Isolation of a ribozyme with 5'–5' ligase activity

Karen B Chapman[†] and Jack W Szostak*

Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA

Background: Many new ribozymes, including sequence-specific nucleases, ligases and kinases, have been isolated by *in vitro* selection from large pools of random-sequence RNAs. We are attempting to use *in vitro* selection to isolate new ribozymes that have, or can be evolved to have, RNA polymerase-like activities. As phosphorimidazolide-activated nucleosides are extensively used to study non-enzymatic RNA replication, we wished to select for a ribozyme that would accelerate the template-directed ligation of 5'-phosphorimidazolide-activated oligonucleotides.

Results: Ribozymes selected to perform the desired template-directed ligation reaction instead ligated themselves to the activated substrate oligonucleotide via their 5'-triphosphate, generating a 5'–5' P^1,P^4-tetraphosphate linkage. Deletion analysis of one of the selected sequences revealed that a 54-nucleotide RNA retained activity; this small ribozyme folds into a pseudoknot secondary structure with an internal binding site for the substrate oligonucleotide. The ribozyme can also synthesize 5'–5' triphosphate and 5'–5' pyrophosphate linkages.

Conclusions: The emergence of ribozymes that accelerate an unexpected 5'–5' ligation reaction from a selection designed to yield template-dependent 3'–5' ligases suggests that it may be much easier for RNA to catalyze the synthesis of 5'–5' linkages than 3'–5' linkages. 5'–5' linkages are found in a variety of contexts in present-day biology. The ribozyme-catalyzed synthesis of such linkages raises the possibility that these 5'–5' linkages originated in the biochemistry of the RNA world.

Chemistry & Biology May 1995, **2**:325–333

Key words: evolution, *in vitro* selection, origin of life, RNA world

Introduction

We have been pursuing the discovery of new ribozymes to provide an experimental foundation for thinking about the potential functions of ribozymes in early cells, and as a way of providing new starting points for attempts to evolve self-replicating RNAs. *In vitro* selection has proven to be a powerful method for the isolation of RNAs with novel functions from large pools of random sequences [1]. *In vitro* selection experiments involving iterative cycles of affinity chromatography and sequence amplification have led to the isolation of specific RNAs that bind a wide range of small molecules with high specificity [2]. Direct selections for catalytic RNAs from partially and completely random sequence RNA pools have yielded several novel classes of ribozymes, including self-cleaving RNAs [3] and a variety of ligases [4,5]. More complex ribozymes that bind two substrates and release two products have recently been isolated from pools of sequences biased towards a binding site for one substrate [6,7]. The chemistry catalyzed by ribozymes has recently been extended beyond phosphotransfer reactions with the isolation of a rotational isomerase [8], a self-alkylating ribozyme [9] and ribozymes that carry out aminoacyl transfer reactions [10] (P. Lohse and J.W. S., unpublished data).

RNA can catalyze several reactions that lead to polynucleotide synthesis [11,12]. We have found that derivatives of group I introns are able to join together a series of short oligonucleotides that have been aligned on a template strand, producing a complementary strand [13]. However, the transesterification reaction catalyzed by these ribozymes conserves the total number of phosphodiester bonds. The joining reaction is therefore largely driven entropically, by the release of guanosine, and the accumulation of this product could ultimately limit the replication of longer templates. Recently, we have isolated ribozymes that catalyze the ligation of oligonucleotides with the same chemistry used by present-day polymerases, that is, the attack of a 3' hydroxyl on the α-phosphate of a 5'-triphosphate [4,5]. Some of these ribozymes may be suitable starting points for the *in vitro* evolution of ribozyme polymerases that use nucleoside (or oligonucleotide) triphosphates as substrates.

In order to isolate other ribozymes that might be evolved into replicases, we have started to explore the use of other phosphate activation chemistries. Nucleoside phosphorimidazolides have been used extensively by Orgel and colleagues [14,15] in studies of non-enzymatic template-directed copying reactions, because they are more highly activated than nucleoside triphosphates and therefore react more rapidly in uncatalyzed reactions. This higher uncatalyzed reaction rate led us to hypothesize that it might be easier to isolate ribozymes that would join phosphorimidazolide-derivatized oligonucleotides instead of triphosphate-activated oligonucleotides, because a smaller rate enhancement would still provide a useful absolute reaction rate. Here we present the results of an attempt to use *in vitro* selection to isolate novel ribozymes that ligate their own 3' ends to a 5'-phosphorimidazolide-activated oligonucleotide substrate. Instead of the desired activity, we found ribozymes that ligated their 5' ends to the substrate.

*Corresponding author. [†]Present address: Geron Corporation, 200 Constitution Drive, Menlo Park, CA 94025, USA.

 ISSN 1074-5521

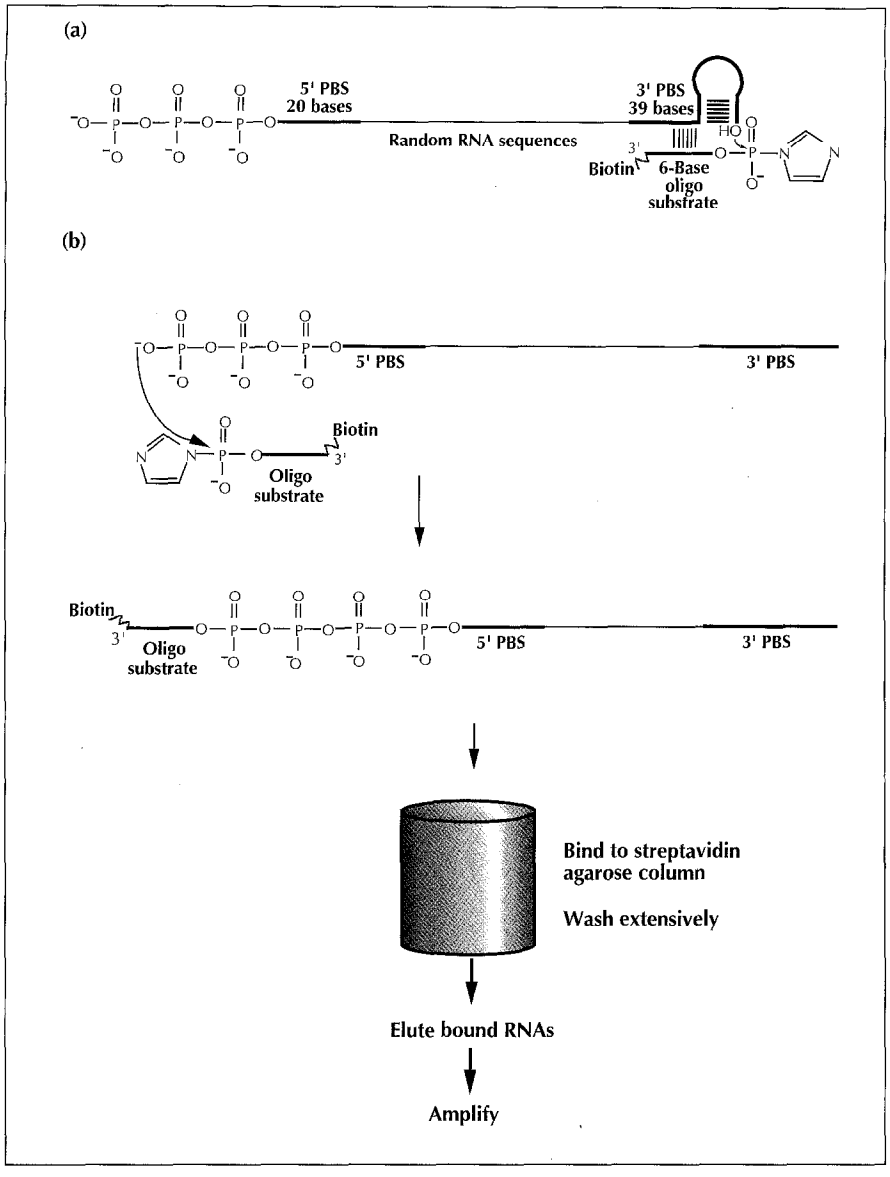

Fig. 1. *In vitro* selection scheme for isolating RNAs that self-ligate to a short substrate oligonucleotide. **(a)** The RNA pool contains 90 nucleotides of random sequences flanked by defined primer binding sites (PBS) for PCR amplification. The substrate is a six-nucleotide synthetic RNA, which is biotinylated at its 3' end and activated with a 5'-phosphorimidazolide. The 3'-PBS of the RNA pool contains a sequence complementary to the substrate oligonucleotide, designed so that binding of the substrate oligonucleotide would form an interrupted hairpin-loop, which would juxtapose the 3'-OH of the pool RNA with the 5'-phosphorimidazolide moiety of the substrate. **(b)** The reaction catalyzed by the RNA that was actually selected from the pool: attack of the γ-phosphate of the 5'-triphosphate of the pool RNA on the 5'-phosphorimidazolide of the substrate to generate a 5'–5' tetraphosphate linkage. Molecules that became ligated to the substrate were purified away from unligated RNAs by binding to streptavidin agarose. These RNAs were eluted, reverse transcribed, PCR amplified and transcribed with T7 RNA polymerase, generating more RNA for further rounds of selection.

Results

In vitro selection

An *in vitro* selection experiment involves the synthesis of a suitable pool of random-sequence RNA molecules, followed by the incubation of the pool with a substrate and the subsequent selection from the pool of those rare molecules that are able to modify themselves by reacting with the substrate. In the selection described here, the substrate was a hexanucleotide with a 5'-phosphate that was activated by reaction with the water-soluble carbodiimide EDC and imidazole to form the reactive phosphorimidazolide (Fig. 1). The substrate was also biotinylated at its 3' end, allowing the reacted molecules to be purified using streptavidin agarose.

A random-sequence RNA pool was constructed by *in vitro* transcription of a synthetic oligodeoxynucleotide that had been amplified by PCR. The pool had a complexity of ~1 x 10^{15} different molecules and each RNA molecule contained 90 random nucleotides flanked by defined primer-binding sites of 20 and 39 nucleotides. The 3' primer binding site was designed to fold into a partial hairpin stem–loop adjacent to the substrate-binding site such that Watson–Crick base pairing with the substrate hexanucleotide would position the 3'-hydroxyl of the pool RNA to attack the 5'-phosphorimidazolide of the activated substrate (Fig. 1a). The template-directed but otherwise uncatalyzed condensation of imidazolide-activated oligonucleotides generates predominantly 3'-5' phosphodiester linkages (R. Rohatgi and J.W.S., unpublished data). The rate of ligation of random pool RNA with saturating concentrations of this substrate was ~7 x 10^{-6} min^{-1}, while reaction with a different substrate, not complementary to the template region, was undetectable after a 24 h reaction.

In the first round of *in vitro* selection, the activated substrate was incubated with the random-sequence RNA for 4 h in a buffer containing high concentrations of salt and magnesium (to stabilize folded RNA structures). Pool molecules that had become covalently linked to the biotinylated substrate were purified by streptavidin

Table 1. Selection Summary

Round	Amount of RNA	Reaction time	Mg^{2+} conc.	% recovered
1	1 mg	4 h	50 mM	0.03 %
2	50 μg	45 min	50 mM	1.5 %
3	50 μg	30 min	50 mM	0.008 %
4	50 μg	30 min	50 mM	0.024 %
5	50 μg	30 min	50 mM	0.02 %
6	50 μg	20 min	50 mM	2.0 %
7	50 μg	2 min	50 mM	0.7 %
8	50 μg	2 min	100 μM	0.03 %

Summary of reaction conditions throughout the selection. Percentage recovered is percentage of input RNA recovered from the streptavidin column. Following round 2, the column washing conditions were altered to include a denaturing wash to reduce non-specific sticking to the column.

affinity chromatography, reverse transcribed and PCR amplified (Fig. 1b). The resulting DNA was transcribed *in vitro* with T7 RNA polymerase to generate more RNA for further rounds of selection, during which the reaction time was gradually reduced to as little as 2 min to increase the stringency of selection for more active ribozymes (Table 1). The concentration of Mg^{2+} was also decreased in the final round to enhance selection of RNA molecules with stable folded structures. After eight rounds of *in vitro* selection, the ligation activity of the selected pool had increased to ~0.8 min^{-1}, an increase of 100 000-fold over the rate exhibited by the initial random pool (Fig. 2).

Sequence and deletion analysis of the selected ribozymes

Individual molecules from the final selected pool were cloned and sequenced. Examination of the sequences of 28 clones revealed a major class of selected RNAs (class 1) represented by six virtually identical clones (Fig. 3a). A second minor class was represented by two individual clones (class 2) while the remaining twenty clones were unique sequences.

Deletion analysis of one of the class 1 ribozymes revealed that as many as 95 nucleotides could be removed from the 3' end (resulting in a 54-nucleotide ribozyme) without greatly affecting ligation activity; however, deletion of 115 nucleotides resulted in an inactive 34-nucleotide derivative. This was a surprising result since it showed that the template provided for substrate binding was not necessary for activity. The 54-nucleotide ribozyme had a k_{cat} for the self-ligating reaction of 0.09 min^{-1} and a K_m of 8 μM for the activated hexanucleotide substrate. This template-independent activity appeared to be a general characteristic of the selected pool. Round 6 RNA was tested for template dependence by changing the designed substrate-binding sequence by PCR. This change did not significantly affect the ligation activity of the selected pool. In contrast to these results, the 5' terminus of the RNA appeared to be critical for ligation, since a PCR-generated deletion of only 10 nucleotides from the 5'-end of the transcript resulted in complete loss of activity.

The secondary structure of the 54-nucleotide deletion mutant is shown in Figure 3b. The existence of the two base-paired stems of the pseudoknot was confirmed by site-directed mutagenesis. Point mutations in the paired regions significantly decreased activity, while double mutations that restored base-pairing also restored activity. Pairing between the 5' sequence G_1GAAC_5, and the sequence $G_{21}UUCU_{25}$ was tested by making the mutants $A_3 \rightarrow U$ and $U_{23} \rightarrow A$, both of which were

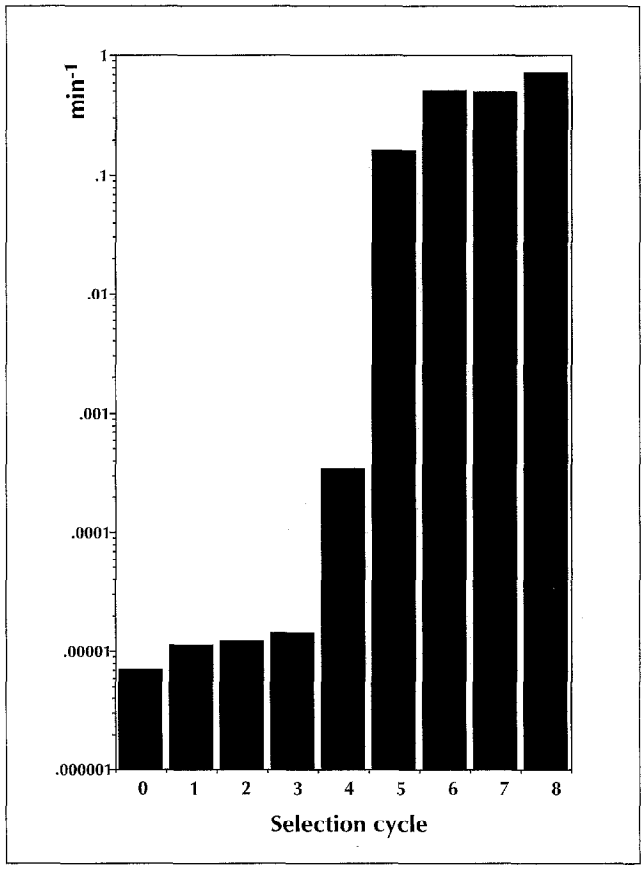

Fig. 2. Rate of ligation of RNA transcribed after each round of selection. The rate observed for the random pool (round 0) reflects the template-directed uncatalyzed reaction, which generates a 3'–5' phosphodiester linkage. Reaction rates seen in later rounds reflect the emergence of RNAs that catalyze the 5'–5' ligation reaction. Initial rates were measured under selection conditions.

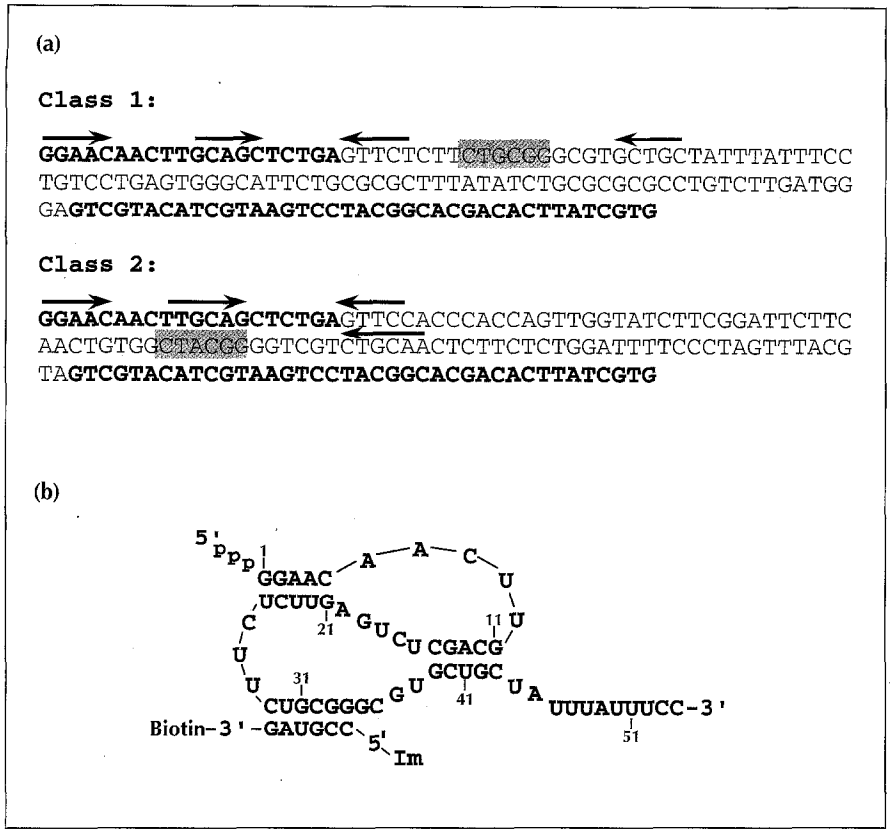

Fig. 3. RNA sequences and structure. **(a)** Sequences of class 1 and class 2 RNAs. Invariant primer binding sites are in bold. Regions complementary to the substrate are boxed, and internal complementary sequences are indicated by arrows of matching color. **(b)** The predicted secondary structure of the smallest active class 1 deletion derivative (cl-54) is depicted below. The class 2 sequence can be folded into a similar structure.

inactive, and the double mutant, which had wild-type activity. Pairing between $G_{11}CAGC_{15}$ and $G_{39}CUGC_{43}$ was tested by making the single mutants $G_{14}\rightarrow C$ and $C_{40}\rightarrow G$, which each had less than 10 % of wild type activity, and the double mutant, which was restored to full activity. Similar experiments failed to provide evidence supporting two other possible pairings, between G_1GAA_4 and $U_{51}UCC_{54}$, and between A_6ACUU_{10} and $G_{19}AGUU_{23}$.

Selected RNAs generate a 5'–5' P^1,P^4-tetraphosphate linkage

Given the apparent importance of the 5'-end of the class 1 ribozyme RNA, its role in the ligation reaction was examined more carefully. *In vitro* transcription with T7 RNA polymerase in the presence of GTP generates RNA transcripts that begin with 5'-triphosphates. This 5'-triphosphate was found to be essential for the ribozyme-catalyzed ligation, as pre-treatment of the RNA with phosphatase results in a complete loss of activity (Fig. 4a). Furthermore, the γ-phosphate of a class 1 transcript was completely removable by treatment with phosphatase, but became resistant to phosphatase after reaction with the substrate oligonucleotide (Fig. 4b). This suggested that the ligation reaction might involve attack of the γ-phosphate of the 5'-triphosphate of the RNA on the 5'-phosphorimidazolide of the substrate oligonucleotide, generating a 5'–5' P^1,P^4 tetraphosphate linkage.

Further evidence that the reaction product is 5'–5' P^1,P^4-tetraphosphate-linked substrate–ribozyme was provided by P1 nuclease digestion of the reacted RNA and thin-layer chromatography (TLC) of the digestion products. P1 nuclease is an exonuclease that digests linear RNA of non-specific sequence to 5'-nucleoside monophosphates; it does not hydrolyze 5'–5' linkages such as those found in the mRNA cap structure [16]. P1 nuclease digestion of substrate oligonucleotide which had been 5'-phosphorylated with γ-[^{32}P]-ATP and T4 polynucleotide kinase yielded the expected product, CMP, as analyzed by TLC (Fig. 4d). After activation of this labeled substrate and reaction with class 1 RNA, P1 nuclease digestion of the product resulted in a more slowly migrating product, consistent with the formation of Cp*pppG (the asterisk follows the position of the ^{32}P label). A similar P1 nuclease digestion product is obtained when the labeled phosphate is the γ-phosphate of the 5'-triphosphate of the ribozyme RNA. When γ-^{32}P-labeled RNA was digested with P1 nuclease, the labeled product GTP was obtained as expected (Fig. 4d, lane 3). However, following reaction with unlabeled substrate oligonucleotide, P1 nuclease digestion of the reaction mixture again yielded a product consistent with Cpp*ppG. The co-migration of the reaction products generated from 5'-labeled substrate or 5'-labeled ribozyme RNA (Fig. 4d) shows that the 5'-phosphate of the substrate and the γ-phosphate of the ribozyme RNA are in the same P1 nuclease-resistant linkage. This linkage is also resistant to ribonuclease T2 digestion and alkaline hydrolysis, as expected for a 5'–5' P^1,P^4-tetraphosphate linkage. A similar nuclease digestion analysis of the round 5 selected pool indicates that the vast majority of the pool molecules (>97 %) catalyze the same type of reaction.

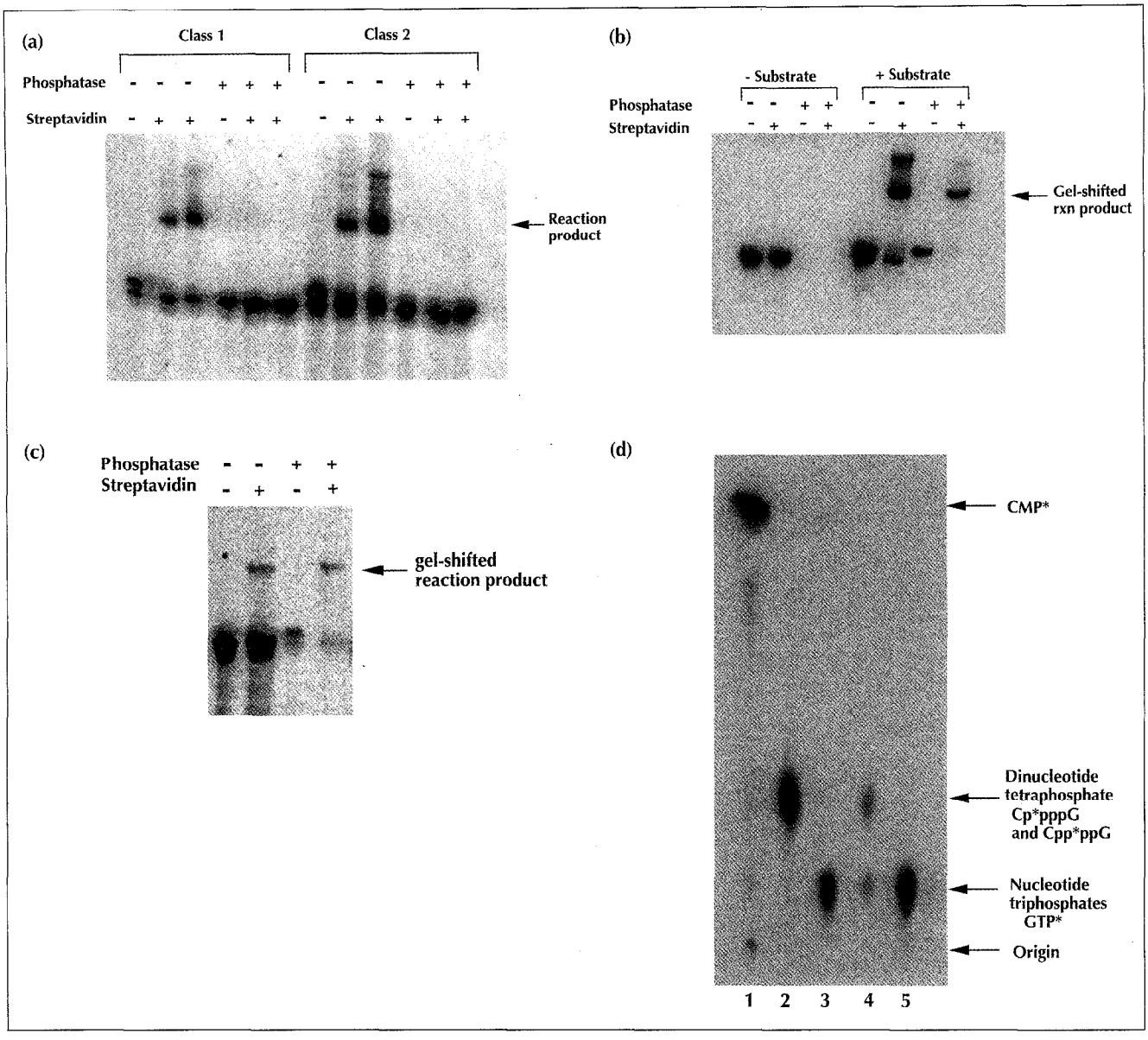

Fig. 4. Characterization of ligation reactions. **(a)** Both class 1 and class 2 ribozymes lose activity if their 5'-triphosphates are removed with phosphatase. Class 1 and class 2 RNAs, internally labeled with ^{32}P, were incubated +/– alkaline phosphatase prior to 10 and 30 min reactions with activated substrate oligonucleotide. Reaction aliquots were incubated +/– streptavidin prior to gel electrophoresis; the position of the ligated, gel-shifted product is indicated. Streptavidin conjugates run more slowly than free RNAs in this gel system; slowest migrating bands are due to complexes with streptavidin multimers. Reaction products were visualized using a phosphoimager (Molecular Dynamics). **(b)** Reaction products are resistant to phosphatase treatment. Class 1 RNA beginning with γ-[^{32}P]-GTP was treated with phosphatase before (lanes 1–4) or after (lanes 5–8) reaction with biotinylated substrate. Aliquots of each reaction were mixed with streptavidin as indicated to distinguish between ligated and unligated RNAs. **(c)** Reaction product formed from 5'-monophosphorylated class 1 RNA is resistant to phosphatase. Class 1 RNA that had been phosphatased and subsequently phosphorylated with polynucleotide kinase and γ-[^{32}P]-ATP was incubated with activated substrate, treated with phosphatase as indicated, and analyzed as above. **(d)** Thin layer chromatography of reaction products following digestion with nuclease P1. Lane 1 is a nuclease P1 digest of the substrate oligonucleotide phosphorylated with polynucleotide kinase and γ-[^{32}P]-ATP, thus yielding p*C, with the asterisk indicating the position of the ^{32}P label. Lane 2 is a nuclease P1 digestion of the same labeled substrate oligonucleotide after reaction with class 1 RNA, thus yielding Cp*pppG. Lane 3 is a nuclease P1 digest of class 1 RNA labeled with γ-[^{32}P]-GTP during transcription, and which yields p*ppG after nuclease P1 digestion. Lane 4 is a nuclease P1 digest of the same labeled class 1 RNA following incomplete reaction with the substrate RNA, yielding both Cpp*ppG and p*ppG. Lane 5 is authentic γ-[^{32}P]-GTP.

The class 1 ribozyme can also generate 5'–5' triphosphate and 5'–5' pyrophosphate linkages

Class 1 ribozyme RNA that begins with a 5'-diphosphate or monophosphate retains partial activity in the 5'–5' ligation reaction. When RNA is transcribed in the presence of either GMP or GDP in addition to the four nucleoside triphosphates, a fraction of the RNAs initiate with either GMP or GDP respectively. When the class 1 5'–5' ligase was transcribed in the presence of GMP, two linkages to substrate were observed after P1 nuclease digestion, consistent with CppppG and CppG being formed from RNAs beginning with 5'-triphosphates or

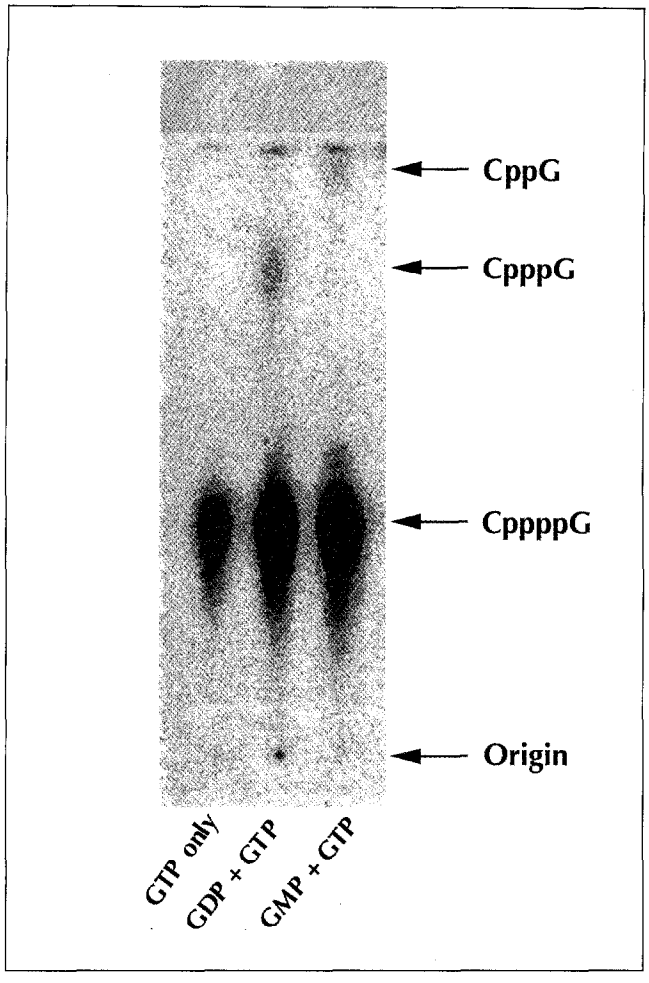

Fig. 5. The selected ribozyme generates a variety of 5'–5' oligophosphate linkages. cI-54 RNA was transcribed in the presence of all four nucleotide triphosphates plus GDP or GMP as indicated. The resulting RNA was reacted with labeled, activated substrate oligonucleotide (r37) and the reacted product was gel-purified and digested with nuclease P1. TLC was carried out as described in Materials and methods.

5'-monophosphates respectively (Fig. 5). A similar result was obtained when this RNA was transcribed in the presence of GDP, consistent with the formation of a 5'–5' triphosphate linkage.

Formation of the pyrophosphate linkage was directly confirmed with class 1 ribozyme RNA that had been treated with phosphatase to remove the 5'-triphosphate and then mono-phosphorylated by incubation with γ-^{32}P-ATP and polynucleotide kinase (Fig. 4c). In this case, the pyrophosphate linked product was observed exclusively. The 5'-monophosphorylated RNA reacted 360-fold more slowly than RNA beginning with a 5'-triphosphate.

Ribozyme rate enhancement

The uncatalyzed rate of formation of 5'–5' tetraphosphate-linked RNA was so slow that we were unable to detect any product after a 24 h incubation under the conditions and reactant concentrations in which the selection was carried out. We therefore developed a more sensitive assay for reaction between the substrate and the 5'-triphosphate of 2',3'-dideoxy ATP. Dideoxy adenosine triphosphate was used to prevent reaction at the 2' and 3' positions, thus simplifying the analysis of product formation. Using labeled substrate and high concentrations (25 mM) of 2',3'-dideoxy ATP, followed by P1 nuclease digestion and thin-layer chromatography, we were able to measure a second-order rate constant k_{uncat} of ~6 x 10^{-4} min^{-1} M^{-1} for 5'–5' tetraphosphate formation.

Comparison of the second-order rate constant for the uncatalyzed reaction with k_{cat}/K_m for the class 1 54-nucleotide deletion derivative (cI-54) RNA (1.1 x 10^4 min^{-1} M^{-1}) is not particularly meaningful, since most of the substrate-binding energy is derived simply from base pairing, and the calculated rate enhancement could be arbitrarily changed by increasing or decreasing the number of ribozyme–substrate base pairs. The ratio k_{cat}/k_{uncat} yields an effective substrate concentration of 1.5 x 10^2 M. This value suggests that the ribozyme could be simply bringing the reactive phosphates close to each other; alternatively the ribozyme may also constrain the relative orientation of the reacting groups. The k_{cat} for the selected ribozyme is faster by a factor of 10^3–10^4 than the rate of the uncatalyzed but template-directed 3'–5' ligation reaction (roughly 10^{-4} min^{-1}; R. Rohatgi and J.W.S., unpublished data), a situation involving the same activated phosphate and highly constrained reacting groups.

The oligonucleotide substrate-binding site

Although the selected RNAs do not use the template provided in the 3' constant region, examination of the sequence of both class 1 and class 2 ribozymes revealed an internal site with the potential to bind the substrate by Watson–Crick base pairing. The class 2 ribozymes contain an internal site that is completely complementary to the substrate oligonucleotide (six Watson–Crick base pairs) while the corresponding site in the class 1 ribozymes contains five Watson–Crick base pairs and one U:G wobble base pair. To assess the role of this putative internal substrate-binding site, five of the six nucleotides of this site in the class 1 ribozyme (cI-54) were altered by site-directed mutagenesis (Fig. 6a). The resulting mutant ribozyme (cI-54T) lacked detectable activity when assayed with the original substrate (Fig. 6b). However, when the mutant ribozyme was incubated with a new substrate oligonucleotide that was complementary to the altered substrate-binding site, ligation activity was restored (Fig. 6b, lane 4). A time course of this reaction indicated that at 10 μM substrate concentration, the reaction rate of this altered ribozyme with its matching substrate was only 6-fold slower than that of the original ribozyme–substrate combination. These results demonstrate that the internal site is in fact the substrate-binding site, and that the 5'–5' ligase is capable of ligating itself to different substrate sequences, provided that they base pair to the internal substrate-binding site.

Discussion

In the course of selecting for a ribozyme with a 3'–5' RNA ligase activity, we have identified a novel class of

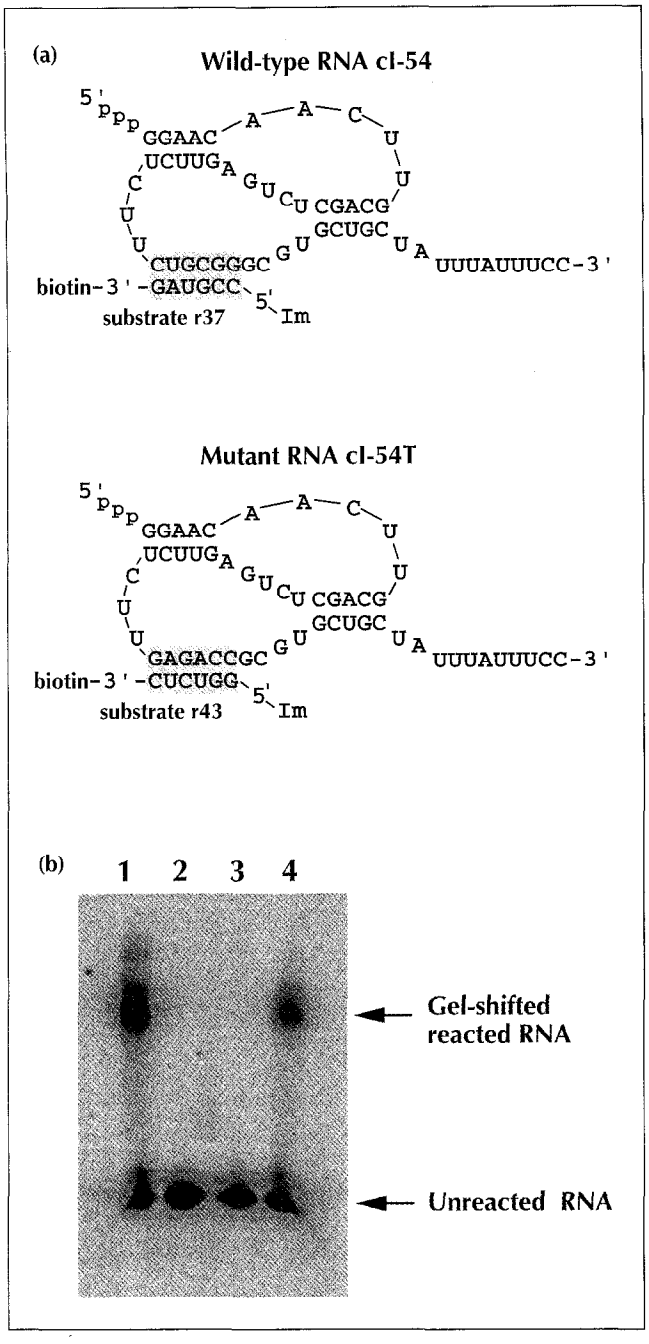

Fig. 6. Importance of the internal substrate-binding site. **(a)** Class 1 RNA contains an internal substrate binding site. The substrate binding site of cl-54 is boxed. The template mutant, cl-54T, contains five nucleotide changes in the shaded region. Base-pairing to the complementary substrate r43 is shown. **(b)** Reaction of [^{32}P]-labeled cl-54 RNA (lanes 1 and 2) or cl-54T RNA (lanes 3 and 4) with substrate r37 (lanes 1 and 3) or r43 (lanes 2 and 4), confirms the substrate-binding site indicated in (a).

ribozymes that ligate RNAs through a 5'–5'-tetraphosphate linkage. The selection of this novel class of ribozymes as opposed to the desired 3'–5' ligation activity is particularly remarkable in that all pool molecules contained a designed and built-in template for the 3'–5' ligation activity, whereas only about 1 % of the pool molecules would be expected to contain by chance an internal sequence complementary to the substrate. Therefore, the 5'–5' P^1,P^4-ligases were selected from a pool that was effectively about 100-fold smaller than the pool from which a template-directed 3'–5' ligase could have arisen. The selection of 5'–5' ligation activity, as opposed to 3'–5' ligation activity, may reflect the inherent ability of RNA to catalyze different types of reactions. For example, a nicked but continuously base-paired duplex may be such a sterically crowded environment that it is difficult to achieve catalysis by an RNA enzyme, whereas bringing a phosphorimidazolide and a triphosphate into proximity may create a reactive center that is much more accessible and around which it is easier to place functional groups that can effect catalysis.

This hypothesis is consistent with the results of previous *in vitro* selection experiments. For example, in the selection of RNA ligases that catalyze the reaction of the 2'- or 3'-hydroxyl of one oligonucleotide with the 5'-triphosphate of another oligonucleotide (which was in turn attached to the ribozyme), alignment of the substrates by continuous Watson–Crick base pairing to a nearby template was disfavored. All of the selected ligases used either partial, internal templates, or competing sequences to disrupt the designed, continuous pairing so that a more open ligation junction was generated [5]. Even a selection for optimal DNA substrates for the protein enzyme, DNA ligase, showed that ligation junctions with nearby base–base mismatches were preferred substrates [17]. None of the ribozymes found in nature efficiently cleave or ligate in the context of perfect RNA duplexes: for example, the group I introns contain a U:G wobble base pair at the cleavage/ligation site, perhaps to widen the major groove and increase accessibility to the reactive phosphate.

While we favor the idea that continuous base pairing to a template restricts access to the site of ligation, we cannot exclude the alternative hypothesis that the 5'–5' ligase we selected was favored for some other reason, for example, by being simpler and therefore more abundant in the initial pool than any possible 3'–5' ligases, by being more easily amplifiable, or because of the greater nucleophilicity of the attacking phosphate compared to the 3'-hydroxyl. Although substrate ligation at the designed site would extend the 3' hairpin and partially occlude the 3' primer binding site, this is unlikely to decrease the efficiency of the reverse transcription step significantly, due to the length of the 3' primer.

The selected RNAs also accelerate the formation of 5'–5' pyrophosphate and 5'–5' triphosphate linkages, although less efficiently than they accelerate the formation of 5'–5' tetraphosphate linkages. It is remarkable that a 5'-monophosphate, necessarily somewhat distant from the position of the γ-phosphate of a 5'-triphosphate, should be able to react at all with the same substrate phosphorimidazolide. The most likely explanation for this phenomenon is that the ribozyme itself is rather flexible and that internal motions in the ribozyme can result in the 5'-monophosphate being brought close to the substrate,

although less frequently than in the case of the γ-phosphate. It will be interesting to see if it is possible to select for mutations that compensate for this change in distance; such mutations could provide useful constraints for efforts to model the structure of this ribozyme.

The 5'–5' tetraphosphate, triphosphate and pyrophosphate linkages are found in a variety of biological contexts. All cell types that have been examined thus far have been found to contain diadenosine 5'–5' P^1,P^4-tetraphosphate (Ap_4A), a molecule that is synthesized by some cellular tRNA synthetases [18,19]. The levels of Ap_4A increase under conditions of oxidative stress and heat shock [19,20] and it has been proposed that Ap_4A acts as a cellular 'alarmone', a signaling molecule that alerts the cell to a condition of stress [18] and modulates the activities of cellular enzymes. Although the precise function of Ap_4A is unknown, Ap_4A has been implicated in the processes of DNA replication [18], platelet aggregation [21], hemoglobin function [22] and extracellular signaling [23].

The structure of the 5' cap of eukaryotic mRNAs is composed of a 7-methyl guanosine residue linked to the 5' end of the mRNA via a 5'–5' triphosphate linkage [16,24,25]. The cellular capping enzyme possesses two enzymatic activities, a 5' triphosphatase activity, which removes the γ-phosphate from the 5' terminus of the mRNA, and a guanylyltransferase activity, which reacts with GTP to form an enzyme–GMP intermediate and then transfers the GMP to the 5' diphosphate of the mRNA generating the structure G-5'-ppp-5'-N [26,27]. The cap is subsequently methylated at the N7 position of the terminal guanosine and at the 2' OH groups of the first and sometimes the second nucleotide of the mRNA chain. This 5' cap structure is found on all eukaryotic mRNAs and serves as the site for the binding of eIF–4E, an early step in the assembly of the ribosome–mRNA complex prior to the initiation of translation [28,29].

The 5'–5' pyrophosphate dinucleotides are a common product of activated nucleotide condensation reactions in the absence of templates [14], in the presence of oligonucleotide templates [14,15] or in the presence of montmorillonite clay [30]. Condensation of the 5'-phosphorimidazolide of adenosine in the presence of montmorillonite has shown that A-5'-pp-5'-A is readily incorporated in RNA oligomers and is most prevalent at the 5' position, thus yielding pyrophosphate 'capped' RNA oligomers [30]. Also, under certain conditions, the most abundant products of oligonucleotide-templated condensation reactions are capped with nucleoside pyrophosphates [15]. This has led to the suggestion that RNA oligonucleotides formed under pre-biotic conditions would be likely to be capped by nucleoside pyrophosphates [31]. Pyrophosphate linkages are also the substrates for modern day DNA and RNA ligases. It has been suggested that the use of pyrophosphate in the ligation mechanism may represent a 'molecular fossil', reflecting earlier times when pyrophosphate linkages were prevalent [32]. Many common metabolic cofactors, such as NAD, FAD and CoA also contain pyrophosphate linkages. Our results show that the di-, tri- and tetraphosphate linkages discussed above can be readily synthesized by ribozymes from activated substrates, and might thus have been used throughout the ribozyme-dominated period known as the RNA world.

Significance

We have described the isolation of a new 54-nucleotide ribozyme that can generate 5'–5' tetraphosphate-, triphosphate- or pyrophosphate-linked RNAs. While it is remarkable that this is achieved by such a small and simple ribozyme, it is equally surprising that no ribozymes were selected that accelerated the seemingly simpler template-directed 3'–5' ligation. An interesting possible explanation for this difference is that the steric environment within a continuous RNA duplex is too crowded to allow effective catalysis by RNA, with its bulky nucleotide subunits. If so, models of RNA replication that invoke a primitive replicase that assembles a cRNA strand by ligating together a series of oligonucleotides may have to be discarded in favor of the use of mononucleotide substrates from the beginning.

The reaction products formed by this novel class of ribozymes, 5'–5' oligophosphate linkages, are common in a variety of biological contexts. Aminoacyl-tRNA synthetases synthesize the signaling molecule Ap_4A, a 5'–5' triphosphate linkage is part of the cap structure of all eukaryotic mRNAs and 5'–5' pyrophosphate linkages serve as an intermediate for both DNA and RNA ligases and are also found in many cofactors. The ability of ribozymes to catalyze reactions that generate biologically relevant molecular structures, and the existence of such ribozymes in a relatively small sampling of sequence space, supports the notion that RNA enzymes had an important role in evolution prior to the emergence of protein enzymes.

Materials and methods

In vitro *selection*

Ligation reactions contained 1 μM random pool RNA, 10 mM activated substrate, 500 mM KCl, 50 mM $MgCl_2$, 30 mM TrisCl pH 7.4, and 10 mM NH_4Cl. All reaction components except the activated substrate were mixed and heated to 80 °C for 2 min to denature the RNA and then allowed to cool to room temperature. Reactions were initiated by the addition of the activated substrate and incubated at room temperature for the time indicated in Table 1. The pool RNA was then ethanol precipitated, resuspended in 100 μl of streptavidin binding buffer (1 M NaCl, 5 mM EDTA, 10 mM HEPES pH 7.4) and incubated with 500 μl of 50 % streptavidin agarose (Pierce) for 30 min with gentle rolling. The streptavidin agarose slurry was transferred to a column and was washed with 5 ml of binding buffer, 5 ml dH_2O, 5 ml 3 M urea (rounds 3–8), and 5 ml

dH_2O. Bound RNAs were eluted by washing the streptavidin agarose with 2 x 0.5 ml of 10 mM biotin pH 7. The streptavidin agarose was then resuspended in 0.5 ml 10 mM biotin and heated to 94 °C for 8 min [9]. Eluted RNAs were ethanol precipitated, reverse transcribed, PCR amplified and T7 transcribed as previously described [4]. The 5' PCR primer, 40.70 is 5'-TTCTAATACGACTCACTATAGGAACAACTTGCAGC TCTGA-3' and the 3' primer, 39.43 is 5'-CACGATAAGTGT-CGTGCCGTAGGACTTACGATGTACGAC-3'. The substrate oligonucleotide was imidazole-activated by incubation in a small volume (typically 20 μl) containing 0.13 M imidazole solution, pH 6 and 0.19 M 1-ethyl-3,3-dimethylaminopropyl carbodiimide (EDC) for 1 h at room temperature.

Oligonucleotide synthesis

All oligonucleotides were synthesized on a Millipore Expedite oligonucleotide synthesizer. The 3' biotinylated RNAs were synthesized on bioTEG columns (Glen research) and chemically phosphorylated on the column with phosphorylating reagent (Glen research). Synthetic RNAs were purified by high performance liquid chromatography (HPLC). Mutant versions of the cI-54 ribozyme were generated by T7 transcription of synthetic oligodeoxynucleotides.

Cloning and sequencing

Round 8 DNA was cloned using pT7 Blue T-Vector Kit (Novagen) and sequenced using standard dideoxy-nucleotide termination techniques. Deletion mutants were constructed from clone s9c31 by PCR amplification using primers complementary to internal regions of the RNA.

Nuclease digestion and thin layer chromatography

Nuclease P1 digests were done in 30 mM sodium acetate, pH 5.3, in a 5 ml volume. Thin-layer chromatography was carried out on PEI cellulose plates and developed in 0.75 M KH_2PO_4, pH 3.5. Products were visualized with a phosphoimager or by autoradiography.

Ribozyme assays

Ribozyme activity was assayed in selection buffer at 1 μM ribozyme and 10 μM substrate. Streptavidin gel shift assays were conducted by mixing 2 μl aliquots from the reaction mixture with excess (2 μg) streptavidin prior to mixing with an equal volume of formamide loading buffer. Samples were then subjected to electrophoresis on an 8 M urea, 6 % polyacrylamide gel, and the separated products were visualized with a phosphoimager or by autoradiography. Streptavidin binding causes biotinylated RNAs to migrate more slowly.

Acknowledgements: We thank B. Cuenoud, J. Lorsch and R. Roberts for discussions and comments on the manuscript. K.B.C. was supported by a postdoctoral fellowship from the Helen Hay Whitney Foundation. This work was supported by a grant from NASA.

References

1. Szostak, J.W. (1992). *In vitro* genetics. *Trends Biochem. Sci.* **17**, 89–93.
2. Lorsch, J.R. & Szostak, J.W. (1994). *In vitro* selection of nucleic acid sequences that bind small molecules. In *Molecular Repertoires and Methods of Selection*. (Cortese, R. ed.) Walter de Gruyter and Company, Berlin.
3. Pan, T., & Uhlenbeck, O.C. (1992). *In vitro* selection of RNAs that undergo autolytic cleavage with Pb^{2+}. *Biochemistry* **31**, 3887–3895.
4. Bartel, D.P., & Szostak, J.W. (1993). Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418.
5. Bartel, D.P., Ekland, E.H. & Szostak, J.W. (1995). Structurally complex and highly active ligases derived from random RNA sequences. *Science*, in press.
6. Lorsch, J.R. & Szostak, J.W. (1994). *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31–36.
7. Chapman, K.B., & Szostak, J.W. (1994). *In vitro* selection of catalytic RNAs. *Curr. Opin. Struct. Biol.* **4**, 618–622.
8. Prudent, J.R., Uno, T. & Schultz, P.G. (1994). Expanding the scope of RNA catalysis. *Science* **264**, 1924–1927.
9. Wilson, C. & Szostak, J.W. (1995). *In vitro* evolution of a self-alkylating ribozyme. *Nature* **374**,777–782.
10. Illangesekare, M., Sanchez, G., Nickles, T. & Yarus, M. (1995). Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643–647.
11. Zaug, A.J. & Cech, T.R. (1986). The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* **231**, 470–475.
12. Been, M.D. & Cech, T.R. (1988). RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the *Tetrahymena* ribozyme. *Science* **239**, 1412–1416.
13. Doudna, J.A. & Szostak, J.W. (1989). RNA-catalyzed synthesis of complementary-strand RNA. *Nature* **339**, 519–524.
14. Sulston, J., Lohrmann, R., Orgel, L.E. & Miles, H.T. (1968). Nonenzymatic synthesis of oligoadenylates on a polyuridylic acid template. *Proc. Natl. Acad. Sci. USA* **59**, 726–733.
15. Inoue, T. & Orgel, L.E. (1982). Oligomerization of (guanosine 5'-phosphor)-2-methyl-imidazole on poly(C). *J. Mol. Biol.* **162**, 201–217.
16. Furuichi, Y., Morgan, M., Muthukrishnan, S. & Shatkin, A.J. (1975). Reovirus messenger RNA contains a methylated, blocked 5'-terminal structure: $m^7G(5')ppp(5')G^mpCp$-. *Proc. Natl. Acad. Sci. USA* **72**, 362–366.
17. Harada, K. & Orgel, L.E. (1993). Unexpected substrate specificity of T4 DNA ligase revealed by *in vitro* selection. *Nucleic Acids Res.* **21**, 2287–2291.
18. Varshavsky, A. (1983). Diadenosine 5', 5'''-P^1,P^4-tetraphosphate: a pleiotropically acting alarmone? *Cell* **34**, 711–712.
19. Johnstone, D.B. & Farr, S.B. (1991). ApppppA binds to several proteins in *Escherichia coli* including the heat shock and oxidative stress proteins DnaK, GroEL, E89, C45 and C40. *EMBO J.* **10**, 3897–3904.
20. Bochner, B.R., Lee, P.C., Wilson, S.W., Cutler, C.W. & Ames, B.N. (1984). AppppA and related adenylylated nucleotides are synthesized as a consequence of oxidation stress. *Cell* **37**, 225-232.
21. Zamecnik, P.C., Kim, B., Gao, M.J., Taylor, G. & Blackburn, G.M. (1992). Analogues of diadenosine 5',5'''-P^1,P^4-tetraphosphate (Ap_4A) as potential anti-platelet-aggregation agents. *Proc. Natl. Acad. Sci. USA* **89**, 2370–2373.
22. Bonaventura, C., Cashon, R., Colacino, J.M. & Hilderman, R.H. (1992). Alteration of hemoglobin function by diadenosine 5',5'''-P^1,P^4-tetraphosphate and other alarmones. *J. Biol. Chem.* **267**, 4652–4657.
23. Hilderman, R.H., Martin, M., Zimmerman, J.K. & Pivorun, E.B. (1991). Identification of a unique membrane receptor for adenosine 5',5'''-P^1,P^4-tetraphosphate. *J. Biol. Chem.* **266**, 6915–6918.
24. Cory, S. & Adams, J.M. (1975). The modified 5'-terminal sequences in messenger RNA of mouse myeloma cells. *J. Mol. Biol.* **99**, 519–547.
25. Adams, J.M. & Cory, S. (1975), Modified nucleosides and bizarre 5'-termini in mouse myeloma mRNA. *Nature* **255**, 28–33.
26. Furuichi, Y. & Shatkin, A. (1989). Characterization of cap structures. *Methods Enzymol.* **180**, 164–176.
27. Fresco, L.D. & Buratowski, S. (1994). Active site of the mRNA-capping enzyme guanylyltransferase from *Saccharomyces cerevisiae*: similarity to the nucleotidyl attachment motif of DNA and RNA ligases. *Proc. Natl. Acad. Sci. USA* **91**, 6624–6628.
28. Kozak, M. (1978). How do eukaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109–1123.
29. Rhoads, R.E. (1988). Cap recognition and the entry of mRNA into the protein synthesis initiation cycle. *Trends Biochem. Sci.* **13**, 52–56.
30. Ferris, J.P. & Ertem, G. (1992). Oligomerization of ribonucleotides on montmorillonite: reaction of the 5'-phosphorimidazolide of adenosine. *Science* **257**, 1387–1389.
31. Rodriguez, L. & Orgel, L.E. (1991). Pyrophosphate formation as the most efficient condensation reaction of activated nucleotides. *J. Mol. Evol.* **32**, 101–104.
32. Orgel, L.E. (1986). RNA catalysis and the origins of life. *J. Theor. Biol.* **123**, 127–149.

Received: **19 Apr 1995**; revisions requested: **2 May 1995**; revisions received: **4 May 1995**. Accepted **4 May 1995**.