A combinatorial method to isolate short ribozymes from complex ribozyme libraries

Joshua T. Arriola and Ulrich F. Müller ⓘ*

Department of Chemistry & Biochemistry, University of California San Diego, La Jolla, California 92093, USA

ABSTRACT

In vitro selections are the only known methods to generate catalytic RNAs (ribozymes) that do not exist in nature. Such new ribozymes are used as biochemical tools, or to address questions on early stages of life. In both cases, it is helpful to identify the shortest possible ribozymes since they are easier to deploy as a tool, and because they are more likely to have emerged in a prebiotic environment. One of our previous selection experiments led to a library containing hundreds of different ribozyme clusters that catalyze the triphosphorylation of their 5′-terminus. This selection showed that RNA systems can use the prebiotically plausible molecule cyclic trimetaphosphate as an energy source. From this selected ribozyme library, the shortest ribozyme that was previously identified had a length of 67 nucleotides. Here we describe a combinatorial method to identify short ribozymes from libraries containing many ribozymes. Using this protocol on the library of triphosphorylation ribozymes, we identified a 17-nucleotide sequence motif embedded in a 44-nucleotide pseudoknot structure. The described combinatorial approach can be used to analyze libraries obtained by different in vitro selection experiments.

INTRODUCTION

In vitro selections were originally established to identify RNAs that bind to specific target molecules [1,2], and developed further to identify catalytic RNAs (ribozymes) from large, combinatorial libraries [3]. *In vitro* selections for catalytic RNAs have usually employed libraries with 10^{12} to 10^{15} different sequences, in which a portion of the library between 40 and 228 nucleotides in length is randomized [3–30]. To select for active sequences, the libraries of randomized RNA sequences are usually covalently coupled to one of the reaction partners of the desired reaction, while the second reaction partner is coupled to a functional group that serves as a handle. When incubating the modified library with the functionalized substrate, catalytically active sequences can covalently link themselves to the handle. The handle is then used to isolate the active sequences. Because the enrichment for active sequences in a single experiment is limited, multiple cycles of incubation with substrate, isolation of active molecules, reverse transcription, PCR amplification, and transcription are necessary to obtain libraries that are dominated by active sequences. The resulting, active library is then analyzed for individual, active sequences.

After active sequences have been enriched by selection steps, it is often desirable to identify ribozymes with specific properties. To enrich for the most active sequences, usually selection cycles with higher selection stringency are added (e.g. [3,12,13]). To identify ribozyme variants that use different metal ion cofactors, further rounds of in vitro selection or in vitro evolution can be employed in the presence of different metal cations, or metal cation concentrations [31]. To generate, and identify smaller variants of *individual* ribozymes a combinatorial method is available that can remove unnecessary fragments from the ribozyme [32]. However, we are not aware of a method to identify the smallest ribozymes from complex libraries with *many* ribozymes.

The desire to identify short ribozymes is partially motivated by attempts to explore how an RNA world could have emerged as an early stage in the origin of life. Short ribozymes are especially helpful because they are more likely to emerge from a prebiotic setting for two reasons: First, short RNA polymers are more abundant than long RNA polymers in model prebiotic reactions [33]. Second, short ribozymes require less information content, and are therefore found more frequently in a given set of randomized sequences. Previously selected ribozymes for prebiotically relevant reactions typically have sizes around 70 nucleotides or larger (e.g. [11,13,20,30,34,35], but see [36]). In contrast, model reactions for prebiotic RNA polymerization in the absence of enzymes or sophisticated ribozymes generates little material above 40 nucleotides, even under optimal conditions [33]. Because RNAs with lengths of 30–40 nucleotides are more prebiotically plausible, the involvement of catalytic RNAs in early stages of life would be more convincing if short ribozymes could be found to emerge from random sequences.

---

*To whom correspondence should be addressed. Tel: +1 858 534 6823; Email: ufmuller@ucsd.edu
Present address: Ulrich Müller, Department of Chemistry & Biochemistry, University of California San Diego, La Jolla, CA 92093, USA.

We previously performed an in vitro selection for ribozymes that triphosphorylate their own 5′-hydroxyl group with cyclic trimetaphosphate (Tmp) (30), a prebiotically plausible energy source (37,38). This selection identified >300 different ribozymes (39), each with a total length of 182 nucleotides. These results showed that catalytic RNA systems are able to use Tmp as energy source. Here we developed a combinatorial method to identify the shortest ribozyme sequences in complex libraries. We show how combinatorial truncation at the pool's 3′-terminus, size fractionation, and re-selection of each size fraction can identify the shortest ribozymes in this complex library. The shortest ribozyme had a truncated length of 44 nucleotides, and its 17-nucleotide conserved core was also identified in two additional sequences of the same selected ribozyme library by the re-analysis of high throughput sequencing data. The method may be useful for the isolation of short ribozyme motifs from other libraries containing hundreds of different ribozymes.

MATERIALS AND METHODS

Generation of truncated pools with a randomized primer

The truncated sub-pools were generated from template DNA using a randomized primer consisting of an N11 randomized region and a new 3′ constant region (5′-TAAG TCGTAGTTACATCANNNNNNNNNNN-3′). The template DNA was a pool consisting of hundreds of active triphosphorylating ribozymes flanked at the 5′ end by the T7 promoter sequence, a hammerhead ribozyme sequence, and a 5′ constant region and at the 3′ end by a 3′ constant region. The annealing and extension conditions were chosen to obtain a broad size distribution of the desired extension products. A mixture containing 30 pmol of template DNA, 5 pmol randomized primer in a volume of 0.2 ml, and a trace amount of radiolabeled randomized primer were denatured at 94°C for 2 min then immediately placed on ice. The Klenow fragment (New England Bioscience) was used to extend this primer according to manufacturer's instructions. The mixture was allowed to incubate overnight at room temperature. Formamide loading buffer containing 30 mM sodium/EDTA was added to the reaction mixture to quench the reaction. The mixture was heat denatured at 80°C for 2 min then separated on a denaturing 15% polyacrylamide gel. A 246 nt and 96 nt marker were used to define the region of interest. This region of interest was divided into 10 fragments and each fragment was excised. Each gel slice was eluted and ethanol precipitated. The resulting DNA sub-pools were PCR amplified using Taq DNA polymerase and transcribed using T7 RNA polymerase. The RNA sub-pools were gel purified, ethanol precipitated, and resuspended in water.

In vitro selection

One round of selection was performed on each purified RNA sub-pool according to previously published methods (30). 100 nM of purified RNA sub-pool was incubated with 50 mM Tris/HCl, 3.3 mM NaOH, 100 mM MgCl$_2$ and 50 mM of freshly dissolved trimetaphosphate. This mixture was allowed to incubate for 3 h at room temperature.

The reaction was quenched by ethanol precipitation. The RNA pellet was washed with chilled water and desalted using P30 Tris/HCl spin columns (Bio-Rad). Active ribozymes were ligated to a biotinylated oligomer (biotin-d(GAACTGAAGTGTATG)rU) using the R3C ligase ribozyme whose arms were designed to anneal to the 5′ constant region of the pool RNA and the biotinylated oligomer. A mixture containing 1.3 uM of desalted pool RNA, 1 uM ligase ribozyme, 1.2 uM biotinylated oligomer, 100 mM KCl, 100 mM Tris/HCl, and 60 pM triphosphorylated RNA was heated to 65°C for 2 min then cooled to 30°C at a rate of 0.1°C per second. The triphosphorylated RNA was used to generate a small amount of ligation product and thereby reduce the number of PCR cycles, and prevent artefacts during PCR amplification. After heat renaturation, an equal volume of a mix consisting of 40% (w/v) PEG8000, 4 mM Spermidine, and 50 mM MgCl$_2$ was added to the ligase reaction mixture. This mixture was incubated at 30°C for 3 h. After the incubation step, the reaction was quenched by adding final concentrations of 13.9 mM sodium/EDTA, 50 mM Tris/HCl, 50 mM KCl, 0.012% (w/v) Triton X-100, and 1.19 uM of a 61 nt oligomer that was complementary to the ligase ribozyme (5′GAACTGAAGTGTATGCTTCAACCCATTCAA ACTGTTCTTACGAACAATCGAGCAAGATGTT-3′). This mixture was heated at 50°C for 10 min. Streptavidin magnetic beads (Promega) were washed thrice with 20 mM HEPES/KOH pH7.2, 0.01% (w/v) Triton X-100 and 50 mM KCl. Biotinylated RNA was then captured by mixing the RNA with the magnetic beads and rotating end-over-end at room temperature for at least 30 min. A magnetic rack was used to focus the beads and the beads were washed twice with a solution containing 0.01% (w/v) Triton X-100 and 20 mM NaOH. Captured RNA was eluted from the beads by incubating the beads in a solution of 25 mM Tris/HCl pH 8.5, 1.56 mM EDTA and 96% formamide at 65°C for 3 min. The beads were removed by immediate centrifugation, and the supernatant was concentrated by ethanol precipitation. 1 ug of tRNA was used as precipitation carrier. Captured RNA was resuspended in 10 mM Tris/HCl, pH 8.3. The RNA was reverse transcribed using Superscript III (Invitrogen) and a reverse transcription primer (5′-TAAGTCGTAGTTACATCA-3′) complementary to the new 3′ constant region, which was implemented after the Klenow extension using the randomized primer. The products were then PCR amplified with 5′ and 3′ primers that could bind to the sequence of the biotinylated oligomer and the 3′ constant region of the pool respectively. A second PCR amplification was performed to add the T7 promoter sequence and the hammerhead ribozyme sequence to the DNA.

Triphosphorylation assays

Purified ribozyme (5 uM) was incubated with 50 mM Tris/HCl, 100 mM MgCl$_2$, and 50 mM of freshly dissolved trimetaphosphate. This mixture was incubated for 3 h at room temperature. The products were ligated to a radiolabeled oligomer (γ ^{32}P-d(GAACTGAAGTGTATG)rU) using an R3C ligase ribozyme whose arms were designed to anneal to the 5′ constant region of the pool RNA and the

oligomer. To do this, the above reaction mixture was diluted 1:10 in a solution that contained a final concentration of 0.5 uM ligase ribozyme, 0.5 uM biotinylated oligomer, 100 mM KCl, 100 mM Tris/HCl, and 15 mM sodium EDTA. This mixture was heat renatured at 65°C for 2 min then cooled to 30°C at a rate of 0.1°C per second. After heat renaturation, an equal volume of a mix consisting of 40% PEG8000 (w/v), 4 mM Spermidine, and 50 mM $MgCl_2$ was added to the ligase reaction mixture. This mixture was incubated at 30°C for 3 h. After the incubation step, the reaction was quenched by ethanol precipitation. The reaction products were resolved on a denaturing 10% polyacrylamide denaturing gel.

Generation of 3′ truncations

3′ truncations were generated by PCR using Taq DNA polymerase and a series of 3′ primers. The 3′ primers annealed to different portions of the ribozyme sequence of interest to generate a series of truncations with desired deletions from the 3′ end. DNA templates of each truncation were transcribed using T7 RNA polymerase, and purified by denaturing PAGE.

Secondary structure analysis using SHAPE

Selective 2′Hydroxyl Acylation Analyzed by Primer Extension (SHAPE) was performed on the 44-nucleotide long ribozyme isolated in this study, with 1-Methyl-7-nitro-2H-3,1-benzoxazine-2,4(1H)-dione (1M7) as the chemical probe ([40]). An adaptor region was added to the 3′ end of the ribozyme by PCR amplification. The sequence of this primer is (5′-GTGTGCTAGGATCACAATGATGTCTC TTTAATAAGA-3′) where the underlined portion is the adaptor region. 20 pmol of ribozyme were heat renatured in 10 ul at 80°C for 2 min, cooled to 50°C for 5 min then left at room temperature for 5 min. To this solution, HEPES/NaOH pH 8.0 and $MgCl_2$ were added to a final concentration of 50 mM each. This mixture was incubated at room temperature for 2 min. A 20 mM stock solution of 1M7 was prepared in DMSO. 1M7 was added to this solution such that its final concentration was 2 mM and the concentration of DMSO was 10% of the solution. A negative control was prepared with ribozyme, HEPES, $MgCl_2$ and 10% final DMSO. Both samples were incubated at room temperature for 3 min. The reaction was quenched by ethanol precipitation and resuspended in 10 ul of 5 mM Tris/HCl pH 8.0. The products were reverse transcribed using Superscript III reverse transcriptase (Invitrogen) according to manufacturer's instructions and trace amounts of a radiolabeled primer that annealed to the aforementioned adaptor region. The sequence of this primer is (5′- GTGTGCTAGGATCACAAT-3′). The RNA template in each sample was degraded by alkaline hydrolysis, by incubating for 5 min at 80°C in a solution containing 750 mM NaOH. The reaction was quenched by adding a two-fold stoichiometric excess of acetic acid to generate a NaOAc/HOAc buffer with a final concentration of 300 mM. The products were ethanol precipitated, resuspended in formamide loading buffer and resolved on a denaturing 20% polyacrylamide gel. Signals were quantified using

the 'rectangles' function in the software Quantity One (Bio-Rad) and background rectangles were subtracted.

To predict the secondary structure based on these SHAPE data the software vsFold5 was chosen because it is able to predict pseudoknots, and because it outperformed four other algorithms in predicting a secondary structure consistent with the SHAPE probing data. Specifically, iterative HFold ([41]) was unable to predict the pseudoknot, IPKnot ([42]) predicted a pseudoknotted structure lacking a part of the P2 helix and had an additional 2-base pair helix that were inconsistent with the SHAPE data, HotKnots ([43]) and CCJ ([44]) introduced two different 2-base pair helices that were not supported by the SHAPE data. None of the five algorithms predicted a structure that showed an extension of the P2 duplex between bases C16/C17 and G37/A38, which would have been consistent with the SHAPE data (Figure 5).

RESULTS

To identify the shortest ribozymes from our previously selected library of self-triphosphorylating ribozymes ([30]) with > 300 different ribozyme clusters ([39]), we followed a five-step procedure (figures 1 - 3). In the first step, 3′-truncated variants of the DNA pool were generated with the help of partially randomized primers and Klenow enzyme (Figure 1A). To do this, the DNA pool was annealed to a DNA primer that contained 11 nt of a randomized sequence at its 3′-terminus (red in Figure 1A). This randomized sequence allowed the primer's 3′-terminus to anneal at any position of the templating DNA library molecules. The primer was then extended by a Klenow Fragment enzyme that lacked exonuclease activity. A portion of the DNA primer was radiolabeled at its 5′-terminus to allow the detection of extension products of this primer.

In the second step, the primer extension products were separated into size fractions, using denaturing polyacrylamide gel electrophoresis (PAGE) (Figure 1B). Two radiolabeled markers indicated the sizes of long Klenow extension products that included the full 150-nucleotide long random region (total length 246 nt), and short Klenow extension products that omitted any portion of the randomized region (total length 96 nt). The extension products with sizes between these markers were separated into ten evenly sized segments, and the DNA was eluted from each segment (red brackets in Figure 1B). As a negative control, a reaction that omitted any DNA pool was performed (Figure 1B, lane labeled 'Rxn-T'). Extension products <96 nt in length were detected but little to no products with >96 nt in length. The short products can be explained by a scenario where the randomized portions of two primers were complementary to each other or the randomized portion of one primer was complementary to the constant region of another. This did not generate problems because the length of the desired products was between 96 and 246 nt in length, and the segments excised from the gel contained little or no contamination from short, non-templated extension products.

In the third step, the DNA molecules eluted from each of the ten size fractions were amplified by PCR using primers complementary to the T7 RNA polymerase promoter at the
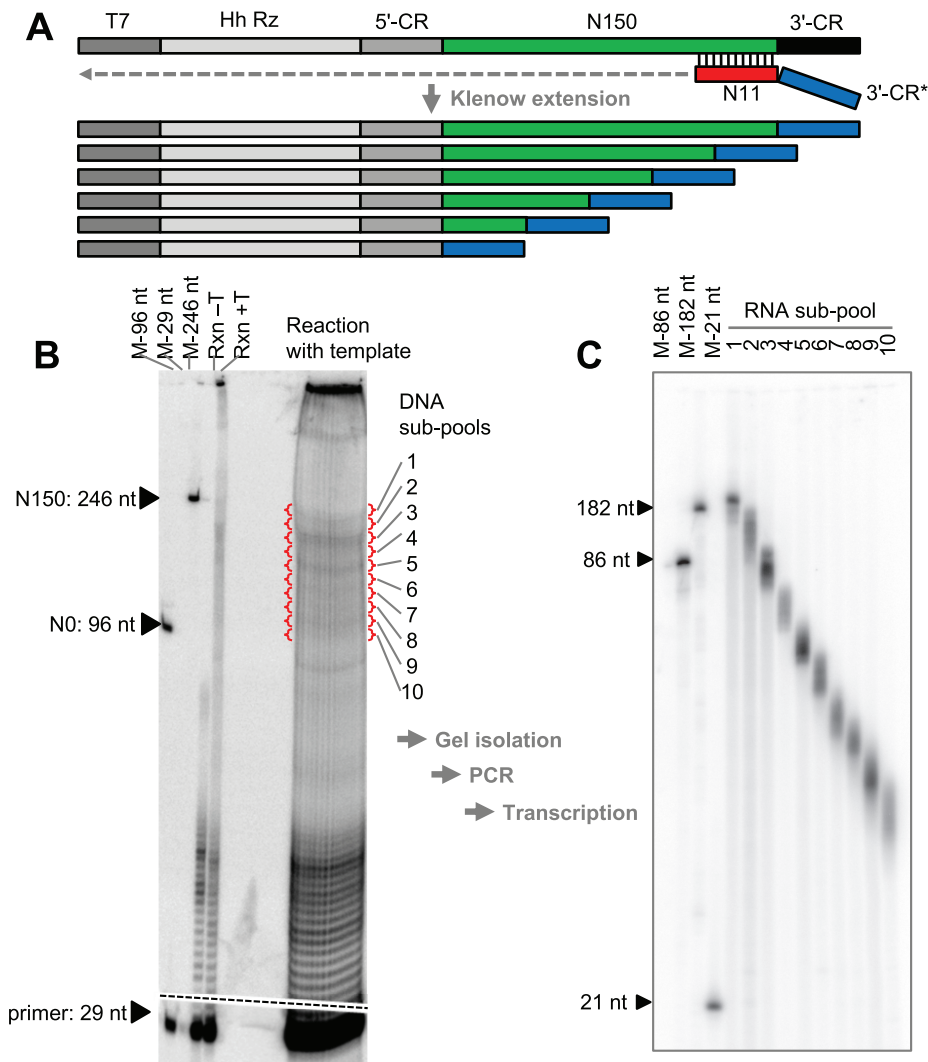
Figure 1. Workflow of generating ribozyme sub-pools with truncations at their 3′-end. (**A**) Schematic describing the extension of a primer with a randomized region (red) that can anneal anywhere on the template pool, and that introduces a new 3′-constant region (3′-CR*, blue). The annealed primers were then extended by the Klenow fragment. Only extension products with a random pool region (green) of 0–150 nucleotides are shown. (**B**) Autoradiogram of a denaturing 15% PAGE of Klenow extension products using 5′-radiolabeled primers. The first three lanes show markers that indicate the position of the unextended primer (29 nt), an extension product corresponding to the full-length pool (246 nt), and an extension product corresponding to the pool without randomized region (96 nt). The next two lanes show the reaction products of a small-scale reaction without template (Rxn-T) and with template (Rxn+T). The broad lane on the right shows the separation of a large-scale extension reaction with template (200 ul). Note that many products are shorter that the length of the full-length random region because the DNA library also contains a promoter for T7 RNA polymerase, a sequence encoding a hammerhead ribozyme, and a 5′-constant region. The dashed line indicates the position where two parts of the same exposure were assembled. We are unclear about the origin of the bands in the gel pockets; they may be an aggregation of template, hot primer, and Klenow fragment. The positions where ten sub-pools were excised from the gel are indicated with red brackets. After elution, each sub-pool was processed separately using PCR with a 5′-primer complementary to the T7 promoter, and a 3′-primer complementary to the new 3′-CR (blue in (A)). (**C**) Autoradiogram of 5′ end radiolabeled RNA sub-pools after transcription of the ten sup-pools, and separation by 8% denaturing PAGE. Three size markers indicate the position of full-length pool RNA (182 nt) and shorter fragments (86 nt and 21 nt).

5′-terminus (dark grey in Figure 1A), and the newly introduced 3′-constant region at the 3′-terminus (blue in Figure 1A). This PCR generated ten sub-pools with distinct size distributions. To test whether this procedure would give rise to RNA sub-pools with distinct sizes, each of the ten DNA sub-pools was transcribed, and a sample of the resulting RNA sub-pool was 5′ end radiolabeled with $\gamma[^{32}P]$ATP. The products were separated by PAGE (Figure 1C) and showed the desired size distributions. All ten sub-pools showed an expected overlap with the neighboring size fraction but no

significant overlap with other size fractions. Note that the RNA sub-pools are significantly shorter than the DNA sub-pools because the hammerhead ribozyme at the 5′-terminus of all transcripts cleaves co-transcriptionally, generating a 5′-hydroxyl group (30).

In the fourth step, the shortest sub-pool with active ribozymes was identified. To do this, all ten RNA sub-pools were subjected to one round of in vitro selection, using the same in vitro selection procedure that was used to select the ribozyme library from random sequence (30)
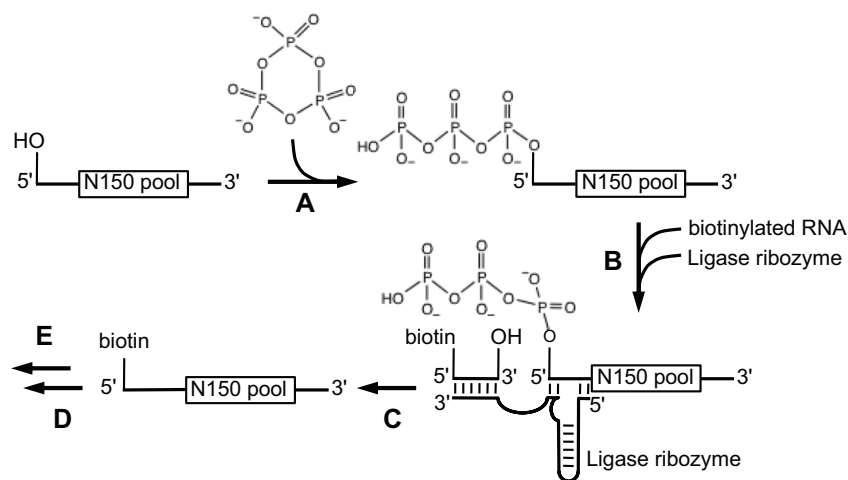
Figure 2. Schematic for selecting active ribozymes from each of the ten RNA sub-pools. (**A**) The RNA sub-pool is incubated with Tmp to allow active ribozymes to triphosphorylate their 5′-hydroxyl group, generating a 5′-triphosphate. (**B**) Pool molecules are annealed with a biotinylated RNA and a ligase ribozyme. The ligase ribozyme covalently links 5′-triphosphorylated pool molecules to biotin. (**C**) After capture of biotinylated RNAs on streptavidin beads, RNAs not linked covalently are washed away. Isolated Sequences are then (**D**) reverse transcribed and (**E**) PCR amplified.

(Figure 2). Specifically, the RNA sub-pools were incubated in the presence of Tmp such that catalytically active pool molecules could convert their 5′-hydroxyl to a 5′-triphosphate group. Pool molecules with a 5′-triphosphate were covalently linked to a biotinylated primer, using the R3C ligase ribozyme (19). Covalently linked pool molecules were isolated via their new biotin 5′-terminus using streptavidin beads, reverse transcribed, and PCR amplified. Importantly, only sub-pools 1–7 gave rise to a clean PCR product of the expected size (Supplementary Figure S1). Sub-pools 8 and 9 gave rise to a diffuse product distribution without a band at the expected size. This suggested that after one round of selection, sub-pool 7 contained the shortest pool sequences with self-triphosphorylation activity.

In the fifth step, catalytically active RNAs were identified from sub-pool 7. To do this, sub-pool 7 was cloned, and 21 clones were arbitrarily chosen for biochemical analysis and sequencing (Supplementary Figures S2 and S3). To measure the biochemical activity of each clone, an assay was used in which a radiolabeled primer was ligated to an equimolar concentration of individual RNAs after their incubation with Tmp, thereby quantifying the fraction of ribozymes that triphosphorylated their 5′-terminus (30) (Figure 3 and Supplementary Figure S3). After denaturing PAGE separation of the products, the quantified bands showed that ten of the 21 clones had biochemical activity (Figure 3A, B). When the sequences of these ribozymes were aligned, all ten active ribozymes fell within one cluster, and all 11 inactive clones belonged to other clusters (Figure 3B, C and Supplementary Figure S2). These results suggest that the cluster with these ten ribozymes represents the shortest ribozymes in the selected library. The individual sequence with highest activity, clone 78, was chosen for further biochemical analysis.

To identify the catalytic core of clone 78, additional truncations were made at its 3′-terminus, and the biochemical activity was tested for each truncation variant (Figure 4). Truncations at the 5′ end were not made since the 5′ terminus is the catalytic site for the selected activity. The av-

erage activity was highest when the ribozyme was truncated to 54 nucleotides, and the average activity of clone 78 (with a length of 72 nucleotides) was maintained when the ribozyme was truncated down to 44 nucleotides. Analysis of secondary structures predicted by vsFold5 (45) found that the ribozymes between 57 and 44 nucleotides in length could fold into a conserved pseudoknot structure, whereas longer and shorter variants of the ribozyme did not (Supplementary Figure S4).

To probe the secondary structure of the 44-nucleotide ribozyme we used selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) (40) (Figure 5). The SHAPE probing data were mostly consistent with the formation of the same pseudoknot as predicted by vsFold5. While the existence of the P2 stem is confirmed, the P1 stem may be shorter than the expected six base pairs, or its ends may be dynamic under the conditions of the SHAPE assay. The P2 stem may be slightly longer than the expected 6 bp through interactions of the bases C16/C17 with A38/G39. However, the SHAPE data of these bases may also reflect a more complex interaction, for example involving the protected bases C23/G24. A tertiary structure determination may help reveal, for example, interesting noncanonical base pair interactions or the three-dimensional shape of the Tmp binding pocket. Previous efforts in RNA 3D structure prediction have not yet been successful in predicting the structure of large ribozymes and most successful predictions have been made on short aptamers, especially on those where the structure of a known homolog has been determined (46–48). The ribozyme identified in this study may be a good candidate for 3D structure prediction due to its small size since the quality of a model correlates with size. Nevertheless, the pseudoknot shown in Figure 5B represents the best working model for the ribozyme's secondary structure.

When the sequence of the 44-nucleotide ribozyme was tested for similarities in high throughput sequencing data of the original selection (39), two additional clusters were identified that contained the same 17-nucleotide long motif
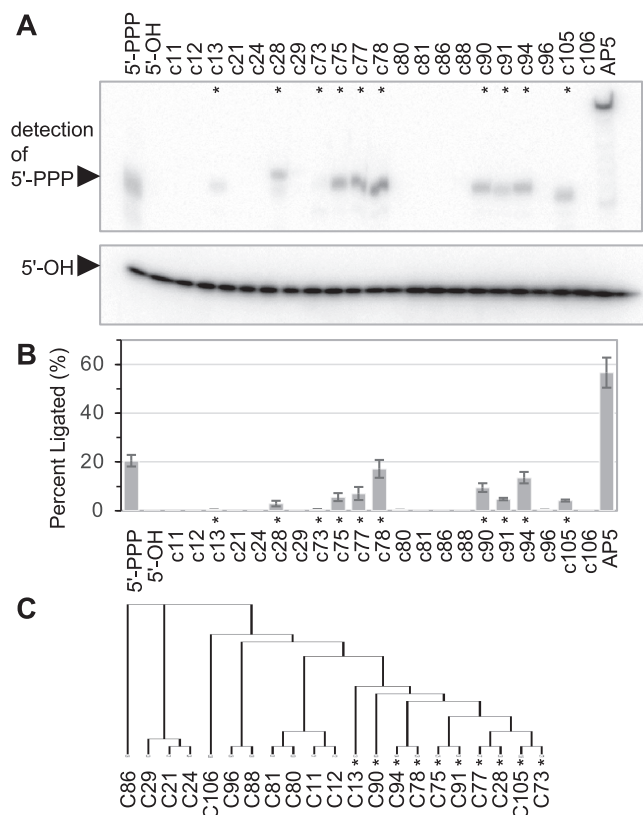
Figure 3. Biochemical and phylogenetic analysis of 21 cloned sequences from sub-pool 7. (**A**) Autoradiograph of a gel-shift assay to detect self-triphosphorylation activity. The lower panel shows unreacted RNAs, whereas bands in the upper panel show triphosphorylation activity. Triphosphorylated RNA (5′-PPP) and RNA with a 5′-hydroxyl group (5′-OH) served as positive and negative control, respectively. The highly active ribozyme AP5 (39) was used as additional positive control. AP5 is 182 nucleotides in length. Clones with an average activity above the detection limit of 0.1% are marked with an asterisk. (**B**) Quantitation of triphosphorylation assays as shown in (A). Error bars are standard deviations from triplicate experiments. (**C**) Phylogenetic comparison of the 21 sequences from sub-pool 7. The phylogenetic tree was generated with the software Geneious, using the neighbor-joining treebuild method. All clones with detectable biochemical activity (higher than 0.1%) are marked with an asterisk.
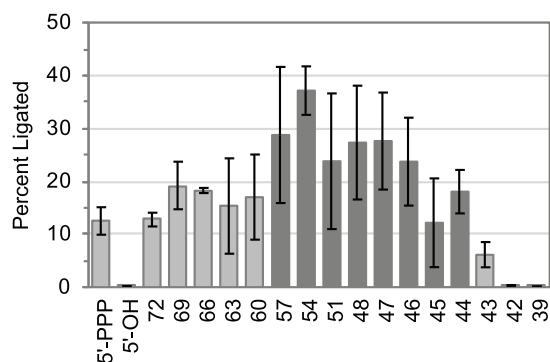

Figure 4. Biochemical activity of variants of ribozyme C78 that are successively truncated at their 3′-terminus. In addition to positive control (5′-PPP) and negative control (5′-OH), the length of the 3′-truncation variants is used as label. Clone C78 has a length of 72 nucleotides. Error bars are standard deviations from triplicate experiments. Columns highlighted in dark grey are predicted to fold into a pseudoknot structure (see Supplementary Figure S4).

(Figure 6). All three sequences are consistent with a model in which the RNA can fold back on itself to generate helix P1, immediately followed by an eight-nucleotide sequence that was identical between the three clusters (red in Figure 6). The six nucleotides preceding the P1 helix contained an additional three conserved nucleotides (green in Figure 6), forming an element of 21 nucleotides with 17 conserved positions. The three clusters containing this motif differed in the sequence and length of insertions before and after this 21-nucleotide long element. Additionally, the 3′-terminus of the pseudoknot forming the P2 helix was heterogeneous in sequence and appeared to be sufficient with 5 base pairs. At the 5′-terminus, the first 14 nucleotides of the ribozyme are defined by the constant region, which includes a two-nucleotide linker between helices P1 and P2 (purple). Because the P2 helix could also form with other sequences downstream, the length of this linker—two nucleotides—is likely required for activity.

To identify the optimal reaction conditions of the 44-nucleotide ribozyme we measured the biochemical activity at different concentrations of Tmp and Mg^{2+}, at different pH values and at different temperatures (Figure 7). At a Mg^{2+} concentration of 100 mM, the optimum of the Tmp concentration saturates at or above 200 mM (Figure 7A). In contrast, at a Tmp concentration of 200 mM, the optimal Mg^{2+} concentration does not exceed 100 mM (Figure 7B). This is different from the other well-characterized triphosphorylation ribozyme TPR1, which shows optimal activity only when the Mg^{2+} concentration exceeds the Tmp concentration (30), and suggests that these two ribozymes use different catalytic mechanisms. The pH dependence of the 44-nucleotide ribozyme shows increased activity at higher pH values (Figure 7C) as expected when the deprotonation of the 5′-hydroxyl group is the likely rate-limiting step of the reaction. The temperature optimum of the 44-nucleotide ribozyme is around 10°C (Figure 7D), different from the optimum of a TPR1 variant at 40°C (49), underlining the different characteristics of these two ribozymes. Together, the optimum reaction conditions for the 44-nucleotide long ribozyme were 200 mM Tmp, 100 mM Mg^{2+}, a pH of 8.5, and a temperature of 10°C.

Together, the results establish a new technique to isolate short ribozymes from combinatorial libraries containing many ribozymes, and provide an initial characterization of the smallest identified self-triphosphorylation ribozyme. The technique to identify small ribozymes from large ribozyme libraries may be employed for different ribozyme libraries and thereby help to identify short ribozymes as tools, as well as explore the feasibility of RNA systems that are small enough to have emerged from a prebiotic environment. For instance, the 44-nucleotide ribozyme identified in this study is close in length to products generated by model prebiotic reactions of RNA polymerization (33) and the 17-nucleotide conserved motif is well within this range.

DISCUSSION

The method described in this study identified a self-triphosphorylation ribozyme with a conserved motif of 17 nucleotides, and likely additional constraints at positions 21, 23, 24 and up to 5 positions that form the P2 duplex
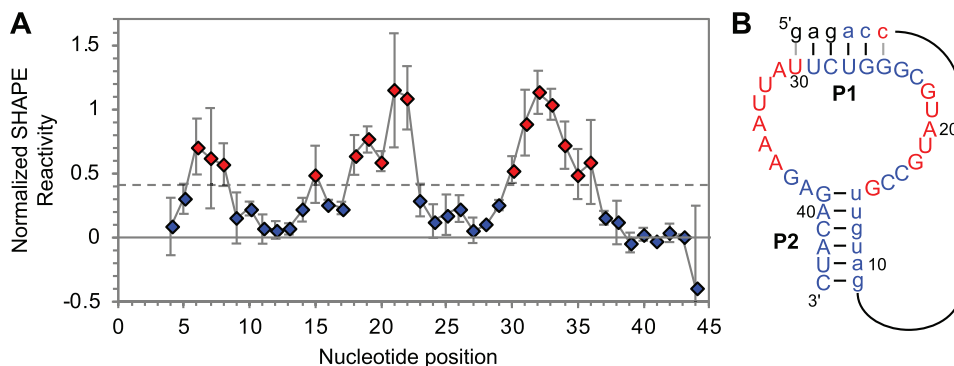
Figure 5. SHAPE probing of the secondary structure of the 44-nucleotide long ribozyme. (**A**) Normalized SHAPE reactivity as a function of the nucleotide position in the ribozyme. The reactivities of the average of the five most reactive positions were set to 1.0. The reactivity of 0.4 was set as the cutoff between 'low reactivity'(blue) and 'high reactivity' (red). The first three nucleotides (black) did not yield SHAPE data because they were too close to the radiolabeled 5′-terminus to be separated on the used denaturing polyacrylamide gels. Error bars are standard deviations from triplicate experiments. (**B**) Pseudoknot structure that is mostly consistent with the SHAPE probing data. Nucleotides with low SHAPE reactivity are shown in blue, positions with high reactivity in red.
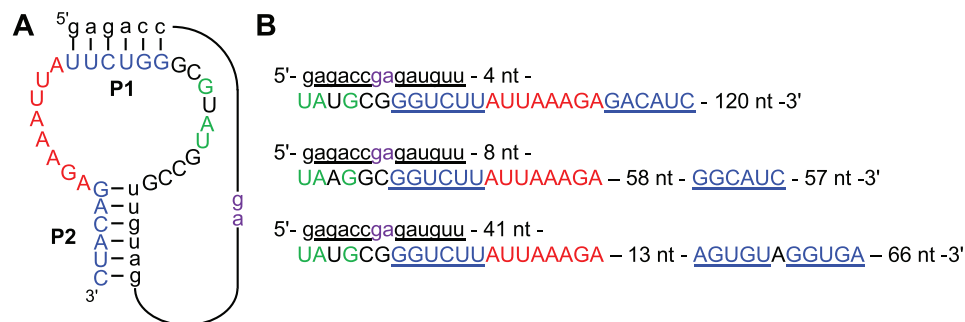

Figure 6. Sequence conservation among three different clusters that appear to form the same catalytic core. (**A**) Secondary structure with annotation of nucleotides that are within the primer binding site of the library (lower case), a two-nucleotide linker (purple), three conserved nucleotides in the L2/1 loop (green), sequences that form P1 and P2 (blue), and an eight-nucleotide sequence that is completely conserved among the three clusters (red). (**B**) Three peak sequences from clusters of the original selection. They could form the same secondary structure but with different sizes of loops between the P1 and P2 duplexes.

(see figures 5 and 6). For a given RNA sequence of length, n, the total sequence space is 4^n. Thus, the conserved motif identified in this study is likely to be found about once in 4^{20} ($\sim 1.1 \times 10^{12}$) sequences. The original library contained an effective complexity of 1.7×10^{14} sequences (30), therefore we would expect to find about 150 such ribozymes. However, we identified only three such ribozymes in the original library. There are several possibilities to explain this discrepancy. First, there may be additional sequence determinants in the ribozyme. For example, the nucleotides at position 15–18 may be constrained, and could contribute up to a 256-fold reduction in complexity. Second, the nucleotides forming the P2 stem (positions 39–44) may not be as flexible as the sequences of the second and third clones in the original library would suggest (Figure 6B). Third, a larger size of the loops between conserved elements may reduce the fraction of ribozyme folding into the active conformation. Therefore, most sequences with the 17-nucleotide motif near the 3′-terminus of the 150 randomized nucleotides may be inactive. This idea is supported by the finding that the conserved regions of all three putative ribozymes are within the first 100 nucleotides of the pool. Awaiting a detailed characterization of the sequence requirement of this

small ribozyme, an incidence number of three in the original library is therefore within the expected range.

It is possible that there exist other, shorter triphosphorylation ribozymes that were not discovered. However, the 17-nucleotide motif of the 44-nucleotide ribozyme was identified three times in the library, suggesting that any smaller motif would have been contained even more frequently, and identified in the used procedure. Two types of self-triphosphorylation ribozymes would not have been discovered by the original selection: First, the random sequence library contained 14 nucleotides of fixed sequence at its 5′-terminus. Any ribozyme with different sequence requirements within the first 14 nucleotides would not have been contained in the original library. Second, the selection procedure requires that after self-triphosphorylation, the ribozyme 5′-terminus is unfolded and based paired to the 5′-terminus of a ligase ribozyme. Any ribozyme with a very stable structure at their 5′-terminus that does not allow annealing with the ligase ribozyme would not be selected. The procedure of this study relies on the same selection principle as the original selection, therefore the procedure described in this study would likely have found any ribozymes in the original library with a smaller sequence motif than the iden-
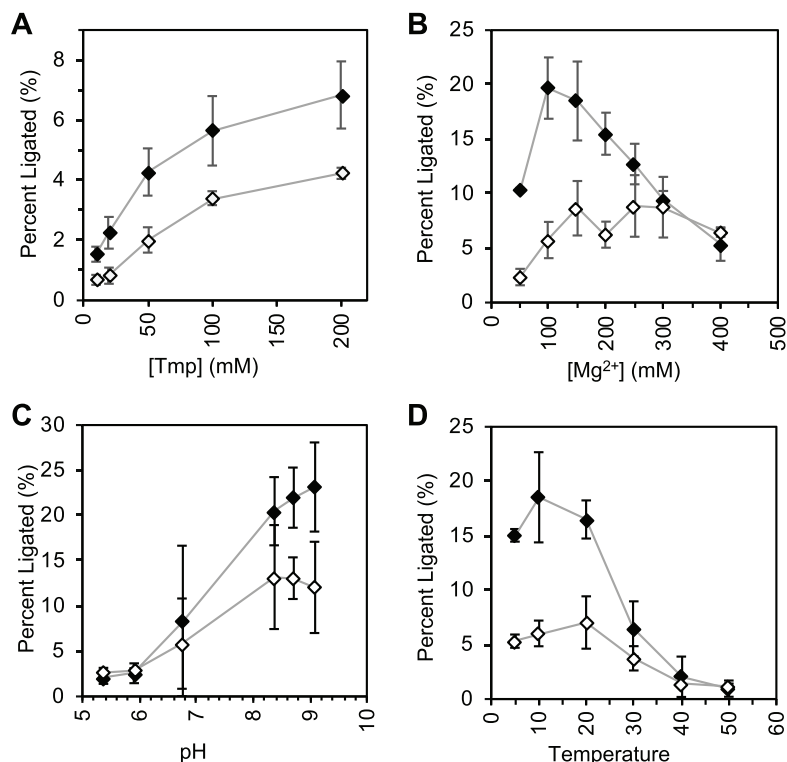
Figure 7. Dependence of ribozyme activity on reaction conditions. (**A**) Dependence of ribozyme activity on the concentration of trimetaphosphate (Tmp) at 100 mM MgCl$_2$ and 50 mM Tris/HCl pH 8.5. Empty diamonds correspond to reaction times of 2 min, filled diamonds to 20 min. (**B**) Dependence of ribozyme activity on the concentration of MgCl$_2$ at 200 mM Tmp and 50 mM Tris/HCl pH 8.5. (**C**) Dependence of ribozyme activity on the pH at 200 mM Tmp and 100 mM MgCl$_2$. (**D**) Dependence of ribozyme activity on reaction temperature at 200 mM Tmp, 100 mM MgCl$_2$, and 50 mM Tris/HCl pH 8.5. In all cases, error bars are standard deviations of triplicate experiments.

tified ribozyme. Similarly, other libraries of in vitro selected ribozymes can now be tested with the same approach, each using their own selection scheme.

It is interesting to deliberate whether computational tools would be able to identify catalytic motifs within High Throughput Sequencing data of in vitro selected RNA populations. One possibility for such a procedure is to use sequence conservation data in each of the clusters, and thereby isolate those sequence fragments that are highly conserved. HTS data on our own in vitro selection were not deep enough to allow such analysis (39). However, the high mutagenic rates, and defined nucleotide bias used in 'DNA shuffling' are able to tease out the secondary / tertiary structure of a specific ribozyme with a limited depth of sequencing (50). With the advance of more powerful High Throughput Sequencing platforms it may become possible to efficiently identify co-variations even from in vitro selected libraries with low mutagenesis. While this may generate a computational alternative to the experimental approach described here, it would be limited by identifying only conserved structural motifs as opposed to catalytic function.

The experimental method described in this study could generate a bias to identify motifs that end in, or are followed by GC-rich sequences. The reason is that such sequences would generate more stable duplexes with the randomized N11-region on the primer used in the Klenow extension reaction (see Figure 1A). Indeed, most of the identified clones carry a C-rich sequence immediately after the motif. While this could reduce the chance of identifying motifs with AU-rich sequences it may not be a serious concern, especially if randomized regions with more than 11 nucleotides, and/or High Throughput Sequencing is used to analyze the selected sequences.

The described procedure uses truncations at the library 3′-terminus, and therefore depends on the 5′-terminus of the library to be the position where the reaction occurs. The majority of in vitro selections for ribozymes have used the 5′-terminus as the reaction site (3,6–10,13,15–20,25,28,29), therefore the described procedure should be widely applicable. However, several in vitro selections focused on internal 2′-hydroxyl groups as reactive groups; for these selections the proposed method would be unsuitable (12,26,27). Other in vitro selections have used the 3′-terminus of the library as the position where the reaction occurs (11,24,36). In these instances, a variation of the described procedure could be successful: Instead of the 3′-primer with a randomized 3′-terminus and a new 3′-constant region for Klenow extension (see Figure 1A), a 5′-primer could be used that truncates the ribozyme 5′-terminus. After PAGE purification of the extension products, necessary elements at the 5′-terminus such as the promoter for T7 RNA polymerase could be added by PCR. In this way, the described procedure would be useful for the isolation of short ribozymes for many types of selections, with reaction centers positioned at the library 5′-terminus and 3′-terminus.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
2. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
3. Bartel,D.P. and Szostak,J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science*, **261**, 1411–1418.
4. Illangasekare,M., Sanchez,G., Nickles,T. and Yarus,M. (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. *Science*, **267**, 643–647.
5. Wilson,C. and Szostak,J.W. (1995) In vitro evolution of a self-alkylating ribozyme. *Nature*, **374**, 777–782.
6. Lohse,P.A. and Szostak,J.W. (1996) Ribozyme-catalysed amino-acid transfer reactions. *Nature*, **381**, 442–444.
7. 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.
8. Tarasow,T.M., Tarasow,S.L. and Eaton,B.E. (1997) RNA-catalysed carbon-carbon bond formation. *Nature*, **389**, 54–57.
9. Wiegand,T.W., Janssen,R.C. and Eaton,B.E. (1997) Selection of RNA amide synthases. *Chem. Biol.*, **4**, 675–683.
10. Zhang,B. and Cech,T.R. (1997) Peptide bond formation by in vitro selected ribozymes. *Nature*, **390**, 96–100.
11. Unrau,P.J. and Bartel,D.P. (1998) RNA-catalysed nucleotide synthesis. *Nature*, **395**, 260–263.
12. Jenne,A. and Famulok,M. (1998) A novel ribozyme with ester transferase activity. *Chem. Biol.*, **5**, 23–34.
13. Seelig,B. and Jaschke,A. (1999) A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.*, **6**, 167–176.
14. Jaeger,L., Wright,M.C. and Joyce,G.F. (1999) A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 14712–14717.
15. Robertson,M.P. and Ellington,A.D. (1999) In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat. Biotechnol.*, **17**, 62–66.
16. Lee,N., Bessho,Y., Wei,K., Szostak,J.W. and Suga,H. (2000) Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Biol.*, **7**, 28–33.
17. Kumar,R.K. and Yarus,M. (2001) RNA-catalyzed amino acid activation. *Biochemistry*, **40**, 6998–7004.
18. Sengle,G., Eisenfuhr,A., Arora,P.S., Nowick,J.S. and Famulok,M. (2001) Novel RNA catalysts for the Michael reaction. *Chem. Biol.*, **8**, 459–473.
19. 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.
20. Tsukiji,S., Pattnaik,S.B. and Suga,H. (2003) An alcohol dehydrogenase ribozyme. *Nat. Struct. Biol.*, **10**, 713–717.
21. Hati,S., Boles,A.R., Zaborske,J.M., Bergman,B., Posto,A.L. and Burke,D.H. (2003) Nickel2+-mediated assembly of an RNA-amino acid complex. *Chem. Biol.*, **10**, 1129–1137.
22. Ikawa,Y., Tsuda,K., Matsumura,S. and Inoue,T. (2004) De novo synthesis and development of an RNA enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 13750–13755.
23. Yoshioka,W., Ikawa,Y., Jaeger,L., Shiraishi,H. and Inoue,T. (2004) Generation of a catalytic module on a self-folding RNA. *RNA*, **10**, 1900–1906.
24. Lau,M.W., Cadieux,K.E. and Unrau,P.J. (2004) Isolation of fast purine nucleotide synthase ribozymes. *J. Am. Chem. Soc.*, **126**, 15686–15693.
25. Fusz,S., Eisenfuhr,A., Srivatsan,S.G., Heckel,A. and Famulok,M. (2005) A ribozyme for the aldol reaction. *Chem. Biol.*, **12**, 941–950.
26. Saran,D., Nickens,D.G. and Burke,D.H. (2005) A trans acting ribozyme that phosphorylates exogenous RNA. *Biochemistry*, **44**, 15007–15016.
27. Li,N. and Huang,F. (2005) Ribozyme-catalyzed aminoacylation from CoA thioesters. *Biochemistry*, **44**, 4582–4590.
28. Ryu,Y., Kim,K.J., Roessner,C.A. and Scott,A.I. (2006) Decarboxylative Claisen condensation catalyzed by in vitro selected ribozymes. *Chem. Commun. (Camb.)*, **13**, 1439–1441.
29. Wang,T.P., Su,Y.C., Chen,Y., Liou,Y.M., Lin,K.L., Wang,E.C., Hwang,L.C., Wang,Y.M. and Chen,Y.H. (2012) In vitro selection and characterization of a novel Zn(II)-dependent phosphorothiolate thiolesterase ribozyme. *Biochemistry*, **51**, 496–510.
30. Moretti,J.E. and Muller,U.F. (2014) A ribozyme that triphosphorylates RNA 5′-hydroxyl groups. *Nucleic. Acids. Res.*, **42**, 4767–4778.
31. Lehman,N. and Joyce,G.F. (1993) Evolution in vitro of an RNA enzyme with altered metal dependence. *Nature*, **361**, 182–185.
32. Wang,Q.S. and Unrau,P.J. (2005) Ribozyme motif structure mapped using random recombination and selection. *RNA*, **11**, 404–411.
33. Huang,W. and Ferris,J.P. (2003) Synthesis of 35–40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem. Commun. (Camb.)*, **12**, 1458–1459.
34. Coleman,T.M. and Huang,F. (2005) Optimal random libraries for the isolation of catalytic RNA. *RNA Biol*, **2**, 129–136.
35. Sabeti,P.C., Unrau,P.J. and Bartel,D.P. (1997) Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem. Biol.*, **4**, 767–774.
36. Chumachenko,N.V., Novikov,Y. and Yarus,M. (2009) Rapid and simple ribozymic aminoacylation using three conserved nucleotides. *J. Am. Chem. Soc.*, **131**, 5257–5263.
37. Pasek,M.A., Kee,T.P., Bryant,D.E., Pavlov,A.A. and Lunine,J.I. (2008) Production of potentially prebiotic condensed phosphates by phosphorus redox chemistry. *Angew. Chem. Int. Ed. Engl.*, **47**, 7918–7920.
38. Pasek,M.A., Harnmeijer,J.P., Buick,R., Gull,M. and Atlas,Z. (2013) Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 10089–10094.
39. Pressman,A., Moretti,J.E., Campbell,G.W., Muller,U.F. and Chen,I.A. (2017) Analysis of in vitro evolution reveals the underlying distribution of catalytic activity among random sequences. *Nucleic Acids Res.*, **45**, 8167–8179.
40. Mortimer,S.A. and Weeks,K.M. (2007) A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.*, **129**, 4144–4145.
41. Jabbari,H. and Condon,A. (2014) A fast and robust iterative algorithm for prediction of RNA pseudoknotted secondary structures. *BMC Bioinformatics*, **15**, 147.
42. Sato,K., Kato,Y., Hamada,M., Akutsu,T. and Asai,K. (2011) IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics*, **27**, i85–93.
43. Andronescu,M.S., Pop,C. and Condon,A.E. (2010) Improved free energy parameters for RNA pseudoknotted secondary structure prediction. *RNA*, **16**, 26–42.
44. Chen,H.L., Condon,A. and Jabbari,H. (2009) An O(n(5)) algorithm for MFE prediction of kissing hairpins and 4-chains in nucleic acids. *J. Comput. Biol.*, **16**, 803–815.
45. Dawson,W.K., Fujiwara,K. and Kawai,G. (2007) Prediction of RNA pseudoknots using heuristic modeling with mapping and sequential folding. *PLoS One*, **2**, e905.
46. Miao,Z., Adamiak,R.W., Blanchet,M.F., Boniecki,M., Bujnicki,J.M., Chen,S.J., Cheng,C., Chojnowski,G., Chou,F.C., Cordero,P. *et al.* (2015) RNA-Puzzles Round II: assessment of RNA structure

prediction programs applied to three large RNA structures. *RNA*, **21**, 1066–1084.

47. Miao,Z., Adamiak,R.W., Antczak,M., Batey,R.T., Becka,A.J., Biesiada,M., Boniecki,M.J., Bujnicki,J.M., Chen,S.J., Cheng,C.Y. *et al.* (2017) RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*, **23**, 655–672.

48. Miao,Z., Adamiak,R.W., Antczak,M., Boniecki,M.J., Bujnicki,J., Chen,S.J., Cheng,C.Y., Cheng,Y., Chou,F.C., Das,R. *et al.* (2020) RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA*, **26**, 982–995.

49. Akoopie,A. and Muller,U.F. (2016) Lower temperature optimum of a smaller, fragmented triphosphorylation ribozyme. *Phys. Chem. Chem. Phys.*, **18**, 20118–20125.

50. Curtis,E.A. and Bartel,D.P. (2013) Synthetic shuffling and in vitro selection reveal the rugged adaptive fitness landscape of a kinase ribozyme. *RNA*, **19**, 1116–1128.