 and N6 positions of an opposing adenosine.

There are some structural similarities between the R3 ligase and the hairpin ribozyme, both of which catalyze RNA ligation, albeit by a different chemical mechanism. Both ribozymes contain an A-G sequence flanking the ligation junction and an unpaired 5′-GAA-3′ sequence on the opposing strand. Both contain extended double helical domains that meet at a junction and interact through tertiary contacts that help form the active enzyme-substrate complex. In the case of the hairpin ribozyme, these contacts involve interactions between the bulge loop at the ligation junction and a bulge loop within the other extended helical domain (for recent review, see Fedor, 2000). Analogous tertiary interactions may occur among the bulge loops of the R3 ribozyme, helping to stabilize an otherwise poorly constrained structure.

There are two major and conflicting obstacles that must be overcome in the development of a cytidine-free ribozyme. The first is the difficulty in forming a stable secondary structure in the absence of G•C pairs. This can be accomplished with longer stem regions based on A•U and G•U pairs. The second obstacle is the need to avoid alternative, inactive conformations that are likely to occur when every purine (A or G) is complementary to every pyrimidine (U). The longer the RNA, as would be required to form a stable secondary structure, the more difficult it will be to avoid alternative conformations. These two constraints may explain why the R3 ribozyme adopts a simple secondary structure that is formed by a modest number of nucleotides. Following the introduction of cytidine, the stem regions became more stable and more compact, allowing a more reactive secondary structure to form that was less susceptible to alternative conformations.

There are subtle differences in the secondary structures of the R3 and R3C ribozymes. Six of the 12 mutations that arose in the R3C ribozyme, including five of the seven added cytidines, stabilize the stem regions by converting A•U or G•U pairs to G•C pairs. Five other mutations result in remodeling of the bulge loops that lie adjacent to the three-way junction. The bulge-remodeling mutations alone were highly detrimental to the catalytic activity of the R3 ligase, but together with the stem-stabilizing mutations proved highly beneficial.

It is possible that one or more of the added cytidines in the R3C ligase play a role in the catalytic mechanism, accounting for some of the 20-fold improvement in catalytic rate compared to that of the R3 ligase. This improvement is rather modest, however, corresponding to only about 2 kcal/mol of transition state stabilization. Thus it is unlikely that an added cytidine plays a direct role in catalysis. In contrast, a particular cytidine residue has been shown to play a key role in the mechanism of the hepatitis delta virus ribozyme, acting as a general acid as a result of an upward shift in the pK_a of its N3 nitrogen (Perrotta et al., 1999; Nakano et al., 2000). Within the R3C ligase, the C30 residue appears to be the most important cytidine for catalysis. This residue lies 2 nt away from the three-way junction in the single-stranded region that joins the P3 and P4 stems (Fig. 4). Changing this cytidine to uridine reduced catalytic activity by 16-fold, although this may have been due to alteration of the P4 stem rather than a direct effect on catalysis.

It is not clear whether more rounds of in vitro evolution would lead to further improvement in the catalytic activity of the cytidine-containing ligase. If a high level of random mutagenesis is applied during the evolution process, then the molecule might adopt a completely different secondary structure, preventing direct comparison to the cytidine-free ligase. If a low level of mutagenesis is applied, then the population would remain confined to the neighborhood of the R3 ligase, where there seems to be little opportunity for further improvement in activity. Nonetheless, the improvement that was obtained demonstrates the benefit of added cytidine to a preexisting functional RNA. RNA is capable of achieving a substantial catalytic rate enhancement in the absence of cytidine, but evolution quickly exploits cytidine once it has been made available.

MATERIALS AND METHODS

In vitro evolution

In vitro evolution was carried out as described previously (Rogers & Joyce, 1999). The cytidine-free RNAs were prepared by in vitro transcription in the presence of 2 mM each of GTP, ATP, and UTP, then purified by denaturing polyacrylamide gel electrophoresis and subsequent ethanol precipitation. The cytidine-containing RNAs were prepared in the same way, except that 2 mM CTP was included in the transcription mixture. Ligation reactions were carried out in the presence of ~1 μM ribozyme, 5 μM substrate having the sequence 5′-d(GAGACTGATCTAGACTCTAATACGACTCACTAT)rA-3′ (T7 promoter sequence underlined), 25 mM $MgCl_2$, 50 mM KCl, 2 mM spermidine, 4 mM dithiothreitol, and 50 mM EPPS (pH 8.5) at 23 °C. The ribozyme was incubated under these conditions for 5 min prior to addition of the substrate. The reaction was quenched by addition of 4 M urea, 10% sucrose, and 25 mM Na_2EDTA. Ligated RNAs were gel purified, then reverse transcribed in the presence of all four dNTPs using a DNA primer with the sequence 5′-CCTAACCCTTCTCAACCT-3′, which was complementary to the 3′ end of the ribozyme. The other primer for PCR amplification had the sequence 5′-GAGACTGATCTAGACTC-3′, corresponding to the first 17 nt at the 5′ end of the substrate. Individuals were cloned from the population using the Invitrogen TA Cloning Kit and sequenced by the cycle sequencing method (Murray, 1989).

Analysis of regiospecificity of ligation

Ligation was carried out using [5′-^{32}P]-labeled substrate and unlabeled ribozyme. The ligated product was purified by denaturing polyacrylamide gel electrophoresis and subsequent ethanol precipitation. Two methods were used to determine whether a 2′,5′ or 3′,5′ linkage had formed at the ligation junction. The first involved reaction of the ribozyme with a short oligonucleotide substrate that contained a single ribonucleotide at its 3′ end. The ligated product was digested to completion with RNase T2 in a mixture containing 10 nM ligated product, 1 U/μL RNase T2, 2 mM Na_2EDTA, and 50 mM NaOAc (pH 5.2), which was incubated at 37 °C for 30 min. The labeled product of RNase cleavage contained a 3′-terminal phosphate. The second method involved reaction of the ribozyme with an all-RNA substrate having the sequence 5′-UUAAUAAAUAUA-3′, followed by treatment of the ligated product with an 8-17 DNA enzyme (Santoro & Joyce, 1997) having the sequence 5′-CATTACTAATCTTCCGAGCCGGTCGAAATATTTATTAA-3′ (substrate-recognition domains underlined). The DNA enzyme recognized the sequence 5′-UUAAUAAAUAU**A**•GAGAUUAGUAAUG-3′, with cleavage occurring following the single unpaired adenosine (shown in bold) immediately preceding the ligation junction. Authentic 25mer RNAs with the same sequence for recognition by the DNA enzyme, but with either a 2′,5′ or 3′,5′ linkage at the cleavage site, were prepared by chemical synthesis using commercially available phosphoramidites (ChemGenes). The DNA enzyme and RNA substrate first were allowed to hybridize by heating to 95 °C for 1 min, then cooling to 37 °C over 5 min in a mixture containing 100 mM KCl, 0.25 mM Na_2EDTA, and 50 mM Tris•HCl (pH 7.5). DNA-catalyzed RNA cleavage was carried out in a mixture containing 10 nM ligated product (or synthetic RNA), 30 μM DNA enzyme, 25 mM $MgCl_2$, 50 mM KCl, 2 mM spermidine, and 50 mM EPPS (pH 8.5), which was incubated at 37 °C for 30 min. Following treatment with either RNase T2 or the DNA enzyme, the [5′-^{32}P]-labeled products were analyzed by denaturing polyacrylamide gel electrophoresis.

Analysis of secondary structure

Unlabeled ribozyme was subjected to partial alkaline hydrolysis by incubation in the presence of 100 mM $NaHCO_3$/Na_2CO (pH 9.2) at 95 °C for 5 min, then allowed to react with [5′-^{32}P]-labeled substrate. The products were visualized by high-resolution polyacrylamide gel electrophoresis and the size of the shortest ligated product that appeared in good yield was determined by comparison to authentic markers. The 74mer 3′-truncated RNA was prepared by in vitro transcription of PCR-amplified DNA that had been generated by internal priming of the full-length DNA. Chemical probing experiments employing either dimethyl sulfate or kethoxal were carried out as previously described (Stern et al., 1988). For chemical probing that was done in the presence of substrate, the ribozyme and substrate were allowed to incubate together for 5 min prior to addition of the modifying agent. The conditions for chemical probing were identical to the standard reaction conditions, except that the buffer was sodium cacodylate rather that EPPS. The catalytic behavior of the ribozyme was the same in the two buffers. Site-directed mutagenesis was carried out by PCR mutagenesis, introducing the desired mutation(s) within the appropriate PCR primer, followed by in vitro transcription.

Kinetic analysis

Ligation of the ribozyme and substrate was carried out using body-labeled ribozyme that had been prepared by in vitro transcription in the presence of [α-^{32}P]ATP. Each reaction mixture contained 10 nM ribozyme, 0.1–2 μM substrate, 25 mM $MgCl_2$, 50 mM KCl, 2 mM spermidine, 4 mM dithiothreitol, and 50 mM EPPS (pH 8.5), which was incubated at 23 °C for various times. The ligated products were separated by polyacrylamide gel electrophoresis and quantitated using a Molecular Dynamics PhosphorImager. Values for k_{obs} were determined for each concentration of substrate, based on at least five data points obtained over the first three half-lives of the reaction. Values for k_{cat} and apparent K_m were obtained from a standard Michaelis–Menten saturation plot of k_{obs} versus substrate concentration. Estimations of k_{cat} for various mutant ribozymes were obtained using ~10 nM ribozyme and 3.3 μM (saturating) substrate, under conditions as above. The intermolecular ligation reaction, which involved the ribozyme and two oligonucleotide substrates, was carried out in the presence of 0.05–1 μM ribozyme, 6.7 μM 12mer substrate having the sequence 5′-d(CGACTCAC)-r(UAUA)-3′, and 10 nM body-labeled 18mer substrate having the sequence 5′-r(GAGACCGUAAUGAGUACC)-3′ and bearing a 5′-triphosphate, incubated under reaction conditions as above. Values for k_{obs} were determined for various concentrations of ribozyme and a fixed concentration of the two substrates.

ACKNOWLEDGMENTS

This work was supported by National Aeronautics and Space Administration grant NAG5-3647 and the Skaggs Institute for Chemical Biology at The Scripps Research Institute.

Received November 10, 2000; returned for revision December 4, 2000; revised manuscript received December 6, 2000

REFERENCES

Antao VP, Tinoco I. 1992. Thermodynamic parameters for loop formation in RNA and DNA hairpin loops. *Nucleic Acids Res 20*:819–824.

Bartel DP, Szostak JW. 1993. Isolation of new ribozymes from a large pool of random sequences. *Science 261*:1411–1418.

Ekland EH, Bartel DP. 1996. RNA-catalyzed RNA polymerization using nucleoside triphosphates. *Nature 382*:373–376.

Ekland EH, Szostak JW, Bartel DP. 1995. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science 269*:364–370.

Fedor MJ. 2000. Structure and function of the hairpin ribozyme. *J Mol Biol 297*:269–291.

Hager AJ, Szostak JW. 1997. Isolation of novel ribozymes that ligate AMP-activated RNA substrates. *Chem Biol 4*:607–617.

Hegg LA, Fedor MJ. 1995. Kinetics and thermodynamics of intermolecular catalysis by hairpin ribozymes. *Biochemistry 34*:15813–15828.

Hertel KJ, Uhlenbeck OC. 1995. The internal equilibrium of the hammerhead ribozyme reaction. *Biochemistry 34*:1744–1749.

Jaeger L, Wright MC, Joyce GF. 1999. A complex ligase ribozyme

evolved in vitro from a group I ribozyme domain. *Proc Natl Acad Sci USA 96*:14712–14717.
Landweber LF, Pokrovskaya ID. 1999. 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.
Levy M, Miller SL. 1998. The stability of the RNA bases: Implications for the origin of life. *Proc Natl Acad Sci USA 95*:7933–7938.
Lohrmann R, Orgel LE. 1978. Preferential formation of (2′-5′)-linked internucleotide bonds in non-enzymatic reactions. *Tetrahedron 34*:853–855.
Murray V. 1989. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res 17*:8889.
Nakano S-i, Chadalavada DM, Bevilacqua PC. 2000. General acid-base catalysis in the mechanism of the hepatitis delta virus ribozyme. *Science 287*:1493–1497.
Perrotta A, Shih I-h, Been MD. 1999. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science 286*:123–126.
Robertson MP, Ellington AD. 2000. Design and optimization of effector-activated ribozyme ligases. *Nucleic Acids Res 28*:1751–1759.
Rogers J, Joyce GF. 1999. A ribozyme that lacks cytidine. *Nature 402*:323–325.
Rohatgi R, Bartel DP, Szostak JW. 1996a. Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J Am Chem Soc 118*:3332–3339.
Rohatgi R, Bartel DP, Szostak JW. 1996b. Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3′-5′ phosphodiester bonds. *J Am Chem Soc 118*:3340–3344.
Santoro SW, Joyce GF. 1997. A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci USA 94*:4262–4266.
Stern S, Moazed D, Noller HF. 1988. Analysis and RNA structure using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol 164*:481–489.
Vartanian J, Henry M, Wain-Hobson S. 1996. Hypermutagenic PCR involving all four transitions and a sizeable portion of transversions. *Nucleic Acids Res 24*:2627–2631.
Wright MC, Joyce GF. 1997. Continuous in vitro evolution of catalytic function. *Science 276*:614–617.