

# ***In vitro* selection of a 5'-purine ribonucleotide transferase ribozyme**

Taek Jin Kang<sup>1</sup> and Hiroaki Suga<sup>1,2,\*</sup>

<sup>1</sup>Research Center for Advanced Science and Technology, The University of Tokyo, 153-8904 Tokyo, Japan and

<sup>2</sup>Department of Chemistry and Biotechnology, Graduate School of Engineering, The University of Tokyo, 113-8656 Tokyo, Japan

Received March 13, 2007; Revised and Accepted May 16, 2007

## **ABSTRACT**

**Here we report *in vitro* selection of a novel ribozyme that catalyzes the 5'-nucleotidyl transfer reaction forming the 2'-5' phosphodiester bond. This ribozyme was retrieved as a sole sequence in the pool enriched for the 5'-triphosphate-dependent activities in incorporating ATP- $\gamma$ S. The originally selected ribozyme consisting of 109-nucleotide (nt) was miniaturized to 45-nt M4 ribozyme via a series of mutation studies, and based on this mini-ribozyme a *trans*-acting system was constructed. One of the most challenging tasks in our study was to determine the chemistry occurring at the 5'-ppp site. We utilized various analytical methods including MALDI-TOF analysis of the product generated by the *trans*-acting system and elucidated the chemistry to be 3' $\rightarrow$ 5' mononucleotide extension forming the 2'-5' phosphodiester bond. Interestingly, M4 ribozyme promiscuously accepts a variety of purine nucleotides bearing 5'-mono-, di- and triphosphates as substrates. This remarkable ability of M4 ribozyme would lead us to the development of a new tool for the 5'-modification of RNAs with unique chemical groups.**

## **INTRODUCTION**

The discovery of naturally occurring RNA enzymes (ribozymes) (1,2) has postulated so-called the RNA world where RNAs could have served both genetic and catalytic roles (3). This notion has motivated us to not only search more ribozymes from nature (*in vivo*) but also artificially generate ribozymes *in vitro* with a wide variety of catalytic functions. Representative examples include phosphoryl transfer (4,5), acylation (6), alkylation (7,8), Diels-Alder (9), aldol (10) and redox (11) reactions. Among them, *in vitro* selection of ribozymes that utilize nucleotide triphosphates (NTPs) has been of great interest

since NTPs are playing diverse but essential roles in biological systems. ATP, for example, is one of the building blocks of RNA and at the same time serves an energy or a phosphate source as well as in cofactors. Because of their potential biological significance, tremendous efforts have been made to evolve NTP-utilizing ribozymes from an RNA pool of random sequences. Indeed, it has been thus far reported that ribozymes catalyze three types of chemistry with nucleotide substrates.

The first example is ribozymes that use ATP as a phosphate source. Polynucleotide kinase ribozymes (4,12) are capable of catalyzing 5'- or 2'-phosphoryl transfer reaction. The second example is a 5'-capping ribozyme (13–15), which utilizes GDP (as well as GMP or GTP) to form a 5'-5' phosphoanhydride bond that is analogous to the 5'-cap structure of mRNA. The third example is those that can elongate their own or other oligonucleotide chain by NTPs (16,17). Earlier ribozymes of this kind were isolated as those capable of ligating an oligonucleotide to its own 5'-end. One of these ribozymes, called Class I ligase ribozyme, was further evolved using various ingenious strategies and successfully turned into an RNA polymerase ribozyme (16). Significantly, this ribozyme is able to add multiple nucleotides to the 3'-end of a substrate RNA according to the external template sequence *in trans*; thus, its 5' $\rightarrow$ 3' polymerization function is analogous to the naturally occurring counterparts.

Most of earlier demonstrations described above have dealt with chemistry where the selection strategy could be rather easily devised; i.e. a selection method allowed ones to fish out desired ribozymes with only one possible function or with a dominant function competing with other potential reactions to occur. However, this does not mean that the ribozyme chemistry using NTPs or other nucleotide phosphates is limited to these reactions. For example, phosphoryl transfer to the 3'-hydroxyl or 5'-triphosphate (5'-ppp), and the 5'-nucleotidyl transfer (3' $\rightarrow$ 5' extension) could be catalyzed by RNA using ATP (or other NTPs) as a substrate. In principle, RNA sequences that catalyze one of these reactions can be

\*To whom correspondence should be addressed. Tel: +81-3-5452-5495; Fax: +81-3-5452-5495; Email: hsuga@rcast.u-tokyo.ac.jp.

selected by a conventional strategy that yielded kinase ribozymes using ATP- $\gamma$ S as a substrate (4,5,12,18); yet, no ribozyme with such functions has been reported. Therefore, a challenge is how ribozymes that catalyze unreported reactions are isolated and identified from a random RNA pool competing with those catalyzing other reactions reported previously. Accordingly, the selection strategy should be the most critical factor toward the success.

To isolate ribozymes capable of catalyzing the above reactions, we developed a strategy involving three layers of selection constraints to enrich 5'-ppp-dependent catalytic species in a random RNA pool. Executing this selection indeed yielded a novel RNA sequence that catalyzes one of the above reactions: 3'→5' nucleotide extension. We report here selection and characterization of this ribozyme that shows intriguing promiscuous recognition to purine nucleotides.

## MATERIALS AND METHODS

### Selection

The pool DNA (5'-GGATCGTCAT TGCATTGAGAN70-GGTGGTATCC CCAAGGGGTA-3'; constant region for the PCR and random sequences are shown) was amplified using the 5' primer containing the T7 promoter (5'-GGTAACACGC ATATGTAATACGACTCACTA TAGGGATCGT CAGTGCATTGAGA-3'; T7 promoter sequences are underlined) and the 3' primer (5'-TACCCCTTGG GGATACCACC-3'). Four copies of each sequence in the initial pool DNA (complexity of about  $1.6 \times 10^{15}$ ) were transcribed and purified on 10% denaturing PAGE. From the second round of selection onward, the pool RNA was body-labeled with [ $\alpha$ - $^{32}$ P]GTP (ICN) during the transcription. Conditions for the transcription are described below.

Reactions were carried out in 50 mM EPPS buffer (pH 7.5), 100 mM KCl, 100 mM MgCl<sub>2</sub> and 10 mM ATP- $\gamma$ S. The RNA pool (4  $\mu$ M for the first round and 1  $\mu$ M thereafter) was placed in EPPS and KCl and was denatured at 95°C and then brought to room temperature. MgCl<sub>2</sub> was added and let RNAs refold for 5 min. Reactions were initiated by adding ATP- $\gamma$ S (Sigma) and let proceed for 24 h at room temperature. Reaction volumes were 4.5 ml for the first round and 45  $\mu$ l thereafter. RNAs were precipitated from the reaction once by isopropanol and once again by ethanol.

From the first to ninth round of selection, thiophosphate-containing RNAs were purified using thiopropyl-activated resin (Sigma); RNAs from the reaction were dissolved in binding buffer (25 mM EPPS, 1 mM EDTA, pH 7.5) and were allowed to react with the resin (1 ml for the first round of selection and 50  $\mu$ l thereafter) for 1 h at room temperature with a gentle agitation. Thiophosphate-containing RNAs were covalently attached to the resin through the disulfide exchange reaction with 2-pyridyl disulfide part of the resin. The resulting resin was extensively washed with 20 volumes of washing buffer A (25 mM EPPS, 1 M NaCl and

5 mM EDTA, pH 7.5) and 40 volumes of 4 M urea. After washing the resin further with 10 volumes of water, bound RNAs were eluted using 2 resin volumes of elution buffer (1%  $\beta$ -mercaptoethanol, 12.5 mM EPPS, 500 mM NaCl and 2.5 mM EDTA, pH 7.5).

From the 10th round of selection, precipitated RNAs from the reaction were dissolved in biotin buffer [4 mM PEO-iodoacetyl biotin (Pierce), 50 mM Tris, pH 8.4]. Biotinylation of thiophosphate-containing RNAs preceded 3 h at room temperature in the dark with about 70% of yield and was followed by ethanol precipitation (twice). Resulting RNAs from the reaction were dissolved in washing buffer B (50 mM EPPS, 500 mM KCl, 5 mM EDTA, pH 7.5) and were allowed to react with 50  $\mu$ l of SAv-agarose resin (Pierce) for 1 h at room temperature with a gentle agitation. The resin was extensively washed successively with 20 volume of washing buffer B, 40 volumes of 4 M urea and 10 volumes of water. Bound RNAs were eluted using 2 resin volumes of biotin buffer (25 mM EPPS, 10 mM biotin, pH 7.0). Eluted RNAs were ethanol precipitated and reverse transcribed. cDNAs were amplified by PCR and transcribed for the next round of selection.

### Ribozyme preparation

5'-ppp RNAs were prepared by transcribing PCR template with 2 mM each of ATP, CTP, UTP and GTP. Total 5–10  $\mu$ Ci of [ $\gamma$ - $^{32}$ P]GTP or [ $\alpha$ - $^{32}$ P]GTP was added to the transcription reaction to prepare end-labeled or body-labeled RNAs, respectively. 5'-ppp RNA was treated with alkaline phosphatase (Promega) to prepare 5'-hydroxyl RNA (5'-OH RNA) according to the manufacturer's manual. The substrate strand for the *trans* reaction was prepared by the action of 10–23 DNA enzyme [3'-CCCTAGCAAGCAACATCGATCGGACCCACGCT-5', priming sequences to the RNA substrate are underlined, (19)] that was raised to the RNA transcript 5'-pppGGGAUCGUA↓UGGGUGCGAAAAAG-3' (priming sequences to the DNA enzyme are underlined and the site of cleavage is marked with an arrow). All RNAs were purified on appropriate PAGE. Usually the aimed transcript was heavily contaminated with *N* + 1 transcript (up to 50–60% in the transcription of M4) and less significantly with *N* + 2 or *N* – 1 transcripts. Since 3'-heterogeneity did not affect the activity of the ribozyme these transcripts were not extensively purified unless otherwise noted.

### General analytical procedures for ribozyme reactions

Reaction conditions were optimized to 100 mM EPPS (or 100 mM HEPES), 72 mM MgCl<sub>2</sub>, 8 mM MnCl<sub>2</sub>, 100 mM KCl, 10 mM ATP- $\gamma$ S (for other substrates, see figure legends), pH 7.5 and 1  $\mu$ M RNA (20  $\mu$ M RNA were used for the *trans* reaction). All reactions were carried out at room temperature for 24 h. Thiophosphate-containing RNAs were separated from the rest using either the biphasic [(*N*-acryloylamino)phenyl]mercury polyacrylamide gel electrophoresis (APM-PAGE) (20) or the SAv gel shift assay with post-biotinylation with

PEO-iodoacetyl biotin. Radiolabeled RNAs on the gel were visualized in a phosphoimager (Bio-Rad). Alternatively, cold RNAs were stained with SYBR Green II to be detected in the fluorescence scanner (Kodak). Software Quantity One (Bio-Rad) was used to quantify the RNAs in the gel. However, it should be addressed that this thiophosphate-containing RNA quantification was 'relative to control reaction' not 'absolute quantification of the reaction', for two reasons; (1) substrates ATP- $\gamma$ S and GTP- $\gamma$ S are contaminated with significant amount of ADP or GDP respectively (probably from the hydrolysis), (2) thiophosphorylated RNAs are not stable in terms of their terminal thiophosphate-phosphate diesters (only 50–70% of GTP- $\gamma$ S primed transcript that was purified from APM-PAGE survived after 24 h in the reaction buffer, data not shown).

### Susceptibility of the product RNA toward alkaline phosphatase

Reaction of cold M4 with [ $\alpha$ - $^{32}$ P]-ATP was quenched by precipitation with isopropanol. Residual [ $\alpha$ - $^{32}$ P]-ATP was removed from RNA by repeated precipitation with ethanol. Resulting RNA was subjected to the dephosphorylation reaction with 0.1 U/ $\mu$ l shrimp alkaline phosphatase (Roche) according to the manufacturer's manual. Samples were taken after 0, 10, 30 and 60 min of incubation at 37°C and were analyzed on 22% PAGE. [ $\alpha$ - $^{32}$ P]-ATP was treated with alkaline phosphatase to be used as a phosphate control. To make sure the phosphate release is the consequence of the enzyme action, the same set of reaction was carried out without alkaline phosphatase.

### Product analysis on DEAE-cellulose TLC

Cold M4 was allowed to react with [ $\gamma$ - $^{32}$ P]-ATP and was loaded on the Bio-Spin 30 column (Bio-Rad) to remove unreacted [ $\gamma$ - $^{32}$ P]-ATP. Resulting RNA was digested with RNase T2 (Sigma) in 50 mM ammonium acetate at 42°C for 2.5 h. 5'-end-radiolabeled M4 was digested with RNase T2 for being used as charge standards in the TLC. The digests were spotted on DEAE-cellulose TLC plate (J.T. Baker). TLC was run at step gradient of ammonium acetate (from 0.2 to 0.25 M) in 9 M urea and 1 mM EDTA, after a brief pre-run in water.

### Product analysis on MALDI-TOF

The thiophosphate-containing RNA was purified from the *trans* reaction of M4 on the APM-PAGE. Either this product RNA or starting substrate strand RNA in 0.5  $\mu$ l water was mixed with 0.5  $\mu$ l of 0.5 M diammonium hydrogen citrate and 1  $\mu$ l saturated solution of 3-hydroxypicolinic acid (TCI) in acetonitrile:water:ethanol (50:45:5, v/v). Samples were spotted on the target before mass values were taken by MALDI-TOF (Bruker). Synthetic RNA standards (6 and 14 nt long) were used to calibrate the mass externally.

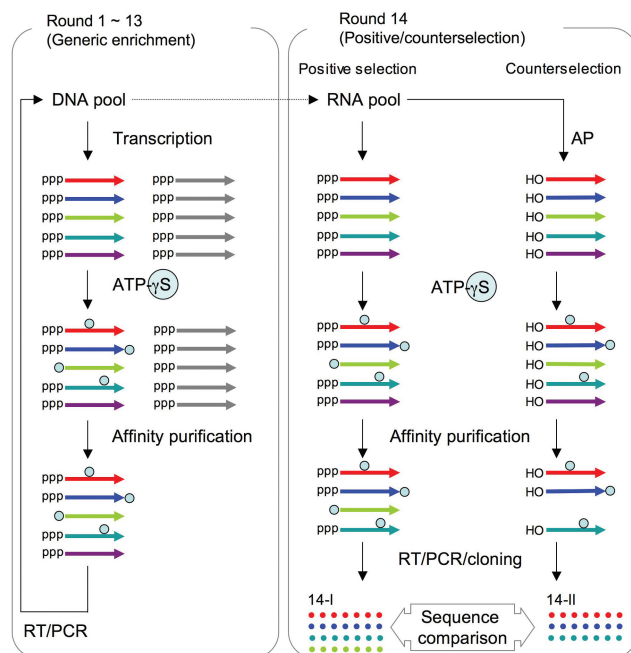
## RESULTS

### Selection of a 5'-triphosphate-dependent ribozyme

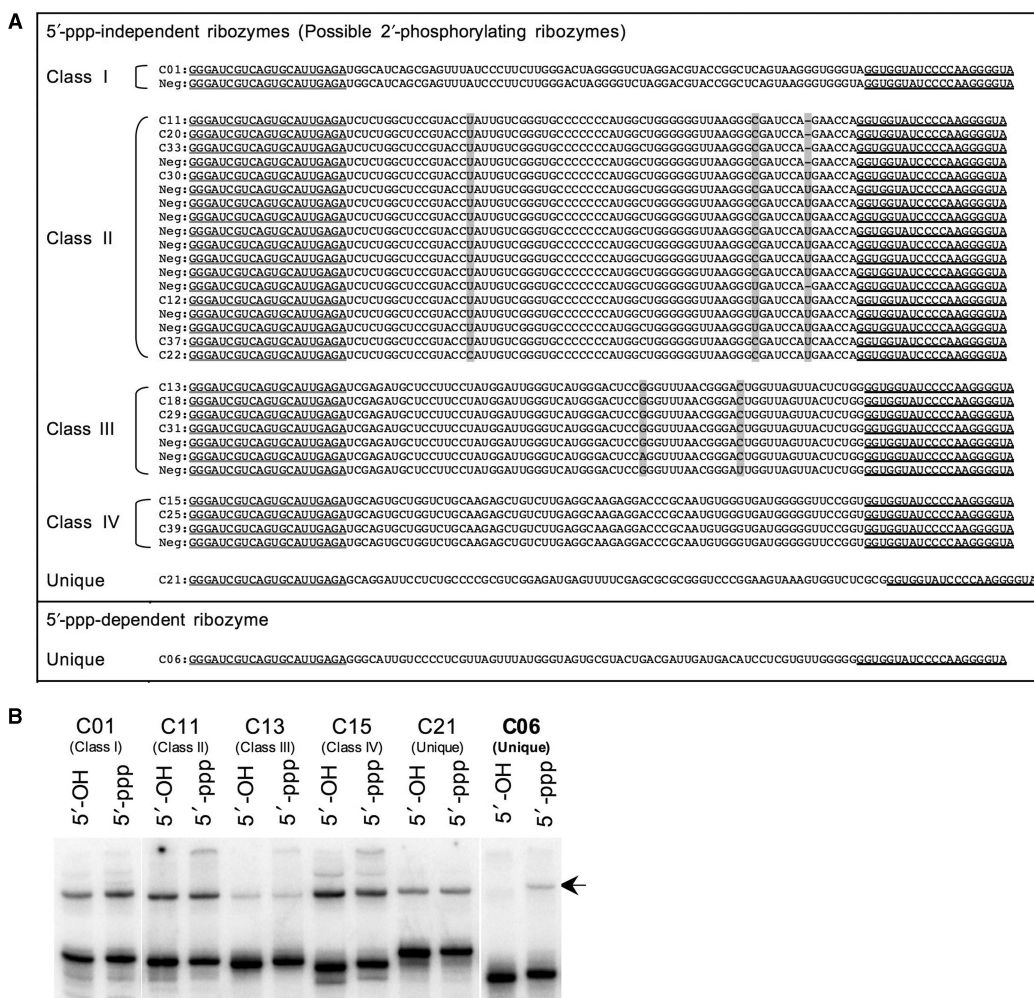
In order to enrich unreported catalytic species from the RNA pool, the selection strategy was built upon three layers of constraints below (Figure 1). The first constraint was to use ATP- $\gamma$ S as a substrate. This reagent had been used in the selection of kinase ribozymes where active RNAs were tagged with a thiophosphate transferred from ATP- $\gamma$ S to its own 5'-OH or internal 2'-OH (4,5). In theory, this reagent should tag any active species that transfer thiophosphate to any sites including the 3'-OH and 5'-ppp (if 5'-ppp-RNA were used) groups, although such ribozymes were not reported previously.

The second constraint was to use a pool of 5'-ppp-RNA as a substrate for self-modification. The triphosphate at the 5'-end of RNA can offer a unique reaction center not only as an electrophile, but also as a nucleophile. At the same time, this should eliminate the appearance of a previously reported catalyst, 5'-OH kinase (4), in the active populations.

The third constraint was aimed at distinguishing 5'-ppp-independent species from 5'-ppp-dependent ones in the enriched pool. The positive selection was performed the same as in earlier rounds (positive selection, Figure 1).



**Figure 1.** Selection of 5'-ppp-dependent *cis*-thiophosphorylating ribozymes from a random pool. 5'-ppp random RNAs were subjected to the reaction with ATP- $\gamma$ S and the resulting thiophosphorylated RNAs were enriched using the following two methods. In round 1–9, active RNAs were isolated on thiopropyl-activated resin. In round 10–14, active RNAs were modified by biotin-iodoacetoamide and isolated on the streptavidin-agarose resin. It should be noted that the combination of these two procedures turned out to be effective for diminishing the appearance of non-specific aptamers that bind to resins. To knock down the population of 5'-ppp-dependent species in pool 14, the counterselection was performed (right); the 5'-ppp group was removed by alkaline phosphatase (AP) to expose the 5'-OH group in pool 14 RNA.



**Figure 2.** Identifying a 5'-triphosphate-dependent ribozyme. **(A)** Sequence alignment of active clones from both selection procedures. According to the sequence similarities, clones were categorized to four classes (classes I-IV) and two unique sequences. The clones found in pool 14-I were named with C followed by number, whereas those found in pool-II were simply categorized as Neg. Sequences in the constant regions are underlined in gray (5') and in black (3'). Sites of sequence variations in each class are shown in gray boxes. **(B)** Thiophosphoryl-transfer activity of representative clones. Body- $^{32}$ P-labeled RNAs bearing 5'-ppp or 5'-OH were subjected to the reaction with ATP- $\gamma$ S for 24 h at room temperature. Resulting RNAs were biotinylated using biotin-I and then analyzed on SAV-dependent mobility shift assay.

On the other hand, the counterselection involved removal of 5'-triphosphate on the enriched RNAs to expose the 5'-OH group, and then active species were selected (counterselction, Figure 1). We carried out these selections at the round 14 in parallel, giving pool 14-I and 14-II, respectively (Figure 1). Since pool 14-I would contain both 5'-ppp-dependent and independent species, whereas pool 14-II would not contain 5'-ppp-dependent species, the comparison of the sequences found in these pools should allow us to identify the 5'-ppp-dependent species existing in the pool 14-I.

In the sequence alignment of clones from pools 14-I and 14-II, we identified two unique sequences, C21 and C06, in the pool 14-I (Figure 2A). The remaining sequences were found multiple times in both pools and classified to class I-IV. These unique clones and representative sequences of each class were then verified for the 5'-ppp dependence (Figure 2B). The clones represented

from classes I to IV were active independent from the presence of 5'-ppp group as expected. The activity of C21 was also independent from the presence of 5'-ppp group. Moreover, it was confirmed that periodate oxidation of the 3'-terminus of these clones did not diminish the activity, suggesting that all these clones likely self-thiophosphorylated onto the internal 2'-OH groups (data not shown). However, removal of 5'-ppp on C06 wiped out the activity, indicating that the 5'-ppp group is essential for activity (Figure 2B). Thus, C06 was chosen for further studies.

At the point when we identified C06, we were still uncertain about the chemistry catalyzed by C06. Since it was very plausible from the selection that 5'-triphosphate would be the site of reaction, possible reactions were (mono, di or tri)-phosphoryl transfer, or nucleotidyl transfer through a phosphoanhydride linkage (capping) or a phosphodiester linkage (extension). To facilitate

defining the product generated at the 5'-end, we first conducted mutation studies aimed at determining the secondary structure of C06 by minimizing the molecule to its essential regions.

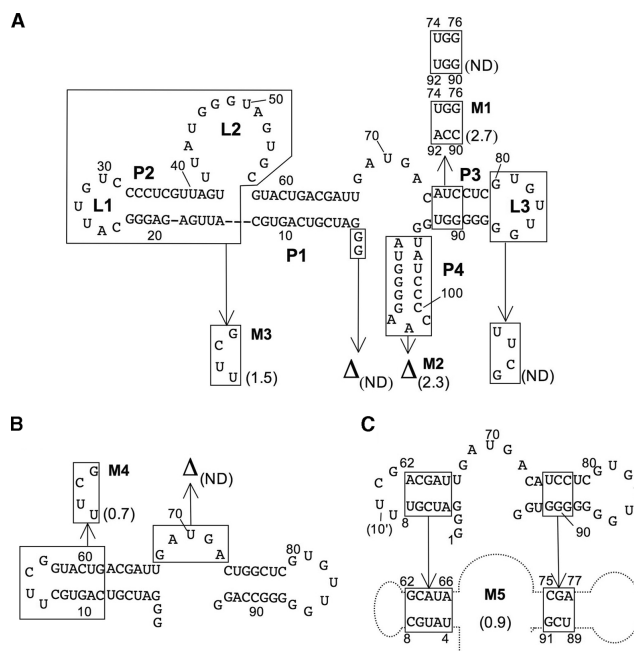
### Secondary structure of C06 and its miniaturization

A potential secondary structure of C06 was predicted by RNAstructure v4.2 (21) to have four stems (termed P1–P4), three loops (L1–L3) and one junction loop as shown in Figure 3A. Based on this structure, a series of mutations/deletions/substitutions were performed on P2, P3 and P4 domains (Figure 3A). Disruption of the P3 stem by making mispairings between A74–C76 and G90–U92 was detrimental to the catalytic activity, while its compensatory pairing gave a mutant M1 with 2.7-fold higher activity than C06. Therefore, the scaffold of M1 was used for further mutation studies toward miniaturization. Deletion of P4 domain in M1 gave only a minor decrease in activity (M2), ruling out a potential modification within this domain such as a hydroxy group at the 3'-end. On the other hand, substitution of L3 with UUCG tetraloop virtually eradicated the activity. Moreover, single and double deletions of the tandem guanines at the 5'-overhang also were detrimental to the activity. These results indicate that both motifs in L3 and the tandem Gs play critical roles in the formation of catalytic core of this ribozyme. Interestingly, the deletion/substitution of P2 domain with UUCG tetraloop reduced the activity, but yet this miniaturized mutant (M3) had an activity 1.5-fold higher than that of the parental ribozyme C06. We therefore used M3 for further miniaturization.

The P1 stem of M3 was further shortened to M4 consisting of a 6 bp stem, resulting in minor reduction in the activity; M4 exhibited a 0.7-fold activity of the wildtype C06 ribozyme (Figure 3B). Substitution of the junction loop G68–A72 between P1 and P3 stems with UUCG tetraloop was detrimental to the activity, indicating the indispensability of the sequence and length of this region (Figure 3B). On the other hand, double substitutions of P1 and P3 stems to unrelated stem sequences afforded M5 with slight increase in the activity presumably due to a better folding (Figure 3C). Taken together, the parental C06 ribozyme was miniaturized to 45-nt M4 ribozyme consisting of the 5'-tandem guanines and two stems that were linked via an indispensable 5-nt junction (Figure 3B). Thus, M4 ribozyme was used to determine chemistry occurring in its active site.

### Product analysis

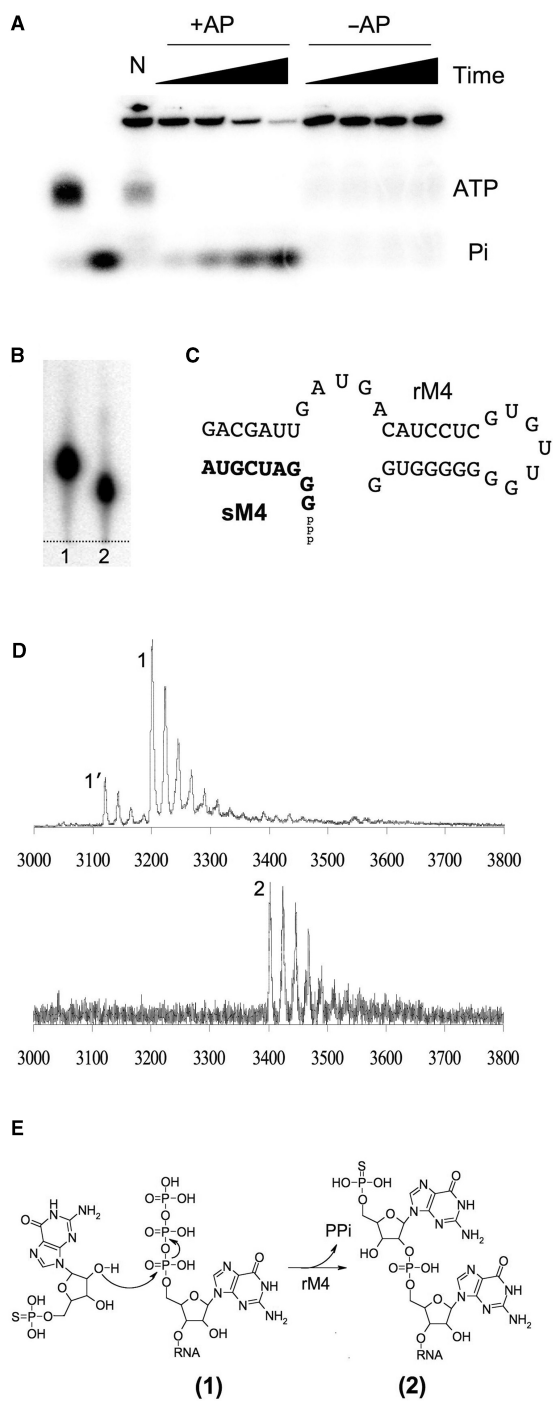
To characterize the M4-catalyzed reaction, we checked the 5'-terminal phosphate, the charge and the mass differences after the reaction. Prior to the series of experiments below to determine the product, we found that M4 ribozyme could be radiolabeled upon treatment with [ $\alpha$ - $^{32}$ P] or [ $\gamma$ - $^{32}$ P]-ATP (data not shown). Moreover, because the reaction took place in the 5'-ppp-dependent manner, the radiolabeling occurred most likely at the 5'-terminus. We therefore took an advantage of this labeling method to characterize the product.



**Figure 3.** Proposed secondary structures of the C06 ribozyme (A) and its derivatives (B and C). The secondary structure of C06 was predicted by mfold simulation and verified by mutation analyses. Mutations are shown in boxes, and only active mutants were named as M1–5. Relative activities compared to C06 ribozyme are shown in parentheses; P, stem; L, loop;  $\Delta$ , deletion mutation; ND, not detected. Nucleotide numberings of M4 (B and C) are kept the same as those in C06.

Reaction product of M4 with [ $\alpha$ - $^{32}$ P]-ATP resulted in radiolabeled RNA as describe above. The resulting RNA was subjected to alkaline phosphatase-mediated dephosphorylation reaction. If M4 catalyzes the capping reaction, the resulting structure that contains a phosphoanhydride bond should be inert toward this enzyme action. However, the transferred [ $^{32}$ P]-radiolabel on M4 was readily removed as a phosphate in an alkaline phosphatase-dependent manner, enabling us to rule out the capping reaction from the possible chemistry (Figure 4A).

For a charge analysis, we used the RNase T2-mediated digestion of the product RNA. Treatment of M4 with [ $\gamma$ - $^{32}$ P]-ATP followed by RNase T2-digestion of the product yielded a mixture of a [ $^{32}$ P]-labeled nucleotide and other non-labeled mononucleotides 3'-phosphate. We then analyzed this mixture on DEAE-cellulose TLC. Since the migration of nucleotides on DEAE-cellulose TLC is mainly determined by the number of negative charges (slower migration indicates more negative charges), this analysis allowed us to estimate the number of phosphates on the radiolabeled substance. As a sample for the comparison with the above product, we carried out [ $\gamma$ - $^{32}$ P]-GTP-primed transcription of M4 generating 5'-p\*pp-M4 (p\* indicates [ $^{32}$ P]-labeled phosphate) and its RNase T2-digestion yielding p\*ppGp (6 negative charges;  $-6$ ). We found that the reaction product migrated slower than the control p\*ppGp, indicating that it has more than 6 negative charges (Figure 4B). If an ATP transfer reaction at the 5'-terminus took place and



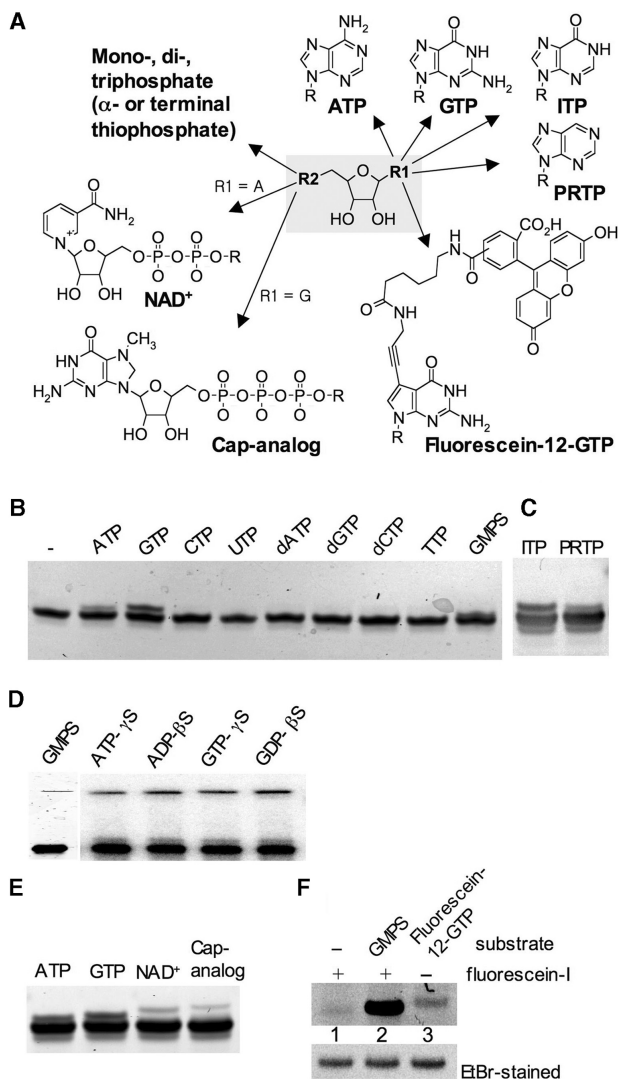
**Figure 4.** Product characterizations. **(A)** Dephosphorylation of the reaction product of M4 and  $[\alpha\text{-}^{32}\text{P}]\text{-ATP}$ . Positions where ATP and phosphate (Pi) migrate are shown; AP, alkaline phosphatase; N, AP-non-treated product RNA. **(B)** DEAE-cellulose TLC analysis of  $[\text{}^{32}\text{P}]$ -labeled nucleotides generated by RNase T2 digestion of the 5'-end-labeled M4 (sample 1), and the reaction product of M4 and  $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$  (sample 2). **(C)** Construct of *trans*-acting ribozyme (rM4) and substrate strand (sM4). **(D)** MALDI-TOF analysis of starting sM4 (upper) and the GMP- $\alpha$ S adduct purified by APM-PAGE (lower); 1, pppGGAUCGUA [ $M/z = 3200.923$  (calc. 3200)]; 2, SpG(2')p(5')GGAUCGUA [ $M/z = 3402.989$  (calc. 3401)]; Sp, thiophosphate. Series of sodium adducts are shown in both spectra as well as dephosphorylated sM4, ppGGAUCGUA (1',  $M/z = 3120.917$ ), in the spectrum for the starting material (upper). **(E)** The chemistry catalyzed by rM4.

the ATP-adduct contained a 3'-5' phosphate linkage, RNase T2-digestion should yield p\*ppAp (-6). The mobility of p\*ppAp should be the same or as close as p\*ppGp, and thus the above observation ruled out this possibility. Likewise, RNase T2-digested product originated from  $[\gamma\text{-}^{32}\text{P}]$ -transfer to an internal 2'-OH should generate N(2'-p\*)pNp (-5), and therefore it was also ruled out as additional evidence to the ppp-dependent activity. Taken together, we could narrow the potential chemistry down to the phosphoryl transfer (p\*pppGp; -7) or 5'-nucleotidyl transfer with 2'-5' phosphate linkage (p\*ppA(2')p(5')Gp; -7) onto the 5'-ppp site.

To distinguish these two products, we decided to conduct direct molecular mass measurement of the reaction product using MALDI-TOF. For this experiment, we used a nucleotide containing thiophosphate, e.g. ATP- $\gamma$ S, as a substrate, since the 5'-thiophosphate adduct could be readily purified by APM-PAGE. Our preliminary attempt using the *cis*-acting system showed increased molecular mass changes, and from the mass change it was likely that 5'-nucleotidyl transfer, rather than phosphoryl transfer, had taken place. However, the observed resolution was not high enough to withdraw a solid conclusion for mass changes (data not shown). To observe mass changes with a higher resolution, we performed reaction (1) with guanosine 5'- $\alpha$ -thiophosphate (GMP- $\alpha$ S) instead of ATP- $\gamma$ S, and (2) using a *trans*-acting M4 system in which the catalytic core and the 5'-acceptor substrate were separated (Figure 4C). Based on its secondary structure (Figure 3C), we synthesized a 9-nt substrate and a *trans*-acting ribozyme via dissection of L1 loop in M4 (Figure 4C, 9-nt sM4 and 34-nt rM4, respectively). These two pieces of RNA would form the complex via seven base pairs and react with GMP- $\alpha$ S to yield the 5'-GMP- $\alpha$ S adduct if the expected reaction took place. The MALDI-TOF analysis of reacted sM4 in comparison with unreacted sM4 showed the average mass change of approximately 202, which conformed to the expected value of 201 for the extension of sM4 by GMP- $\alpha$ S at the expense of pyrophosphate (Figure 4D). Recall the result of RNase T2-digestion experiment in which the M4-catalyzed self-incorporation of  $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$  increased negative charges in the product (Figure 4A). Taken all results together, we concluded that M4 ribozyme catalyzes 5'-nucleotidyl transfer, i.e. 3' $\rightarrow$ 5' mononucleotide extension, forming a 2'-5' linkage that is resistant against RNase T2 digestion (Figure 4E).

### Promiscuous specificity toward purine nucleotide substrates

During the course of the above product determination, we found that M4 ribozyme was able to incorporate GMP- $\alpha$ S to its own 5'-terminus. This observation extended our curiosity to search a full spectrum of substrate specificity. Structures of substrates tested in our studies are shown in Figure 5A. We first examined the substrate specificity of M4 toward canonical NTPs and dNTPs. Interestingly, both ATP and GTP could be a substrate while GTP was a superior substrate to ATP (Figure 5B). On the other hand, neither CTP nor UTP were viable substrates, suggesting that purine base is



**Figure 5.** Substrate specificity of M4. For the single nucleotide resolution, extensively purified M4 was used throughout the experiments shown here. (A) Structures of purine nucleotides used in this study. (B) Reaction with canonical nucleotides. Reactions were carried out using 4 mM of each substrate. (C) Reactions were carried out using 2 mM of ITP- $\alpha$ S or 0.76 mM of PRTP- $\alpha$ S. (D) Reaction with various purine thiophosphates. (E) Reactions were carried out using 4 mM of NAD<sup>+</sup> or the cap analog. (F) Reaction with fluorescein-12-GTP; lane 1, M4 was fluoresceinylated with iodoacetamido-fluorescein following the same protocol for the biotinylation (Materials and methods); lane 2, M4 was allowed to react with 4 mM GMP- $\alpha$ S, and the product was fluoresceinylated as lane 1; lane 3, M4 was allowed to react with 0.2 mM of fluorescein-12-GTP. The gel was scanned directly (upper) or after the staining with EtBr (lower) for the fluorescence. Even though there was a certain level of non-specific fluoresceinylation of (unreacted) 5'-ppp-M4 (lane 1, upper), the decreased mobility in the lane 3 clearly depicts that fluorescein-12-GTP was transferred to M4 (lane 3, upper). When the same gel was stained with EtBr, there was no difference in mobility throughout samples, which means the reaction yield with fluorescein-12-GTP was very low.

critical for the recognition of M4 (Figure 5B). Moreover, it was determined that the M4-catalyzed 5'-nucleotidyl transfer yielded the 2'-5' linkage. This predicted that none of dNTPs would be the substrate, and this was indeed the

case (Figure 5B). This observation added our additional confidence for the selective formation of the 2'-5' linkage.

The promiscuous activity toward GTP and ATP motivated us to investigate a greater variety of purine NTPs (Figure 5C). We found that inosine 5'-triphosphate (ITP) was a substrate as good as GTP. Similarly, purine-riboside 5'-triphosphate (PRTP) served as an active substrate, but less efficiently than GTP or ITP did; yet it was comparable to ATP (Figure 5B and C). The above results indicate that the functional groups of ATP 6-amino group, GTP 4-amino and 6-keto groups do not play vital roles in the recognition by M4 ribozyme.

Incorporation of GMP- $\alpha$ S to the 5'-terminus of M4 was readily detected by running an APM-PAGE (Figure 5D). Using this methodology, we investigated if M4 could incorporate ATP- $\gamma$ S, ADP- $\beta$ S, GTP- $\gamma$ S and GDP- $\beta$ S to its own 5'-terminus. All substrates turned out to be active, indicating that the activity of M4 did not rely on the number of 5'-phosphates in the substrate. Along similar lines, we also tested whether NAD<sup>+</sup> and m7GpppG could be incorporated into the 5'-end of M4. These nucleotides were also active substrates but less efficient than their canonical nucleotide counterparts, ATP and GTP, respectively (Figure 5E). This suggests that some steric restriction may exist in the substrate recognition site.

Lastly, we performed M4-catalyzed 5'-incorporation of fluorescein-12-GTP in which N7 on the guanine ring was modified with a linker that connected to the fluorescein molecule (Figure 5A). We were interested in this molecule since the incorporation of fluorescein-12-GTP would allow us to detect the product by a direct fluorescent gel scanning and at the same time the importance of purine N7 could be verified. Although the reaction did take place (Figure 5F), the efficiency was significantly low compared with the reaction with GMP- $\alpha$ S. It cannot be ruled out that the steric hindrance of the fluorescein linker interferes with the active site of M4, but this may suggest that the purine N7 plays some positive roles in catalysis.

In conclusion from the studies on substrate specificity, M4 ribozyme has unique promiscuous recognition toward purine nucleotides, including purine base compositions, mono-, di-, triphosphate, thiophosphate modification at the  $\alpha$ -/ $\beta$ -/ $\gamma$ -position and even modifications on  $\beta$ -/ $\gamma$ -phosphate with nucleotides. On the other hand, 2'-OH on the ribose is essential for chemistry to occur since M4 catalyzes the 5'-nucleotidyl transfer onto its own 5'-end to form the 2'-5' phosphodiester linkage.

## DISCUSSION

Here we have shown *in vitro* selection of a novel ribozyme C06 that catalyzes the 5'-nucleotidyl transfer reaction forming the 2'-5' phosphodiester bond (Figure 4E). The C06 sequence was found in pool 14-I as a minor population compared to possible internal kinases (Figure 2B). Finding of such a sequence was made possible by executing selection under the layers of selection criteria and the careful sequence comparison in two pools at the round 14 generated by the procedures

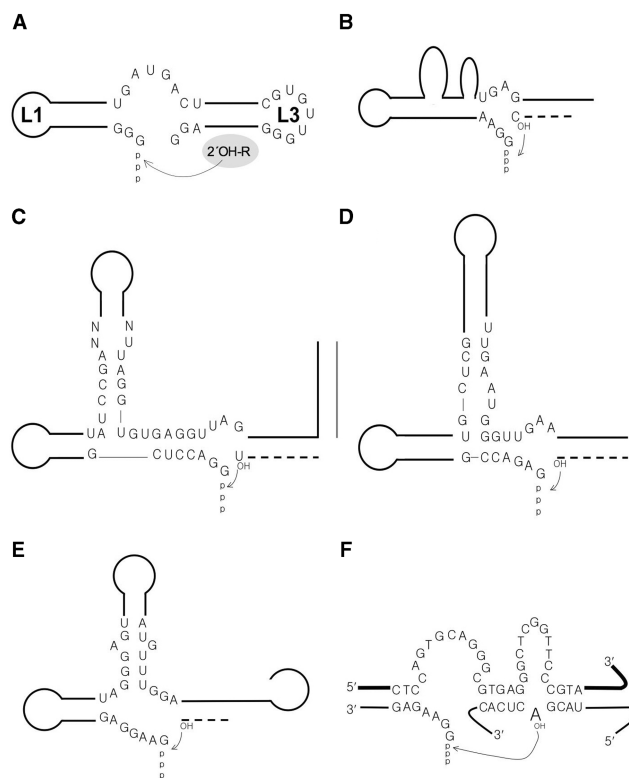
involving 5'-ppp or 5'-OH RNA for the selection. Especially, the latter counterselection strategy facilitated the confirmation of the desired activity without checking all clones for their activities. The 5'-nucleotidyl transfer reaction is considered as a 3'→5' mononucleotide extension, and thus no natural counterpart is known to exist in the present enzyme world. Nonetheless, an RNA molecule reported here is able to perform such a unique chemistry in *cis* as well as in *trans*.

Joyce *et al.* have reported a ribozyme capable of catalyzing both 5'- and 3'-nucleotidyl transfer reactions (17). This ribozyme, called E278-19, is a variant of Class I ligase ribozyme isolated by a continuous *in vitro* evolution procedure aimed at evolving 5'→3' mono- or di-nucleotide extension. Thus, the 3'→5' nucleotidyl transfer activity was coincidentally discovered in the collection of ligase or polymerase ribozymes, but it certainly kept the same signatures as the parental class I ligase ribozyme, where the incoming nucleotide formed a base pair with the template nucleotide and the 3'-5' phosphodiester linkage was generated.

In contrast, C06 and its miniaturized variant M4 promiscuously recognize the incoming purine nucleotides for the 3'→5' extension (Figure 5). Although it has not been ruled out if a nucleotide templating to the incoming purine nucleotide resides in the active site, the observed promiscuous activity toward various purine bases suggests that specific hydrogen bonding to purine bases is very unlikely. We rather think that the incoming substrate is guided by base-stacking interactions to the active site of ribozyme and somehow the ribozyme 5'- $\alpha$ -phosphate is projected to the 2'-OH of the incoming nucleotide for the nucleophilic attack.

Even though we cannot identify a direct counterpart enzyme in the contemporary protein world, terminal deoxyribonucleotide transferase (TdT) catalyzes a similar reaction. This enzyme can add nucleotide(s) to the 3'-end of DNA in a template-independent manner. Also TdT is well-known to have a preference to dG over other deoxyribonucleotides even though it basically has a promiscuous ability to accept nucleotide substrates (22). TdT apparently uses base-stacking interaction at least for its second nucleotidyl transfer step (23). Although TdT shares some of key features of M4, distinct differences exist regarding directionality and substrate specificity of the nucleotidyl transfer event. As TdT plays an indispensable role in enhancing the diversity of the immunoglobulin repertoire (23), we can imagine a beneficial role of terminal nucleotide transfer at either end of RNA in the RNA world with regard to a genetic diversity, where the recombination could have been the primary source to achieve an evolution or adaptation (3,24).

During the course of mutation studies, we found that indispensable motifs in M4 reside in the regions of the 5'-overhang, P2-P3 junction and L3 loop (Figures 3 and 6A). Since M4 catalyzes essentially the same chemistry as RNA ligase ribozymes except for the directionality of extension, it is of interest to compare the secondary structure of M4 with those of reported ligase ribozymes. Four representative ligases are selected for comparison; Bartel's class II ligase, Ellington's L1 ligase, Joyce's R3C



**Figure 6.** Proposed secondary structures of M4 ribozyme and ligases. (A) M4 from this study. (B) Bartel's class II ligase ribozyme (25). (C) Ellington's L1 ligase ribozyme (26). (D) and (E) Joyce's R3C and cytidine-free ligase ribozyme, respectively (27,28). (F) Siverman's 7S11 RNA ligase DNA enzyme (30). Bases likely involved in catalysis are shown. Other parts of ribozyme sequences are shown in solid line; DNA parts are shown in thicker solid line. The hydroxyl nucleophile of incoming purine (R) or oligonucleotide substrate (dashed lines) attacks the  $\alpha$ -phosphate on the ribozyme.

and cytidine-free ligases (Figure 6B–E) (25–28), all of which form a 2'-5' phosphodiester linkages like M4 ribozyme. These four ligases share the structural features of 5'-overhang and internal guide sequence for the incoming oligonucleotide. An obvious similarity found in all ribozymes including M4 is that the junction domain likely consisting of a part of catalytic core is dominated with A and G, occasionally containing U (note that G is dominated at the 5'-end because the consecutive Gs generally give higher transcription efficiency and therefore such RNA pools have been often used for selection). This suggests that such AG-rich motifs are suitable for composing the tertiary space where 5'- $\alpha$ -phosphate is positioned to the 2'-OH of the incoming nucleotide. On the other hand, a significant difference between M4 and these ribozymes is that M4 lacks the internal guide sequence for specific base pairing. Probably, the P2-P3 junction and L3 region would be located in close proximity, creating a 3D space for bringing the incoming purine nucleotides to the active site via base-stacking interactions.

A similar discussion could be extended to a DNA enzyme 7S11 that ligates RNA sequences through 2'-5' phosphodiester linkage. Unlike the ligase ribozymes discussed earlier, this DNA enzyme produces a branched



or lariat RNA (29,30) using an internal unpaired nucleotide as a branch-site. Though 7S11 is a DNA enzyme that reacts with a RNA substrate, it shares a similarity in structure to M4 where the electrophilic 5'-ppp-end of substrate is needed to be unpaired (Figure 6F). Moreover, it was shown that the branch-site nucleotide should be a purine nucleotide with 2'-hydroxyl even though adenosine was preferred to guanosine kinetically (30), which makes this reaction analogous to the M4-catalyzed purine nucleotidyl transfer.

The *trans*-acting activity of rM4 and the feasibility of designing the substrate strand guide sequence can lead us to the development of a new tool for the 5'-end modification using purine nucleotides. Particularly, the unnatural 2'-5' phosphodiester linkage at the 5'-end makes the oligonucleotide resistant toward 5'-exonucleases. Unfortunately, the current system is not yet sustainable for such applications due to the poor activity (using GMP- $\alpha$ S as a substrate,  $k_{cat}$  and  $K_M$  were  $0.070\text{ h}^{-1}$  and  $34\text{ mM}$ , respectively). However, re-evolution of M4 ribozyme with appendix domains would potentially produce M4 variants with higher activities, so that a new tool for the 5'-nucleotide modification of RNA can be devised.

## ACKNOWLEDGEMENTS

This work was supported by grants from Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research (S) (16101007) and from the US National Institutes of Health (GM59159). Funding to pay the Open Access publication charges for this article was provided by JSPS.

*Conflict of interest statement.* None declared.

## REFERENCES

- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA - auto-excision and auto-cyclization of the ribosomal-RNA intervening sequence of tetrahymena. *Cell*, **31**, 147–157.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.
- Gilbert, W. (1986) Origin of Life - the RNA World. *Nature*, **319**, 618.
- Lorsch, J.R. and Szostak, J.W. (1994) In-vitro evolution of new ribozymes with polynucleotide kinase-activity. *Nature*, **371**, 31–36.
- Saran, D., Nickens, D.G. and Burke, D.H. (2005) A *trans* acting ribozyme that phosphorylates exogenous RNA. *Biochemistry*, **44**, 15007–15016.
- Saito, H. and Suga, H. (2001) A ribozyme exclusively aminoacylates the 3'-hydroxyl group of the tRNA terminal adenosine. *J. Am. Chem. Soc.*, **123**, 7178–7179.
- Wilson, C. and Szostak, J.W. (1995) In-vitro evolution of a self-alkylating ribozyme. *Nature*, **374**, 777–782.
- Wecker, M., Smith, D. and Gold, L. (1996) In vitro selection of a novel catalytic RNA: characterization of a sulfur alkylation reaction and interaction with a small peptide. *RNA*, **2**, 982–994.
- Tarasow, T.M., Tarasow, S.L. and Eaton, B.E. (1997) RNA-catalysed carbon-carbon bond formation. *Nature*, **389**, 54–57.
- Fusz, S., Eisenfuhr, A., Srivatsan, S.G., Heckel, A. and Famulok, M. (2005) A ribozyme for the aldol reaction. *Chem. Rev.*, **12**, 941–950.
- Tsukiji, S., Pattnaik, S.B. and Suga, H. (2003) An alcohol dehydrogenase ribozyme. *Nat. Struct. Biol.*, **10**, 713–717.
- Lorsch, J.R. and Szostak, J.W. (1995) Kinetic and thermodynamic characterization of the reaction catalyzed by a polynucleotide kinase ribozyme. *Biochemistry*, **34**, 15315–15327.
- Huang, F.Q. and Yarus, M. (1997) 5'-RNA self-capping from guanosine diphosphate. *Biochemistry*, **36**, 6557–6563.
- Zaher, H.S. and Unrau, P.J. (2006) A general RNA-capping ribozyme retains stereochemistry during cap exchange. *J. Am. Chem. Soc.*, **128**, 13894–13900.
- Zaher, H.S., Watkins, R.A. and Unrau, P.J. (2006) Two independently selected capping ribozymes share similar substrate requirements. *RNA*, **12**, 1949–1958.
- Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. and Bartel, D.P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science*, **292**, 1319–1325.
- McGinness, K.E., Wright, M.C. and Joyce, G.F. (2002) Continuous in vitro evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem. Biol.*, **9**, 585–596.
- Saran, D., Held, D.M. and Burke, D.H. (2006) Multiple-turnover thio-ATP hydrolase and phospho-enzyme intermediate formation activities catalyzed by an RNA enzyme. *Nucleic Acids Res.*, **34**, 3201–3208.
- Santoro, S.W. and Joyce, G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA*, **94**, 4262–4266.
- Igloi, G.L. (1988) Interaction of transfer-RNAs and of phosphorothioate-substituted nucleic-acids with an organomercurial - probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry*, **27**, 3842–3849.
- Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA*, **101**, 7287–7292.
- Basu, M., Hegde, M.V. and Modak, M.J. (1983) Synthesis of compositionally unique DNA by terminal deoxynucleotidyl transferase. *Biochem. Biophys. Res. Commun.*, **111**, 1105–1112.
- Gauss, G.H. and Lieber, M.R. (1996) Mechanistic constraints on diversity in human V(D)J recombination. *Mol. Cell. Biol.*, **16**, 258–269.
- Riley, C.A. and Lehman, N. (2003) Generalized RNA-directed recombination of RNA. *Chem. Biol.*, **10**, 1233–1243.
- Ekland, E.H., Szostak, J.W. and Bartel, D.P. (1995) Structurally complex and highly-active RNA ligases derived from random RNA sequences. *Science*, **269**, 364–370.
- Robertson, M.P., Hesselberth, J.R. and Ellington, A.D. (2001) Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings. *RNA*, **7**, 513–523.
- Rogers, J. and Joyce, G.F. (2001) The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA*, **7**, 395–404.
- Rogers, J. and Joyce, G.F. (1999) A ribozyme that lacks cytidine. *Nature*, **402**, 323–325.
- Wang, Y.M. and Silverman, S.K. (2003) Deoxyribozymes that synthesize branched and lariat RNA. *J. Am. Chem. Soc.*, **125**, 6880–6881.
- Coppins, R.L. and Silverman, S.K. (2004) A DNA enzyme that mimics the first step of RNA splicing. *Nat. Struct. Mol. Biol.*, **11**, 270–274.