Isolation of novel ribozymes that ligate AMP-activated RNA substrates

Alicia J Hager and Jack W Szostak

Background: The protein enzymes RNA ligase and DNA ligase catalyze the ligation of nucleic acids via an adenosine-5′-5′-pyrophosphate 'capped' RNA or DNA intermediate. The activation of nucleic acid substrates by adenosine 5′-monophosphate (AMP) may be a vestige of 'RNA world' catalysis. AMP-activated ligation seems ideally suited for catalysis by ribozymes (RNA enzymes), because an RNA motif capable of tightly and specifically binding AMP has previously been isolated.

Results: We used *in vitro* selection and directed evolution to explore the ability of ribozymes to catalyze the template-directed ligation of AMP-activated RNAs. We subjected a pool of 10^{15} RNA molecules, each consisting of long random sequences flanking a mutagenized adenosine triphosphate (ATP) aptamer, to ten rounds of *in vitro* selection, including three rounds involving mutagenic polymerase chain reaction. Selection was for the ligation of an oligonucleotide to the 5′-capped active pool RNA species. Many different ligase ribozymes were isolated; these ribozymes had rates of reaction up to 0.4 ligations per hour, corresponding to rate accelerations of $\sim 5 \times 10^5$ over the templated, but otherwise uncatalyzed, background reaction rate. Three characterized ribozymes catalyzed the formation of 3′-5′-phosphodiester bonds and were highly specific for activation by AMP at the ligation site.

Conclusions: The existence of a new class of ligase ribozymes is consistent with the hypothesis that the unusual mechanism of the biological ligases resulted from a conservation of mechanism during an evolutionary replacement of a primordial ribozyme ligase by a more modern protein enzyme. The newly isolated ligase ribozymes may also provide a starting point for the isolation of ribozymes that catalyze the polymerization of AMP-activated oligonucleotides or mononucleotides, which might have been the prebiotic analogs of nucleoside triphosphates.

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

Correspondence: Jack W Szostak
E-mail: szostak@frodo.mgh.harvard.edu

Key words: ATP aptamer, catalytic RNA, *in vitro* selection, ligase, ribozyme

Received: **23 June 1997**
Accepted: **17 July 1997**

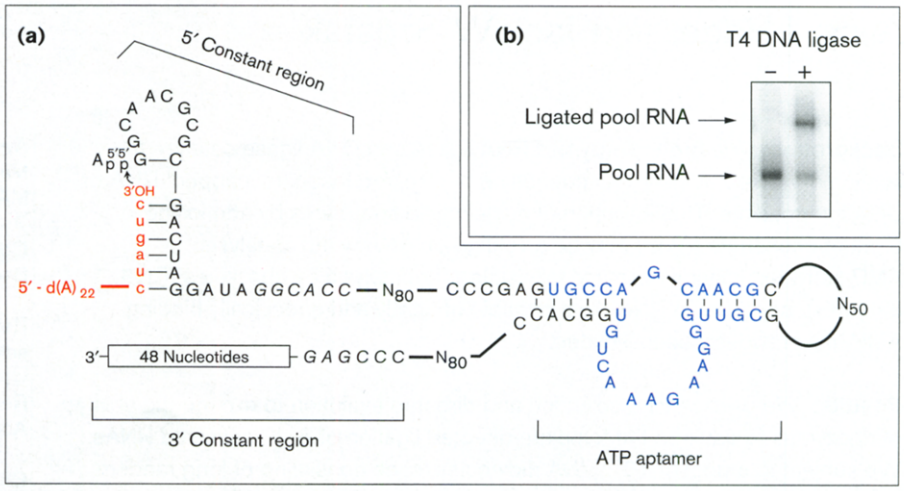
Chemistry & Biology August 1997, 4:607–617
http://biomednet.com/elecref/1074552100400607

Introduction

In vitro selection techniques have facilitated the isolation and optimization of numerous RNA and DNA enzymes, several of which catalyze the ligation of oligonucleotides. Ribozymes that ligate oligonucleotides activated by pyrophosphate (i.e. with 5′-triphosphates) have been isolated using *in vitro* selection from a random pool [1,2]. Further *in vitro* evolution of one of these ribozymes has led to an extremely efficient ligase that is capable of catalyzing the template-directed polymerization of nucleoside triphosphates [2,3]. Similar selection techniques have been used to manipulate and optimize the ligation activity of a ribozyme derived from group I introns [4]. More recently, a DNA enzyme that ligates imidazole-activated oligonucleotides has been isolated from a pool of random sequences [5].

Much of the interest in ligase ribozymes stems from the similarity between oligonucleotide ligation and mononucleotide polymerization. This similarity suggests that ribozymes that catalyze these two types of reaction might be closely related to each other, such that either function might be derived from the other by evolution [6]. The mechanism by which RNA self-replicated in the 'RNA world' is a matter of much speculation. It has been proposed that 5′-5′-pyrophosphate linkages may have played a significant role in prebiotic ligation and polymerization, based upon the accumulation of 5′-5′-pyrophosphate-linked dinucleotides in reactions containing highly activated nucleotides such as phosphorimidazolides [7]. Adenosine-5′-5′-pyrophosphate moieties are also found in some of the most important enzymatic cofactors, such as coenzyme A, NAD^+, and FAD, which may be relics of the RNA world [8,9]. Similarly, it has been proposed that the mechanism of the protein enzymes DNA ligase and RNA ligase, which use nucleic acid intermediates bearing adenosine-5′-5′-pyrophosphate moieties, might be a vestige of the RNA world [7].

Figure 1



Pool RNA. **(a)** A schematic of starting pool RNA and the designed substrate-binding site. The 5′-phosphate of the first guanosine of pool transcripts was activated with adenosine 5′-monophosphate (AMP). Substrate oligonucleotides (like substrate I, shown here in red) were designed to bind a region near the 5′-end of the pool. This stem–loop structure templates the possible attack of the 3′-hydroxyl of the substrate onto the 5′-phosphate of the guanosine, resulting in ligation and the loss of AMP. N represents a random nucleotide. The *Ava*I and *Ban*I sites are shown in italics. Adenosine triphosphate (ATP) aptamer sites that were mutagenized at 15% (5% to each of the three possible other bases) are shown in blue. **(b)** Ligation of substrate to (AppG)-primed pool RNA (AppG: P^1-[adenosine-5′]-P^2-[guanosine-5′]-pyrophosphate) by T4 DNA ligase in the absence of ATP. When substrate oligonucleotide and internally ^{32}P-labeled AppG-primed pool RNA were pre-annealed to a DNA template oligonucleotide that spanned the ligation junction (and usurped the designed stem–loop structure), T4 DNA ligase efficiently ligated the substrate to the pool. No ATP was included in the reaction buffer. In the reaction shown here, substrate III (see the Materials and methods section) was ligated to AppG-primed round 4 pool RNA (see Figure 2 and Table 1). Similar results were obtained from pool RNA of other rounds, including round 0 (the starting pool).

We sought to investigate the ability of RNA enzymes to catalyze the template-directed ligation of AMP-activated (AMP: adenosine 5′-monophosphate) RNA substrates, a reaction analogous to that catalyzed by the protein enzymes DNA ligase and RNA ligase. In an effort to favor interaction between the ribozyme and the activating group at the ligation junction, a previously isolated adenosine triphosphate (ATP) aptamer [10] was incorporated within the context of a complex pool of long, random RNAs. Here, we describe the isolation and initial characterization of multiple, independent AMP-activated ligase ribozymes from the pool using *in vitro* selection.

Results

Pool design and preparation

The initial sequence pool consisted of RNAs containing an ATP aptamer (which also binds AMP) [10–12] embedded within random regions consisting of two 80 nucleotide segments and one 50 nucleotide segment (Figure 1a). The ATP aptamer was mutagenized at a rate of 15% per position. Regions of constant sequence flanked the pool on both ends. Two 6 nucleotide regions overlapping one of the aptamer stems also remained invariant, because these regions served as sites for restriction digestion and ligation during the construction of the pool. The complexity of the completed pool was ~2×10^{15} sequences (~760 μg isolated ligated pool DNA). The total transcript length of the completed pool RNA was 332 nucleotides.

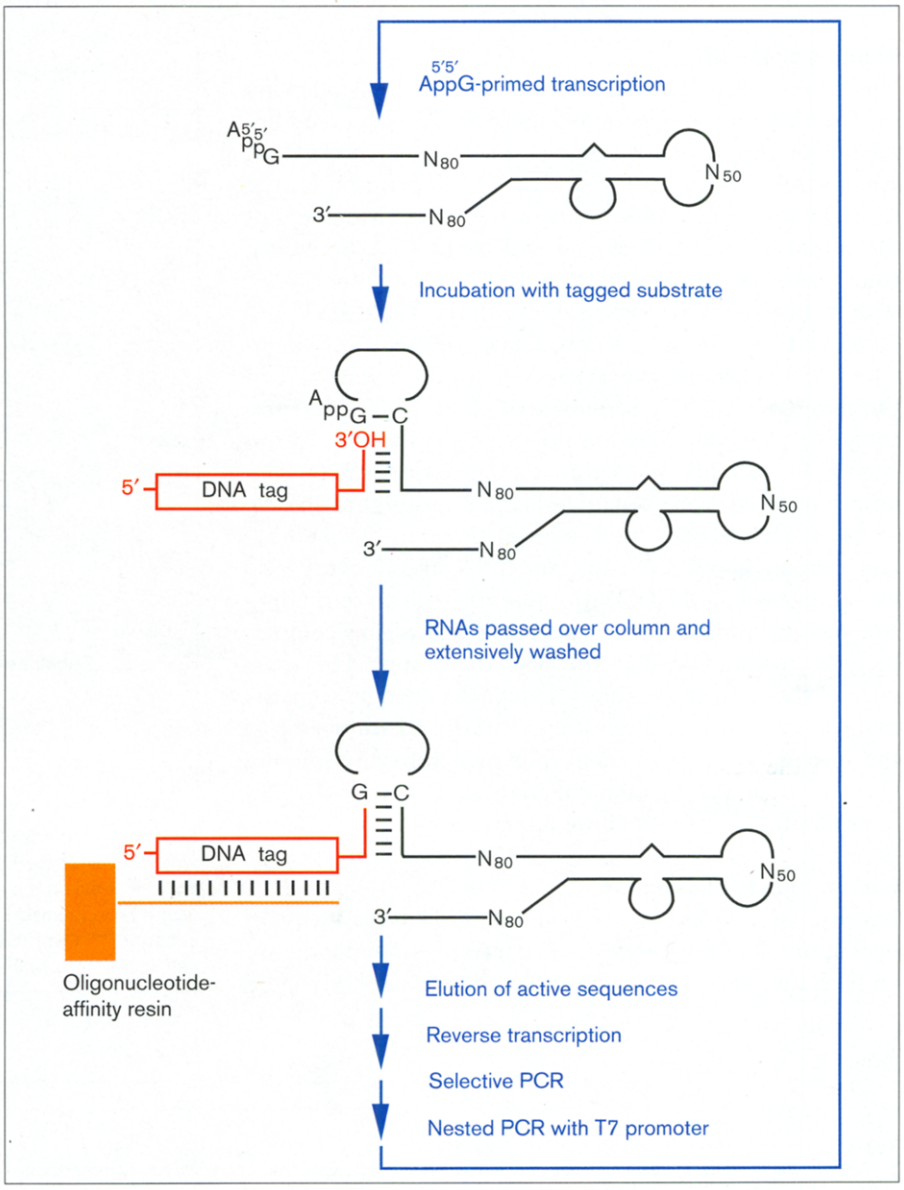
The selection protocol we used required that pool RNA molecules be synthesized with P^1-(adenosine-5′)-P^2-(guanosine-5′)-pyrophosphate (AppG) at their 5′-termini (Figure 1). This was achieved by priming transcription with AppG at a concentration threefold higher than that of the guanosine triphosphate (GTP) present in the transcription reaction. AppG was readily formed by the reaction of the 5′-phosphorimidazolide of adenosine (ImpA) with guanosine 5′-monophosphate (GMP) [13]. AppG incorporation at the beginning of transcripts was verified by demonstrating that AppG-primed pool RNA transcripts serve as substrates for T4 DNA ligase in the absence of ATP (Figure 1b). Typically, 61–65% of pool RNA transcripts were rapidly ligated by T4 DNA ligase. Model studies with short transcripts, the size of which allowed for separation of 5′-AppG-primed and GTP-primed transcripts, indicated that ~70% of the RNA should have been correctly activated under the transcription conditions used (data not shown).

Selection

The initial phase of the selection proceeded for four cycles (Figure 2, Table 1). Following transcription, pool RNA was annealed to a biotinylated oligodeoxynucleotide complementary to its 3′ constant region and the complex was applied to a streptavidin agarose column. This immobilization was designed to prevent aggregation of the pool RNA in the presence of Mg^{2+} [1]. The solid-support-bound pool RNA was then incubated with a buffer containing substrate oligonucleotide and high concentrations of Mg^{2+} and KCl. The substrate oligonucleotides consisted of 6 nucleotides of RNA sequence at the 3′ end and DNA 'tag' sequences of 20–25 nucleotides at the 5′ end. The RNA portion of the substrate was complementary to a constant region near the 5′ end of the pool transcripts

Figure 2

The scheme for *in vitro* selection of pool RNAs that ligate tagged substrate oligonucleotides to their AMP-activated 5′ end. Abbreviations as in Figure 1; round 4 pool is the pool generated by four successive rounds of selection.



(Figure 1a). After the reaction incubation, the pool RNA was eluted from the streptavidin agarose. Active RNAs that had covalently attached themselves to a substrate oligonucleotide were separated from inactive and unreacted RNAs on an affinity selection column. The selection column contained immobilized oligodeoxynucleotides complementary to the 5′ DNA sequence (tag) of the substrate oligonucleotides. The pool RNAs were annealed with the selection column, RNAs lacking the tag sequence were removed by a series of washes, and specifically bound RNAs were then eluted. Following reverse transcription, polymerase chain reaction (PCR) with a forward primer sequence identical to portions of the substrate was used to selectively enrich for active sequences. The selection column resin and corresponding DNA-tag portion of

Table 1

Selection conditions for rounds 1–4.

Round	Amount of pool RNA	Incubation time	Selection column	Proportion of pool RNA selected (%)
1	1.5 mg	21 h	oligo d(T) cellulose	0.08
2	35 μg	21 h	d(A) magnetic beads	0.4
3	15 μg	16 h	d(T) magnetic beads	0.4
4	15 μg	16 h	dGAGATG...teflon	0.9

The proportion of pool RNA selected on the column (%) reflects the percentage of pool RNA that was retained on the column after extensive washings and recovered after elution.

the substrate oligonucleotide were varied from round to round, to discourage the propagation of selection artifacts.

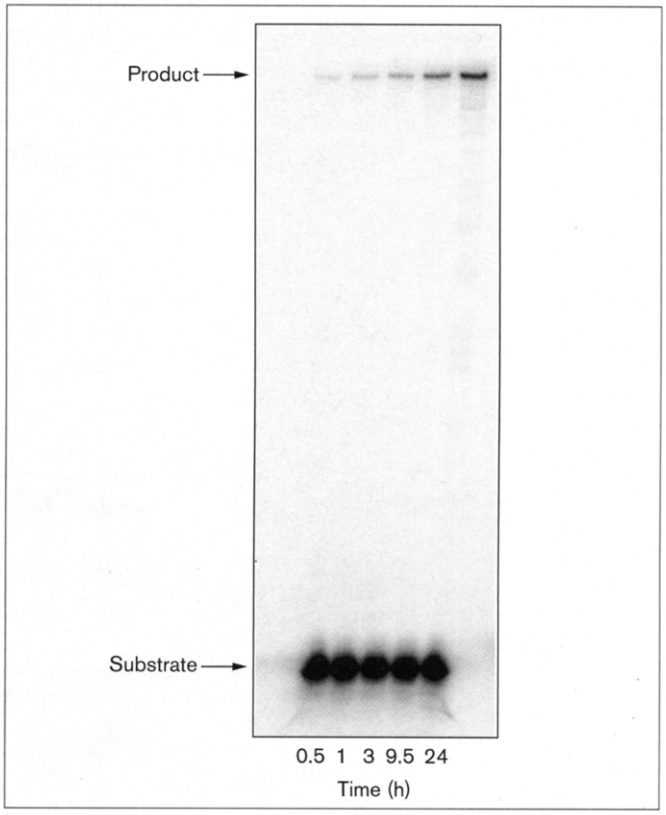
Round 4 pool RNA

Ligation activity by the pool was first detected following the fourth round of selection (Figure 3). The observed ligation rate (k_{obs}) for ligation by the round 4 pool RNA ranged from 1×10^{-3} h^{-1} to 7×10^{-3} h^{-1} depending on the substrate oligonucleotide used in the assay. Members of the round 4 pool were cloned and sequenced. Sequence comparison, as well as digestion of the round 4 pool DNA with a mix of three restriction enzymes (*Taq*I, *Mse*I, and *Hinf*I; data not shown), showed that the pool was dominated by a single ribozyme that constituted at least 80% of the round 4 pool. The sequence of clone H4004, a representative of this dominating ribozyme family, is shown in Figure 4a. In addition to this dominant ribozyme, seven other ribozymes were isolated (Table 2 and Figure 4b). All of the ribozymes contained an internal region complementary in sequence to the substrate RNA, and in every case except one (clone H4003) this potential substrate-binding site was positioned directly adjacent to a region complementary to the 5′ end of the pool transcripts. The presence of these complementary sequences strongly suggests that the ribozymes did not use the designed stem–loop at the 5′ end of the pool transcripts as a substrate-binding site, and that the ligation junction instead relied on an internal ribozyme sequence as a template.

Further rounds of selection

Because of the low activity and limited complexity of pool 4, the selection process was continued independently on both the pool and on one of the minor clones. Six more

Figure 3



Self-ligation by AppG-primed round 4 pool RNA. The pool RNA was first pre-annealed to the biotinylated primer used for pool immobilization in the selection. AppG-primed round 4 pool RNA was then incubated in 1 × ligation buffer with 6 μM 5′ ^{32}P-labeled substrate III. Over the incubation times, the round 4 pool RNA self-ligated to give the product observed on the gel.

Figure 4

(a)
```
24          34          44          54          64          74          84
ACCTTCGAAC CCATTTGGAG CCTTCTTGGC TTGTTAAAAA CCGAGGGGAT GCCTCTGCAA ACGCATGCAT

94          104         114         124         134         144         154
CTTGCACCCT CCCAACCCCG AGTGCCCGCA ACGCATCACT TGCATCCTCC CGCCGAATTG ATTTGAAATG

164         174         184         194         204         214         224
AACTATTCTA GTCCGTTCCG AAAAACCTGT GGCACC CGAG CACGGAATT[redacted]TTCATTT

234         244         254         264         274         284         294
CCTGACATTC GTCTAGCACT ATGACACGTT GGATAAGTCT CCCGAG ACCG TATAGCATGT C
```

(b)
```
24          34          44          54          64          74          84
ACC TCTATTG TAACACCTTT TTACAATATC TGCGAACCCC AAGGGTCTCG GTGGATATGA GCGTCACGTT

94          104         114         124         134         144         154
GGCCATAATG TCTGCCCGAG TCCCACAACA CGACCAGAAG CATTCTCCAG AGGCGATTCG GACTCCACGG

164         174         184         194         204         214         224
GACTCGTTAA CCGCGCTGGC AACAAACTGT GGCACC CACC AAATTTCGAA CACCGCTACC GCGTCCCCGA

234         244         254         264         274         284         294
TGCCCAAACC TACTTCCCTA GCTCTAGAAA CTCAGAGAAG [redacted]T ACCCGAGACC GTATAGCATG

304
TC
```

Sequences of selected round 4 clones. The first 23 and last 33 nucleotides of the sequences are not shown; these regions consist entirely of constant sequence complementary to PCR primers used in the selection. Regions that were initially constant, but which were free to mutate over the course of the selection, are underlined; the restriction digest sites are italicized; originally mutagenized aptamer regions are in blue; the region in the shaded box is likely to act as a template for the ligation junction; internal sites complementary to the substrate RNA are in green; magenta nucleotides are complementary to the 5′ end of pool transcripts. **(a)** Internal sequence of clone H4004. **(b)** Internal sequence of clone H4853.

Table 2

Activities of AppG-primed and GTP-primed round 4 clones with substrate oligonucleotides II and III.

	k_{obs} (h^{-1})		
Clone	AppG/substrate II	AppG/substrate III	GTP/substrate III
H4004	7×10^{-4}	1×10^{-2}	n.d.
H4853	6×10^{-4}	6×10^{-3}	n.d.
H4823	4×10^{-4}	3×10^{-3}	n.d.
H4742	6×10^{-4}	3×10^{-3}	n.d.
H4009	9×10^{-4}	2×10^{-3}	n.d.
H4028	2×10^{-4}	2×10^{-3}	n.d.
H4003	1×10^{-3}	6×10^{-4}	1×10^{-4}
H4736	4×10^{-3}	1×10^{-4}	n.d.

k_{obs}: observed ligation rate. n.d.: no product was detected. See the Materials and methods section for details of substrates II and III.

rounds of selection (selection A) were carried out on the round 4 pool. In the first three of these additional selection cycles, the pools were subjected to mutagenic PCR. In addition, increasingly stringent selection conditions were used to select for more efficient ribozymes (Table 3). Mutagenic PCR conditions in each of the rounds 5, 6 and 7 resulted in an estimated 2% mutagenesis rate per position per round, giving a total of 6% per position. In addition to selection A, a separate selection (selection B) was started from the isolated clone H4853. This clone was evolved separately in order to ensure the propagation of at least one sequence other than the pool-dominating clone H4004. Clone H4853 was chosen because of its relatively high ligation activity; selection B was otherwise identical in procedure to selection A (Table 3). The complexity of the pool in selection B was entirely derived from the PCR mutagenesis performed in rounds 5–7.

Round 10 pool RNA

After 10 rounds of selection, a 10–100-fold increase in ligation rate was observed with the pools from both selections A and B. We also observed a decrease in the effect on the ligation rate of changing the DNA-tag region of the substrate. The final selection A pool ligated to both substrates II and III (see the Materials and methods section) at a rate of 0.10 h^{-1}, while the final selection B pool ligated at rates of 0.04 h^{-1} and 0.038 h^{-1}, respectively.

DNA from both round 10 pools was cloned and 74 sequences from selection A and 37 sequences from selection B were obtained. From the selection A pool, a family of 38 sequences unrelated to any of the ribozymes identified at round 4 was discovered. Like all the other sequences, however, the members of this ribozyme family contained an internal template for the ligation reaction. The originally dominant sequence in round 4 — that of clone H4004 — was represented by 24 clones of the 74 round 10 sequences. The remaining clones from selection A were all related to the other ribozymes identified at round 4, and 34 of the selection B pool sequences were related to the founding clone H4853. Three other selection B sequences were unrelated to H4853, and were believed to be contaminants from selection A.

Sequencing across the ligation junction of the PCR-amplified cDNA from round 4 and round 10 confirmed that ligation occurred at the correct site in all clones. In all cases, the substrate was directly ligated to the initial guanosine at the 5′ end of the pool transcript, and therefore occurred by displacement of the AMP-activating group.

Based in part on the patterns of conserved and variable bases observed in the aligned sequences (data not shown), a series of deletion constructs of ribozymes H4853, H4004 and H10140 (a member of the major ribozyme family from selection A, round 10) were generated and assayed for

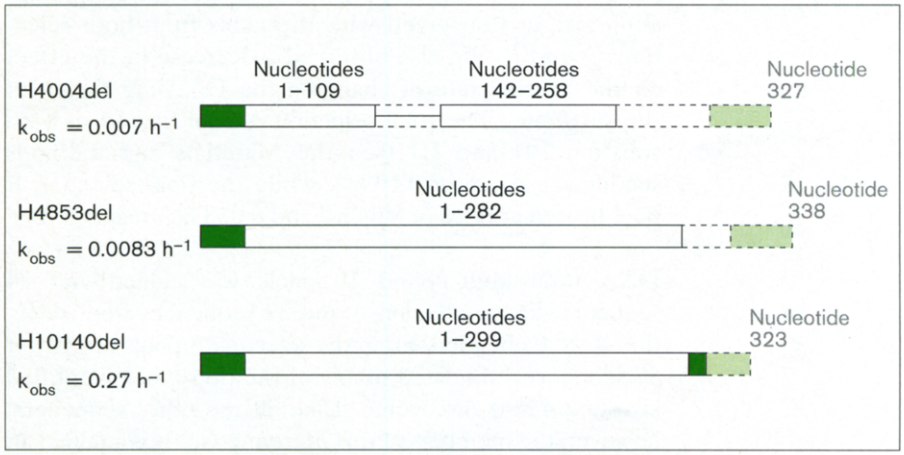
Table 3

Conditions for rounds 6–10 of selections A and B.

Round	Amount of pool RNA	Incubation time	Selection column	Proportion of pool RNA selected (%) A	Proportion of pool RNA selected (%) B	Mutagenic PCR
Round 5	10 μg	3.5 h	d(A) magnetic beads	0.5	0.4	30 doublings
Round 6	10 μg	3.5 h	d(A) magnetic beads	8.4	1.2	30 doublings
Round 7	5 μg	2.5 h	dGAGATG...teflon	0.3	0.2	30 doublings
Round 8	4–10 μg	30 min	dGAGATG...teflon	0.1	0.4	No
Round 9	10 μg	10 min	dGAGATG...teflon	0.2	0.1	No
Round 10	15 μg	2 min	d(T) magnetic beads	0.2	0.4	No

Selection A: rounds 6–10 of selection with pool RNA. Selection B: rounds 6–10 of selection with clone H4853. 30 doublings indicates that the template DNA in the PCR reaction was amplified an estimated 2^{30} times.

Figure 5



Deletion mutants representing the three main round 10 sequence families and their activities. Portions of the ribozymes H4004, H4853 and H10140 were deleted to form H4004del, H4853del and H10140del, respectively. Each of the deletion ribozymes is shown schematically along with the original full-length sequence (lighter colors and dashed lines). The green portions correspond to the flanking constant regions of the pool. Full-length H10140 self-ligated at a rate of 0.2 h^{-1}.

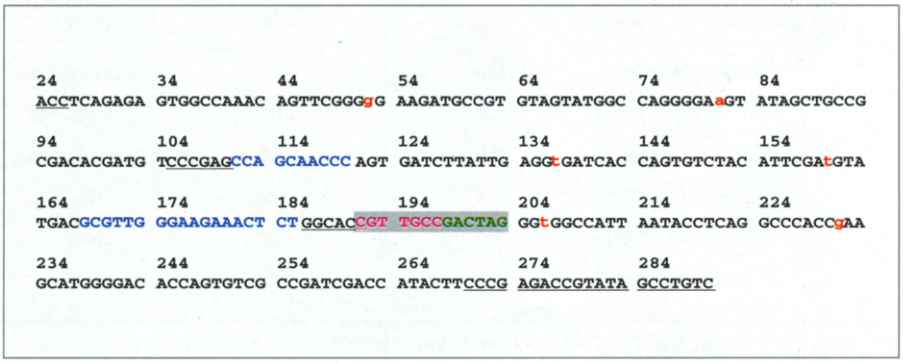
ligation activity (Figure 5). Only limited 3′-terminal deletions of H4853 and H10140 retained full activity. The remaining internal sequences were quite conserved, so no further deletion analysis was undertaken. Although little of the H4004 family sequence was strongly conserved, deletions in most of the unconserved regions resulted in decreased activity. The internal deletion of nucleotides 110–141, which occurred in a significant fraction of the sequenced clones from round 10, was an exception and resulted in no decrease in activity.

Ribozymes CM2g and CM3 were then generated by introducing a subset of the most highly enriched mutations observed in the aligned round 10 sequences into deleted versions of the H10140 family consensus sequence and the H4853 sequence, respectively. The sequence of CM2g is shown in Figure 6. The k_{obs} for ligation of this ribozyme was 0.38 h^{-1}. The activity of CM2g is almost twofold better than that of H10140 (0.2 h^{-1}), which contains many, but not all, of the enriched mutations. CM3 is identical to H4853del except for the mutations T32→C, A58→T, A59→G, A126→T, A201→T, A263→G, and A269→G. These mutations were each present at a high frequency among round 10 members of the H4853 sequence family. CM3 self-ligated at a rate of 0.046 h^{-1}, compared to 0.0058 h^{-1} for the ancestral sequence H4853. Incorporation of enriched mutations from round 10 sequences either decreased or had no effect on the initial rate of ligation observed for clone H4004. Therefore, the original full-length H4004 sequence was used in all further studies of this ribozyme.

Activating-group specificity

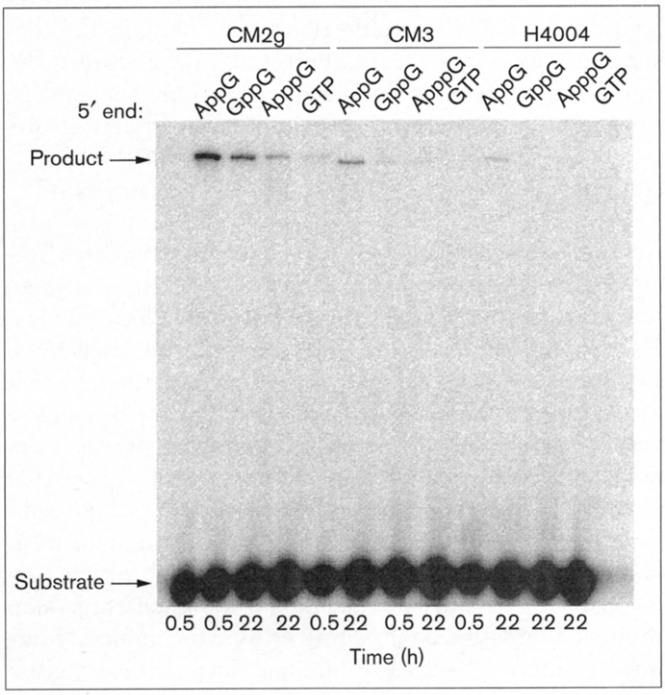
A significant fraction of pool transcripts would be expected to have GTP rather than AppG at their 5′-termini (up to 35%, Figure 1b). To establish the identity of the activating group involved in the observed ligation reactions (and the specificity of the various ribozymes for the activating group), transcription of CM2g, CM3, and H4004 was primed with GTP only, P^1-(adenosine-5′)-P^3-(guanosine-5′)-triphosphate (ApppG), or P^1,P^2-(diguanosine-5′)-pyrophosphate (GppG), and the ligation activity of the resulting RNAs was compared to the activity of the corresponding AppG primed RNAs (Figure 7, Table 4). GTP-primed CM2g exhibited a decrease in ligation rate of > 4000-fold, while GTP-primed CM3 and H4004

Figure 6



The internal sequence of the mutant ribozyme CM2g. The sequence of CM2g is identical to the consensus sequence of the H10140 sequence family except for the nucleotides shown in red lower-case letters and the deletion of nucleotides 300–323. The red lower-case letters represent sites where a significant fraction of the members of the sequence family showed the same, non-consensus base. These enriched mutations were incorporated into CM2g. Other markings in the text are as described in Figure 4. The first 23 and last 9 nucleotides were invariant in the selection and are not shown.

Figure 7



Self-ligation of ribozymes activated by AMP, GMP, adenosine diphosphate (ADP), or pyrophosphate. CM2g, CM3 and H4004 RNAs, each primed in transcription with AppG, P^1,P^2-(diguanosine-5′)-pyrophosphate (GppG), P^1-(adenosine-5′)-P^3-(guanosine-5′)-triphosphate (ApppG), or GTP, were incubated with 5′ ^{32}P-labeled substrate III under standard assay conditions; 0.5 h and 22 h reactions are shown.

showed no detectable ligation activity. In fact, none of the ribozymes isolated from round 4 preferred GTP at the ligation junction (Table 2). Similarly, none of the ribozymes CM2g, CM3, or H4004 ligated efficiently when their transcription was primed with ApppG. GppG-primed CM2g, however, ligated at a rate only threefold lower than the AppG-primed ribozyme, while GppG-primed CM3 or H4004 showed little or no detectable activity.

Ribozymes catalyze 3′-5′ ligation

Because many previously isolated ribozyme ligases catalyze the formation of 2′-5′ phosphodiester bonds [14], we wished to establish whether the current generation of ligases, which uses a different activating group, is predominantly composed of 2′-5′ ligases or 3′-5′ ligases. Ligases that generate 2′-5′ linkages are not strongly selected against during the reverse transcription step of amplification, because reverse transcriptase is able to copy templates containing 2′-5′ linkages under the conditions used [15]. The nature of the linkage formed by the ribozymes CM2g, CM3, and H4004 was determined with substrates carrying modified 3′-termini (Table 5). Each of the three ribozymes was able to ligate to substrates possessing either a 2′-methoxycytosine or a 2′-deoxycytosine at the

Table 4

Comparison of activities of alternatively activated ribozymes.

	k_{obs} (h^{-1})			
	AppG	GppG	AppG	GTP
CM2g	3.8×10^{-1}	1.4×10^{-1}	4.8×10^{-4}	8.9×10^{-5}
CM3	4.6×10^{-2}	9.5×10^{-5}	7.2×10^{-5}	n.d.
H4004	1.2×10^{-2}	n.d.	2.5×10^{-5}	n.d.

k_{obs}: observed ligation rate. n.d.: no product was detected ($k_{obs} < \sim 1 \times 10^{-5}\ h^{-1}$).

3′-terminal position, although the rates decreased by 10–20-fold and 100–400-fold, respectively. In contrast, none of the ribozymes were able to ligate detectably to a 3′-deoxy substrate. These results strongly suggest that the CM2g, CM3, and H4004 ribozymes all catalyze the synthesis of 3′-5′-phosphodiester bonds.

Comparison to background reaction rate

Ligation was not detectable when a stem–loop, directly analogous to the designed stem–loop at the 5′ end of pool transcripts, was incubated with substrate oligonucleotide (Figure 8a). The rate for templated, but otherwise uncatalyzed, ligation in the context of an extended stem–loop structure (Figure 8b) was determined to be $7 \times 10^{-7}\ h^{-1}$. Previous comparisons indicated that this stem–loop fully adopts the closed conformation, so the ligation rate should be comparable to that observed for two oligonucleotides aligned on a separate template strand. Each of the ribozymes CM2g, CM3, and H4004 appeared to catalyze ligation at internal sites within the context of extended duplexes. Therefore, the rate enhancement achieved by

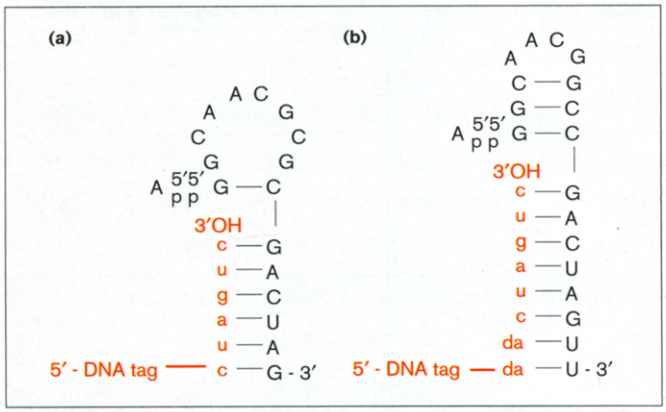
Table 5

Ribozyme activities with modified substrates.

	k_{obs} (h^{-1})			
	Ribose	2′-OMe	2′-deoxy	3′-deoxy
CM2g	3.8×10^{-1}	1.6×10^{-2}	9.1×10^{-4}	n.d.
CM3	4.6×10^{-2}	3.7×10^{-3}	4.4×10^{-4}	n.d.
H4004	1.2×10^{-2}	1.5×10^{-3}	1.1×10^{-4}	n.d.

A very faint product band was seen in reactions of the CM2g and CM3 with the 3′-deoxy-substituted substrate. But in both cases this small amount of product formed rapidly and then leveled off with only 4×10^{-5} to 6×10^{-5} of the total ribozyme being ligated. We believe that this nominal product formation was a result of contamination of the 3′-deoxysubstrate oligonucleotide with minute amounts of the ribose or 2′-deoxysubstrate. This contamination presumably resulted from impurities in the 3′-deoxycytosine CPG used in the automated synthesis procedures. Any product formed in the reaction of H4004 with the 3′-deoxysubstrate was below the detection limit ($k_{obs} < 3 \times 10^{-7}\ h^{-1}$).

Figure 8



The uncatalyzed ligation of RNAs. The background rate of ligation of template RNAs was investigated using the sequences shown. Substrate III and one of the AppG-primed template RNAs shown were annealed and then incubated in standard ligation buffer conditions at 25°C. **(a)** Duplex analogous to the stem–loop originally designed for the pool (Figure 1a). **(b)** Extended duplex configuration.

CM2g was ~5×10^5 relative to the templated, but uncatalyzed, background rate. CM3 and H4004 showed rate enhancements > 10^4-fold above the uncatalyzed rate.

Discussion

We have isolated ribozymes that catalyze the template-directed ligation of AMP-activated RNAs in a reaction analogous to the final step in the mechanism of the protein enzymes DNA ligase and RNA ligase. In this study, a total of nine different ribozymes have been uncovered (eight ribozymes after four rounds of selection and one ribozyme after ten rounds); in an earlier study, a much larger collection of ribozyme ligases that used pyrophosphate-activated substrates was isolated from a pool of similar complexity [1]. This difference raises the question of whether it is easier to select for ligases that use pyrophosphate activation. The background rate for the templated ligation of oligonucleotides activated with pyrophosphate is twofold to threefold higher than that of AMP-activated substrates under similar conditions [1]. Because a smaller rate enhancement would be required to achieve the same ligation rate for pyrophosphate-activated RNAs as for AMP-activated RNAs, ligases that use the former class of substrates might be more abundant in sequence space than those that use the latter class. The present selection was internally controlled for comparison of isolation of different classes of ligases, because up to about a third of the pool RNAs could have been activated with pyrophosphate, rather than AMP, during each round of selection. In this study, no ribozymes were isolated that used pyrophosphate activation to any significant extent, suggesting that ribozymes that use pyrophosphate-activated substrates are not much more common than ribozymes that use AMP-activated substrates. Also, ribozymes that could use either activating group would have held a selective advantage over the AMP-specific ribozymes; because no non-specific ribozymes were detected, these must be significantly less common in sequence space than specific ribozymes. Alternatively, it may be that our attempt to bias the pool in favor of ribozymes that use AMP-activated substrates (by incorporating the aptamer) was successful.

Another remarkable aspect of the ligase ribozymes isolated here is the unusual 'preference' for a ligation junction that is part of a continuous Watson–Crick duplex. DNA and RNA ligase enzymes previously isolated by *in vitro* selection have generally used ligation junctions that are not part of an extended duplex [2,5]. For instance, a very similar selection for 5′-triphosphate ligases from random RNAs yielded only RNAs which ignored or altered the 5′-stem–loop that was designed to bind substrate oligonucleotides as part of a long, contiguous stem [2]. Every ligase isolated from that selection either partially disrupted and opened up the designed stem–loop through alternative base pairing or used a completely different, internal, substrate-binding site that formed a Watson–Crick duplex on only one side of the ligation junction. Presumably, catalysis was favored by enhanced access to the ligation site. In the present selection, a substrate-binding stem–loop structure with a shorter stem and a larger loop than in the triphosphate ligase selection was placed at the 5′ end of the pool. Surprisingly, none of the isolated ligases appeared to use the designed stem–loop. All but one of the isolated ribozymes have sequences that would place the ligation junction within an internal, extended duplex structure with at least two base pairs (usually more) on both sides of the ligation junction. The one exception, clone H4003, was the only round 4 ribozyme with detectable pyrophosphate-activated ligation. The availability of the activating AMP for ribozyme–substrate interactions may have decreased the need for closer contacts to the ligation site which would otherwise have been incompatible with a contiguous duplex. If this speculation is correct, AMP-activation may be intrinsically better suited than pyrophosphate activation for a replication strategy that involves the sequential ligation of short oligonucleotide substrates.

The isolation of ligase ribozymes that use AMP-activated, templated RNA substrates strengthens the notion that RNA may be capable of the ligation or even the polymerization of similarly activated, templated substrates in a sequence-independent manner. The AMP 'handle' at the ligation junction may provide a route to strong enzyme–substrate interactions that would eliminate the sequence-specific contacts often observed with ribozymes. Evolutionary pressures for increased fidelity and for the ability to use more abundant shorter substrates may have favored the use of AppN (where N is a nucleoside) substrates,

which could ultimately be replaced by NTP substrates. The ribozymes isolated here are candidates for further *in vitro* evolution into more efficient ribozymes that catalyze partial RNA replication, as has been done with the previously isolated Class I ligase ribozyme [3], which can extend a primer by incorporation of NTPs.

An interesting but unresolved question is whether incorporation of the mutagenized ATP aptamer region within the pool was advantageous. The presence of frequent mutations in critical parts of the aptamer structure suggests that many of the isolated ribozymes may not use the aptamer region for AMP binding (Figure 4a,b). Presumably these ribozymes have adopted structures elsewhere in their sequence that specifically recognize the AMP leaving group. No new AMP-binding motif was identified in a comparison of different ribozymes, although a lack of knowledge of their secondary structures may have prevented the identification of any such consensus sequence. Whether or not those ribozymes with more conserved aptamer regions, such as H10140 and the derivative CM2g (Figure 6), are in fact using the aptamer to bind AMP is currently under study. Even weak binding of the AMP-activating group may be sufficient due to its high local concentration [16].

A previous selection for kinase ribozymes that used the ATP aptamer in a pool produced an estimated 10,000 kinase ribozymes [17]. This high yield suggested that predisposing the pool towards AMP binding by the incorporation of the aptamer had aided in the isolation of ribozymes, although this remains uncertain in the absence of a control selection from a random pool. Because fewer ligase ribozymes were obtained in the selection presented here than in the closely related selection for 5′-triphosphate ligases from a completely random pool [1,14], it appears that little advantage was achieved by incorporating the aptamer in the starting pool in the selection for AMP-activated ligases. It is possible that binding of the AMP handle by the aptamer might block access to the 5′-5′-pyrophosphate, thereby blocking efficient catalysis of ligation. Alternatively, the 15% mutagenesis of the aptamer region, which was designed to allow for some flexibility in the binding domain, may have simply resulted in the great majority of pool molecules having no functional aptamer domain so that the pool was functionally equivalent to a fully random pool.

Several results strongly suggest that factors other than catalytic efficiency determined the overall fitness of individuals in the pool population. H4004 only showed a slight rate advantage over several of the other ribozymes isolated at round 4, yet made up 80% of the entire pool. It seems unlikely that such a complete level of dominance at an early round could have resulted from such a minor rate advantage. H4004 was probably a superior template for PCR amplification, reverse transcription, or transcription. Furthermore, preliminary analysis suggested that some mutations found at high frequency in round 10 clones of the major ribozymes do not detectably enhance ligation rates. These mutations may have had more of an effect on replication efficiency than catalytic efficiency.

Significance

The successful isolation of ribozyme ligases that use AMP-activated (AMP: adenosine 5′-monophosphate) substrates expands the repertoire of biologically relevant reactions known to be efficiently catalyzed by RNA enzymes. These ribozymes use the same kind of substrate activation employed by the protein enzymes DNA ligase and RNA ligase and accelerate the attack by the 3′-hydroxyl of a templated RNA on the α-phosphate of an adenosine-5′-5′-pyrophosphate linkage of an adjacent templated RNA, resulting in 3′-5′-ligation and loss of AMP. The use of adenosine-5′-5′-pyrophosphate structures may be a relic of ribozyme catalysis in the 'RNA world'. The isolated ligases provide a promising starting point for the evolution of more efficient template-directed ligases or polymerases capable of sequence-independent recognition of substrate RNAs or monomers equipped with an AMP 'handle'.

Materials and methods

Oligonucleotide synthesis and purification

Synthetic DNAs and DNA/RNA substrate molecules were synthesized on PerSeptive Biosystems Expedite 8900 (Framingham, MA, USA) nucleic acid synthesizers according to the manufacturer's instructions. DNA/RNA substrate oligonucleotides were deprotected in ammonium hydroxide, 3:1 (v:v), and then thoroughly dried. The substrate oligonucleotides (each 1 μmol scale synthesis) were then incubated in 150 μl of 1 M tetrabutylammonium fluoride (Sigma, St Louis, MO, USA) at room temperature for 24 h to remove silyl groups. The tetrabutylammonium fluoride was removed by ethanol precipitation of the oligonucleotide. The 2′-methoxycytosine and 3′-deoxycytosine CPG support columns used for modified substrate synthesis were obtained from Glen Research (Sterling, VA, USA). All synthesized oligonucleotides were purified on denaturing polyacrylamide gels.

Pool construction

Two pool oligonucleotides were synthesized: 5′-GAACCTCTAG-GTCACCC*GAG*tgccagcaacgc-N_{50}-gcgttgggaagaaactgt**GGCACC**AC-GTCAGAAGGATC-3′ and 5′-GCGCGACTAGGATA**GGCACC**-N_{80}-*CCCGAG*ACCGTATAGCATGT-3′ (lower-case letters represent ATP aptamer bases randomized at a rate of 15%, upper-case letters represent constant regions, N represents a randomized site, bold letters indicate *BanI* sites, and italics indicate *AvaI* sites). Synthesis of random regions and analysis of pool complexity was done as previously reported [1]. The pool oligonucleotide containing the 80-nucleotide random region was PCR-amplified using the forward primer 5′-GCGC-GACTAGGATAGGCACC-3′ and the reverse primer 5′-ACATGCTAT-ACGGTCTCGGG-3′. The pool oligonucleotide containing the doped ATP aptamer was PCR-amplified using the forward primer 5′-GAAC-CTCTAGGTCACCCGAG-3′ and the reverse primer 5′-GATCCTTCT-GACGTGGTGCC-3′. Half of the N_{80}-containing pool PCR DNA was digested with *BanI* (NEB, Beverly, MA, USA) and the other half was digested with *AvaI* (NEB). The aptamer-containing pool PCR DNA was digested with both *AvaI* and *BanI*. A mixture of each of the three digested pool fragments was combined and ligated using T4 DNA Ligase (NEB). The ligated pool DNA was purified on 5% denaturing polyacrylamide gels and ~760 μg double-stranded DNA was recovered.

The ligated pool DNA was then PCR-amplified using the forward primer 5′-ATGTAATACGACTCACTATAGGCAACGCGCGACTAGGATAGGC-3′ (ah42.97) and the reverse primer 5′-GCCTGGTGATTTGCTGAACCTGATGTATCTAAGGACATGCTATACGGTCTC-3′.

Synthesis of 5′-5′-pyrophosphate dinucleotides

The phosphorimidazolide of adenosine (ImpA) and the phosphorimidazolide of guanosine (ImpG) were prepared on the gram scale as published [18]. The 5′-monophosphates were purchased from Sigma. Typically, AppG was synthesized by reacting ImpA (0.1 M) with 5′-GMP (0.05 M) in 0.1 M PIPES (piperidine-N,N′-bis[2-ethanesulfonic acid) pH 7.0 and 0.15–0.20 M $MgCl_2$ at 50°C for 4 h [13]. GppG was synthesized by reacting ImpG (50 mM) with 5′-GMP (25 mM) under the same conditions. The 5′-5′-pyrophosphate dinucleotides were analyzed and purified by high performance liquid chromatography (HPLC; Beckman Instruments (Columbia, MD, USA), Dionex NucleoPac column, 10% CH_3CN, gradient: 10–400 mM ammonium acetate, pH 5.5, in 30 min). The identity of the desired dinucleotides was confirmed enzymatically. The AppG dinucleotide, for instance, was cleaved by nucleotide pyrophosphatase (Sigma) into phosphatase-sensitive products, which coeluted with 5′-AMP and 5′-GMP under anion exchange HPLC conditions. The 5′-5′-pyrophosphate dinucleotide was resistant to alkaline phosphatase (Boehringer-Mannheim, Indianapolis, IN, USA).

Transcription reactions

Pool transcription of round 0 pool (10 ml) contained the following: ~40,000 units T7 RNA polymerase; ~1200 units RNasin RNase inhibitor (Promega, Madison, WI, USA); pool template DNA (~1 μM); 5 mM DTT (dithiothreitol); 5 mM uracil triphosphate (UTP), ATP, and cytosine triphosphate (CTP); 2.3 mM GTP; 7 mM AppG; 26 mM $MgCl_2$; 0.01% triton x-100; 3 mM spermidine; 40 mM Tris pH 7.9. Transcription reactions for subsequent pools were performed in smaller volumes (25–100 μl). Round 0 pool transcription was at 37°C for 8 h. Subsequent pools were typically transcribed in overnight incubations. All transcribed RNAs were gel purified on denaturing polyacrylamide gels. 'GTP only' transcription reactions were done as above except that they contained 5 mM GTP and no 5′-5′-pyrophosphate dinucleotide. Transcriptions primed with ApppG (NEB) were identical to the AppG reactions except for the substitution of ApppG for AppG. GppG-primed transcriptions contained 4.6 mM GppG and 2.3 mM GTP. (Less GppG was used due to its inhibition of transcription.)

DNA ligase reactions

DNA ligase reactions were performed on internally ^{32}P-labeled AppG-primed pool (0.5 μM) transcripts in ATP-free buffer (50 mM Tris pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, and 50 μg/ml bovine serum albumin). 1.5 μM DNA splint oligonucleotide (5′-CCTAGTCGCGCGTTGCCGACTAGTTTTTTTTTTTTTT-3′, 5′-CCTAGTCGCGCGTTGCCGACTAGAAAAAAAAAAAAAA-3′, or 5′-CCTAGTCGCGCGTTGCCGACTAGTTTGCTTGAGATGC-3′) and 3 μM substrate oligonucleotide were annealed to the pool RNA in distilled H_2O prior to the addition of buffer or enzyme. 60 units of T4 DNA ligase (NEB) were added to the 20 μl reactions. Reactions were incubated for 15 min at 25°C.

In vitro *selection*

Pool RNAs were immobilized on streptavidin agarose (Pierce, Rockford, IL, USA) by first annealing the pool RNA to a biotinylated primer 5′-AA-(biotin dT)-GCCTGGTGATTTGCTGAACC-3′ (biotin-dT phosphoramidite purchased from Glen Research) in 600 mM KCl, 25 mM Tris pH 7.4, and 1 mM EDTA. The poured column was washed with 1 × ligation buffer: 60 mM $MgCl_2$; 30 mM Tris pH 7.4; 600 mM KCl. The column was then stopped and 1 × ligation buffer and substrate DNA/RNA oligonucleotide added (6–8 μM final concentration). DNA/RNA substrate sequences were as follows: 5′-$d(A)_{22}$-cuaguc-3′ (substrate I); 5′-$d(T)_{22}$-cuaguc-3′ (substrate II); 5′-d(AAGCATGTAAGCATCTCAAGCAAA)-cuaguc-3′ (substrate III); 5′-d(AAGCATGTAAGCATCTCAAGCAAC)-cuaguc-3′ (substrate IV). The entire incubation took place at 25°C. After the incubation was completed the pool RNA was eluted off the column with 15 mM NaOH and immediately neutralized.

The active RNA species were selected on either oligo (T) cellulose columns (Pharmacia, Piscataway, NJ, USA), teflon oligo-affinity support columns (Glen Research), poly d(T) magnetic beads (Dynal, Lake Success, NY, USA), or poly d(A) magnetic beads depending on which substrate oligonucleotide had been used in that round. The poly d(A) magnetic beads were constructed from streptavidin magnetic beads from Dynal and the biotinylated oligonucleotide 5′-$d(A)_{25}$-BioTEG-3′ (BioTEG CPG support column from Glen Research). The DNA oligonucleotide sequence of the teflon oligo-affinity column was 5′-GAGATGCTTACATGCA-3′. Selection column binding of active RNAs took place in 5 × SSC (saline-sodium citrate buffer) and 0.5% sodium dodecyl sulfate. Multiple column volume washings were done with 2 × SSC. Active RNAs were eluted with distilled H_2O (or 15 mM NaOH, if necessary).

Superscript II (Gibco BRL, Grand Island, NY, USA) was used in all rounds for reverse transcription. The reverse primer 5′-GCCTGGTGATTTGCTGAACCTGATGTATCTAAG-3′ was used for both reverse transcription and PCR. In selective PCR, 5′-$(A)_{22}$-CTAGTC-3′, 5′-$(T)_{22}$-CTAGTC-3′, or 5′-AAGCATGTAAGCATCTCAAGCA-3′ was used as the forward primer. Prior to round 4, 20 cycles of selective PCR were done in each round. Starting with round 4, selective PCR was performed until a double-stranded DNA product of the correct size could be observed on a 2% agarose gel with ethidium staining. The DNA primer ah42.97 was used in nested PCRs to regenerate the T7 promoter. In rounds 6–10 of selection B, the following clone-specific forward primer was used in nested PCRs to discourage the amplification of contaminants: 5′-ATGTAATACGACTCACTATAGGCAACGCGCGACTAGGATAGGCACCTCTAT-3′.

Mutagenic PCR conditions were created by increasing Mg^{2+} and polymerase concentrations, increasing deoxyCTP (dCTP) and deoxythymidine triphosphate (dTTP) concentrations, and adding Mn^{2+} [1,19]. Analysis of mutations in a nonessential 32 nucleotide region in round 10 H4004 family clones (found deleted in several of the sequences and, when deleted in others, had no effect on activity) confirmed that the average rate of mutagenesis was 6% per site after three rounds of 30 doublings under mutagenic PCR conditions. There was a substantial bias, however, among the 51 mutations examined. A·T to T·A mutations were significantly overrepresented (5.3% frequency) and G·C to T·A and G·C to C·G mutations were significantly underrepresented (0.4% and 0.14% frequency, respectively).

Ligation assay conditions

Ribozymes were assayed for initial ligation rate activity in 1 × ligation buffer with 6 μM 5′ ^{32}P-labeled substrate oligonucleotide at 25°C. Unless otherwise indicated, all assays were done using substrate oligonucleotide III. Ribozyme concentration was 1.5 μM. If the ribozyme was full-length, the biotinylated oligonucleotide used to immobilize the pool during the selection was preannealed to the ribozyme prior to addition of substrate. Reactions were typically analyzed and bands quantitated on 8% polyacrylamide denaturing gels using a PhosphorImager from Molecular Dynamics (Sunnyvale, CA, USA).

Uncatalyzed reaction rate assay

5′ ^{32}P-labeled substrate oligonucleotide III (0.5 μM) and 10 μM of one of the template RNAs shown in Figure 8 were annealed and then incubated in 1 × ligation buffer for 1–2 weeks at 25°C. The template RNAs were AppG-primed and because of their small size we were able to separate the transcripts beginning with AppG from those beginning with GTP on a denaturing polyacrylamide gel.

Cloning and sequencing

All cloning was done as directed in the pT7Blue T-Vector Kit (Novagen, Madison, WI, USA). All sequencing was performed using ABI 373 DNA sequencers. Sequences of round 4 isolates were

submitted to the databases. Genbank accession numbers for the isolates: H4004, AF000122; H4853, AF000123; H4823, AF000124; H4742, AF000125; H4009, AF000126; H4028, AF000127; H4003, AF000128; and H4736, AF000129. The sequences of CM2g and CM3 were also submitted to the databases (accession numbers AF000130 and AF000131, respectively).

Acknowledgements

We thank David Wilson, Jonathan Urbach, Ananya Das and Jonathan Davis for helpful comments on the manuscript. This work was supported by a grant from NASA and an NSF predoctoral fellowship (A.J.H.).

References

1. Bartel, D.P. & Szostak, J.W. (1993). Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411-1418.
2. Ekland, E.H. & Bartel, D.P. (1995). The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231-3238.
3. Ekland, E.H. & Bartel, D.P. (1996). RNA-catalyzed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373-376.
4. Green, R. & Szostak, J.W. (1992). Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science* **258**, 1910-1915.
5. Cuenoud, B. & Szostak, J.W. (1995). A DNA metalloenzyme with DNA ligase activity. *Nature* **375**, 611-614.
6. Hager, A.H., Pollard, J.D., Jr. & Szostak, J.W. (1996). Ribozymes: aiming at RNA replication and protein synthesis. *Chem. Biol.* **3**, 717-725.
7. Orgel, L.E. (1986). RNA catalysis and the origins of life. *J. Theor. Biol.* **123**, 127-149.
8. White, H.B. (1976). Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101-104.
9. White, H.B. (1982). Evolution of coenzymes and the origin of pyridine nucleotides. In *The Pyridine Nucleotide Coenzymes.* (Everse, J., Anderson, B. & You, K.-S., eds), pp. 1-17, Academic Press, New York.
10. Sassanfar, M. & Szostak, J.W. (1993). An RNA motif that binds ATP. *Nature* **364**, 550-553.
11. Dieckmann, T., Suzuki, E., Nakamura, G.K. & Feigon, J. (1996). Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA* **2**, 628-640.
12. Jiang, F., Kumar, R.A., Jones, R.A. & Patel, D.J. (1996). Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature* **382**, 183-186.
13. Chu, B.C.F. & Orgel, L.E. (1984). Preparation of ligation intermediates and related polynucleotide pyrophosphates. *Biochim. Biophys. Acta* **782**, 103-105.
14. Ekland, E.H., Szostak, J.W. & Bartel, D.P. (1995). Structurally complex and highly active ligases derived from random RNA sequences. *Science* **269**, 364-370.
15. Lorsch, J.R., Bartel, D.P. & Szostak, J.W. (1995). Reverse transcriptase reads through a 2′-5′ linkage and a 2′-thiophosphate in a template. *Nucleic Acids Res.* **23**, 2811-2814.
16. Lorsch, J.R. & Szostak, J.W. (1995). Kinetic and thermodynamic characterization of the reaction catalyzed by a polynucleotide kinase ribozyme. *Biochemistry* **34**, 15315-15327.
17. Lorsch, J.R. & Szostak, J.W. (1994). *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31-36.
18. Joyce, G.F., Inoue, T. & Orgel, L.E. (1984). Non-enzymatic template-directed synthesis on RNA random copolymers. *J. Mol. Biol.* **176**, 279-306.
19. Cadwell, R.C. & Joyce, G.F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* **2**, 28-33.