Published online 16 June 2021

Nucleic Acids Research, 2021, Vol. 49, No. 12 ***6971–6981***
https://doi.org/10.1093/nar/gkab504

Single-round deoxyribozyme discovery

Tereza Streckerová[1,2], Jaroslav Kurfürst[1,3] and Edward A. Curtis[1,*]

[1]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague 160 00, Czech Republic, [2]Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague 160 00, Czech Republic and [3]Department of Informatics and Chemistry, University of Chemistry and Technology, Prague 166 28, Czech Republic

Received April 06, 2021; Revised May 14, 2021; Editorial Decision May 26, 2021; Accepted May 31, 2021

ABSTRACT

Artificial evolution experiments typically use libraries of ∼10^{15} sequences and require multiple rounds of selection to identify rare variants with a desired activity. Based on the simple structures of some aptamers and nucleic acid enzymes, we hypothesized that functional motifs could be isolated from significantly smaller libraries in a single round of selection followed by high-throughput sequencing. To test this idea, we investigated the catalytic potential of DNA architectures in which twelve or fifteen randomized positions were embedded in a scaffold present in all library members. After incubating in either the presence or absence of lead (which promotes the nonenzymatic cleavage of RNA), library members that cleaved themselves at an RNA linkage were purified by PAGE and characterized by high-throughput sequencing. These selections yielded deoxyribozymes with activities 8- to 30-fold lower than those previously isolated under similar conditions from libraries containing 10^{14} different sequences, indicating that the disadvantage of using a less diverse pool can be surprisingly small. It was also possible to elucidate the sequence requirements and secondary structures of deoxyribozymes without performing additional experiments. Due to its relative simplicity, we anticipate that this approach will accelerate the discovery of new catalytic DNA and RNA motifs.

INTRODUCTION

The discovery of catalytic RNA in the early 1980s raised the intriguing possibility that artificial enzymes made of nucleic acids could be created in the lab. This led to the development of methods of artificial evolution which made it possible to isolate DNA and RNA molecules with desired functional properties from large random sequence pools (1–3). Application of these methods revealed that both DNA and RNA molecules can bind ligands with high affinity and specificity and catalyze a wide range of reactions (4–7). An advantage of constructing functional motifs from nucleic acids rather than proteins is that they are typically less expensive to synthesize and simpler to prepare (8). They are also easier to optimize and modify using artificial evolution.

The starting point in a typical selection experiment is a library of ∼10^{15} random sequence DNA or RNA molecules. Although libraries of this size might be needed to isolate motifs with complex functions, accumulating evidence suggests that they are not required to find simple aptamers and catalysts. For instance, a number of simple functional motifs have been identified which would be expected to occur multiple times in libraries of 10^{15} random sequences (9). One example is a 25 nucleotide ATP aptamer made up of a short stem and a conserved bulge of 12 nucleotides (10). Given reasonable assumptions about the sequence requirements of this aptamer, it should occur ∼10^{4} times in a 25-nucleotide random sequence library of ∼10^{15} sequences (or about once in a library of 10^{11} sequences), and might be even more abundant in longer random sequence pools (11–12). Even simpler motifs have also been reported, including a self-aminoacylating ribozyme made up of nine nucleotides (13), a manganese-dependent self-cleaving ribozyme containing seven nucleotides (14–15), a lead-dependent self-cleaving ribozyme consisting of six essential nucleotides in an asymmetric bulge flanked by short stems (16–18), and a GTP aptamer likely formed by homodimerization of a seven-nucleotide motif (19). This GTP aptamer could in principle be isolated from a 1.6×10^{4}-member library containing all possible 7-nucleotide sequences.

These observations indicate that under certain conditions it should be possible to isolate functional motifs from random sequence libraries that are significantly less complex than those typically used in selection experiments. In the case of aptamers this has been demonstrated experimentally: thrombin aptamers were successfully isolated from a library containing only 10^{9} sequences (20). In light of recent developments in high-throughput nucleic acid sequencing technologies (21), the possibility of reducing library sizes to ∼10^{7} unique sequences (corresponding to about twelve randomized positions) is of particular interest. If a library

*To whom correspondence should be addressed. Tel: +420 733 169 654; Email: curtis@uochb.cas.cz

of this size turned out to contain functional motifs, these could in principle be identified by performing a single round of selection followed by high-throughput sequencing. In addition, the number of reads obtained (10^7 using standard methods but up to 10^9 using platforms like NovaSeq) would be sufficient to provide information about most active variants in the library. However, the structural diversity of short oligonucleotides is limited, and it seems unlikely that, despite the odd GTP aptamer or self-cleaving ribozyme, an N_{12} library would contain a wide range of functional motifs. To address this possible limitation, here we investigated the functional potential of a DNA architecture in which twelve randomized positions were placed in a bulge flanked by constant sequences predicted to form stable stems. We hypothesized that the $\sim 10^7$ possible bulge sequences in our DNA architecture would provide sufficient diversity to form a range of simple binding pockets and active sites, while the stems would stabilize these functional elements in a sequence-independent manner. Our architecture was inspired by a number of known functional nucleic acids with similar folds (9–10,22–27). It was also influenced by high-resolution structural studies which show that, in at least some cases, single-stranded nucleotides in bulges of aptamers make direct contacts to ligands while flanking stems provide nonspecific stabilization (22–27). Quantitative analysis of the relationship between information content and activity for a series of GTP aptamers and ligase ribozymes also suggests that 24 bits of information in unpaired regions of structured motifs (corresponding to 12 invariant positions or a larger number of less conserved positions) can be sufficient to form a binding pocket or active site (9).

To test the functional potential of this DNA architecture, we performed single-round selections for the ability of library members to cleave an RNA linkage under several different conditions. Analysis of high-throughput sequencing data revealed hundreds to tens of thousands of enriched sequences, some of which enhanced RNA cleavage by factors of up to ~5000. Because each of the 1.7×10^7 possible sequence variants of the randomized bulge were present in the starting library, our single-round selections also provided information about the sequence requirements and secondary structures of these motifs. A second library containing fifteen randomized positions also yielded deoxyribozymes, but provided less information about their sequence requirements. This is probably due to less complete sampling of this more complex library by high-throughput sequencing. Because such selections can be performed easily and analyzed quickly relative to conventional methods, this approach has the potential to simplify the discovery and characterization of novel catalytic nucleic acid motifs.

MATERIALS AND METHODS

Preparation of constructs containing a single RNA linkage

All sequences were purchased from IDT (Supplementary Table S1). Chimeric DNA/RNA constructs shorter than 50 nucleotides were ordered directly and gel-purified on 8% denaturing PAGE gels. Longer constructs (including libraries used for selections) were generated by ligation of two oligonucleotides: 15 μM of the 5′ part of the sequence and 5 μM of the 5′ phosphorylated 3′ part of the sequence (containing the RNA nucleotide) was mixed with 15 μM of a splint oligonucleotide, 1× of T4 DNA ligase buffer (NEB) and T4 DNA ligase (NEB) at a final concentration of 4000 units/ml. The mixture was incubated at 16°C for 1 hour and gel-purified on a 7% denaturing PAGE gel. The ligated product (82 nucleotides for the N_{12} library and 85 nucleotides for the N_{15} library) could be readily separated from the splint (32 nucleotides) under these conditions.

Single-round selections

All buffer components were purchased from Sigma-Aldrich. Two different buffers were used. The first one contained lead and was similar to that used in a previous selection for RNA-cleaving deoxyribozymes (28). This buffer contained 0.5 M KCl, 50 mM $MgCl_2$, 50 mM HEPES pH 7.0 and 1 mM $PbCl_2$. All ingredients except $PbCl_2$ were mixed to make a 2× solution. 10 mM $PbCl_2$ was added directly into the reaction. The second buffer was prepared as described in (29), and contained 1 M NaCl, 10 mM $MgCl_2$, and 50 mM Tris-HCl pH 7.5. The stock solution of this buffer was prepared at a 2× concentration.

The selection step was performed as followed: 83 μl of 10 μM stock solution of the library (5×10^{14} molecules) was mixed with 332 μl of water, heated to 65°C for 5 minutes, and cooled at room temperature for 5 minutes. 415 μl of 2× selection buffer was added (or 332 μl of 2.5 × buffer and 83 μl of 10 mM $PbCl_2$) and the reaction mixture was incubated at 37°C for 30 minutes. The reaction was stopped by desalting using SigmaSpin™ Sequencing Reaction Clean-up columns (Sigma-Aldrich). Afterwards, the DNA was ethanol precipitated and the pellet resuspended in 50 μl of gel-loading buffer (4 M urea; 1 mM Tris, pH 7.5; 10 mM EDTA; 0.028 g/l of bromophenol blue). The sample was then heated at 70°C for 2 minutes, and purified on a 6% denaturing PAGE gel run at 300 volts for 70 minutes. Bands corresponding to reacted library members (54 nucleotides for the library containing 12 randomized positions and 57 nucleotides for the library containing 15 randomized positions) were excised, eluted in 450 μl of 0.3 M NaCl, and ethanol precipitated together with 5 μg of yeast tRNA (Ambion). Pellets were resuspended in 50 μl of water.

Elongation and amplification of the library

Because reacted members of the library were too short to be sequenced by Illumina, they needed to be elongated by primer extension. To do this, the entire 50 μl sample was mixed with 2.8 μl of 10 μM elongation oligo (TS_49.2_OSS_PEX_rev), 21.5 μl of 5 × Q5 buffer, 21.5 μl of 5 × GC enhancer, 10.8 μl of 5 mM dNTPs, and 1.1 μl of Q5 Hot Start High-Fidelity DNA Polymerase (NEB) and subjected to three thermal cycles of 98°C for 1 minute, 56°C for 1 minute, and 72°C for 2 minutes using a Bio-Rad T100™ Thermal cycler. The mixture was then diluted 3-fold into water and used as a template for PCR. 15 μl of this template was used in a 50 μl pilot reaction to determine the minimum number of cycles needed to generate a sufficient amount of material for sequencing. After every 5 cycles, 5 μl of the PCR reaction was removed and analyzed on an agarose gel. Subsequently, a preparative PCR reaction was

performed. This contained 308 μl of template, 10.3 μl of 100 μM TS_17.7 fw, 10.3 μl of 100 μM TS_20.16 rev, 185 μl of 5 × Q5 buffer, 185 μl of 5× GC enhancer, 20.5 μl of 25 mM dNTPs, 298 μl of water, and 10.3 μl of Q5 Hot Start High-Fidelity DNA Polymerase (NEB) in a total volume of 1027 μl. The thermocycler was set as follows: Initial denaturation 98°C 1 minute; cycles: 98°C 20 seconds; 52°C 20 seconds; 72°C 1 minute; final extension 72°C 2 minutes. The choice of the polymerase and the presence of GC enhancer were important as none of the other conditions we tested generated product efficiently in the PCR. Reaction products were analyzed on agarose gels and purified using NucleoSpin Gel and PCR Clean-up columns (Macherey-Nagel). The product was again loaded on an agarose gel and the concentration was measured by Qubit assay (Invitrogen). Purified material was characterized using Illumina sequencing (Seqme) with a 100 bp single-end setting (results summarized in Supplementary Table S2).

Analysis of sequencing data

Sequencing data were preprocessed by adaptor trimming, primer clipping, length filtering, and quality filtering using cutadapt (v2.8). Unique sequences were counted and sorted using basic bash commands. Further analysis was done using in-house scripts.

Identification of correlated pairs of positions

Heat maps were constructed by first determining the read numbers of each of the possible single and double mutants of a given reference sequence. For each of the double mutants, an 'expected read number' was computed by dividing the read number of each of the single mutants that make up the double mutant by the read number of the reference sequence and multiplying these two numbers. The observed read number of the double mutant was then divided by this 'expected read number'. The result of this 'observed/expected' value was scaled to a <0, 1> interval as well as to the double mutant read number. Finally, the scaled double mutant read number was multiplied by the scaled 'observed/expected' value. The resulting value was added to the corresponding value in a matrix describing all possible pairs of positions. The heat maps are visualizations of this matrix.

Initial velocity and activity measurements

All activity measurements were done in triplicate. Most experiments were performed using trimmed variants of the scaffold used for selection (Figure 1), in which the possible primer binding sites at the 5′ and 3′ ends were deleted. 1 μl of a 10 μM oligonucleotide solution was mixed with 4 μl of water, heated at 65°C for 5 minutes, and cooled for 5 minutes at room temperature. The sample tube was then preheated to 37°C and 5 μl of selection buffer (or all the components in a total volume of 5 μl) was added. The sample was briefly vortexed and incubated for a given amount of time. In the case of velocity measurements, six time points were taken for each sample. The reaction was stopped by adding 1 μl of 0.5 M EDTA and briefly vortexing. After adding 11 μl

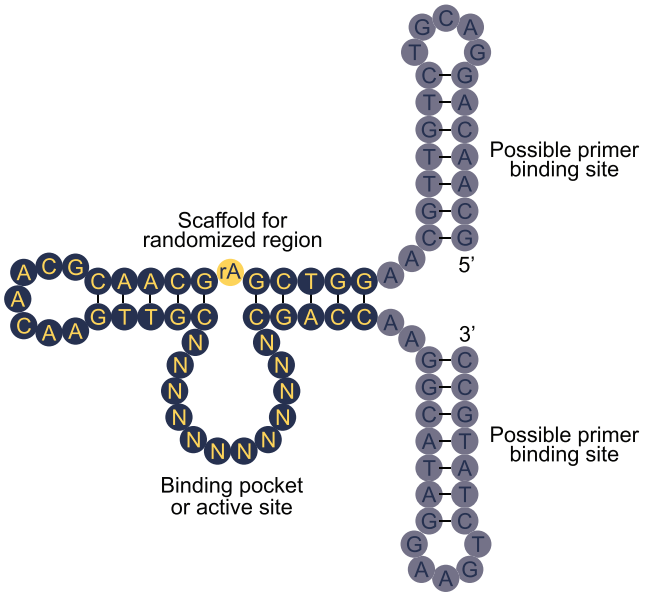

Figure 1. A DNA architecture for single-round selections. The ribonucleotide at the expected site of cleavage is indicated by rA. Bases shown in a pale background ('possible primer binding site') were deleted in most experiments (see also Table 1).

of 2× gel loading buffer (described above), the sample was heated to 70°C for 2 minutes and loaded on either a 7% (full-length constructs; 300 volts, 80 minutes) or 12.5% denaturing PAGE gel (300 volts, 130 minutes). When electrophoresis was finished, DNA was stained by GelRed (Biotium), scanned using a GE Typhoon FLA-9100 gel scanning device and analyzed using ImageQuant software. Rates were calculated from the initial phase of the reaction using the following equation, where F is the fraction reacted, k is the rate, and t is time.

$$F = 1 - e^{-kt} \quad (1)$$

Analysis of substrate sequence requirements

To compare the substrate requirements of our deoxyribozymes with that of the previously described 8–17 motif, variants of Dvanactka and 8–17 were prepared containing either a C–G or T.G pair flanking the cleavage site. These variants were then tested for activity over a range of pH values and lead concentrations using a 100 second time point. The initial velocity of the reaction was also measured for both Dvanactka and the 8–17 T.G construct under optimal conditions.

Analysis of base pairs using compensatory mutations

Approximately twenty variants of the Dvanactka motif (isolated from the N_{12} library in the presence of lead) were tested for the ability to cleave RNA. Most contained mutations at positions 1–7 or 2–6 of the randomized region, and were designed to obtain additional evidence for base pairs at these positions. Each variant was incubated at 37°C in the $+Pb^{2+}$ selection buffer for 1 hour, analyzed by PAGE, and the amount of RNA cleavage was determined.

RESULTS

The importance of bulges in functional nucleic acid structures

The secondary structures and sequence requirements of a variety of functional nucleic acid motifs have been characterized by multiple research groups using artificial evolution and comparative sequence analysis. In some cases, high-resolution structures have also been determined by NMR or X-ray crystallography. These motifs often consist of highly conserved bulges or loops flanked by stems containing Watson–Crick base pairs. Well-characterized examples of such bulged structures include aptamers that bind ATP, GTP, FMN and theophylline (9,10,22–27). A characteristic shared by some of these aptamers, especially those that form binding sites for small molecules, is that contacts with the ligand are made primarily by nucleotides in bulges rather than with those in stems. For example, in the NMR structure of an RNA aptamer bound to ATP, the primary role of the stems appears to be to stabilize the overall fold of the structure, while specific contacts with ATP are made by G7, G8, A10, G11, A12, G17 and G30, all of which occur in an asymmetric bulge (Supplementary Figure S1) (23–24). This suggested to us that, for at least some functional motifs with bulged structures, the presence of a stem is important but the sequence of the stem is not. To the extent to which this is true, one way to more readily access such motifs would be to place the randomized nucleotides in a library in the context of an arbitrarily chosen but stable scaffold which is present in all library members. Several studies have shown that the presence of a constant stem or patterned regions with the propensity to form hairpins can enhance the functional potential of libraries of 10^{15} random sequences in conventional selection experiments (30–31). In addition, a library containing 15 randomized nucleotides in the loop of a hairpin was previously used to identify thrombin aptamers in a single round of selection (20). However, the extent to which such architectures can be used in combination with small libraries to identify new catalytic nucleic acid motifs has not been investigated.

To explore this idea experimentally, we constructed a library in which 12 randomized positions were embedded in a bulge flanked by two 5 bp stems (Figure 1). This scaffold was flanked by hairpins at both the 5′ and 3′ ends that could function as primer binding sites. These were designed to be stable enough to minimize interactions with the random sequence region of the library, but not so stable that they would interfere with PCR amplification. Only 12 positions were randomized because the amount of diversity in such a library (1.7×10^7) corresponds to the number of reads that can be obtained in a standard high-throughput sequencing experiment. In the absence of an enrichment step, each sequence in the library would be expected to occur at a frequency of $\sim 10^{-7}$, and on the average would be expected to occur approximately once in a high-throughput sequencing experiment with $\sim 10^7$ reads. On the other hand, after enriching the library 100-fold for variants with a desired biochemical function (which can normally be achieved in several hours using standard separation methods such as affinity chromatography or PAGE), the most active sequences would be expected to be present at a frequency of $\sim 10^{-5}$, and therefore represented ~100 times more frequently in the sequencing data. These considerations suggested to us that, if catalytic motifs occur in a library of $\sim 10^7$ sequences, it should be possible to identify them in a single round of selection.

Identification of deoxyribozymes that cleave RNA in a single round of selection

As an initial test of this method, we investigated whether deoxyribozymes that cleave RNA could be isolated from our library in a single purification. We chose this activity in part for historical reasons: the ability to cleave RNA was the first catalytic activity of DNA to be identified (28). In addition, although a range of relatively simple RNA-cleaving deoxyribozymes have been reported, including the well-known 8–17 and 10–23 motifs (29), the bulge in our library was too small to contain canonical versions of these motifs. As was the case in the original selection for RNA-cleaving deoxyribozymes (28), our library contained a single RNA linkage (in this case, opposite to the twelve-nucleotide randomized bulge). Because the nonenzymatic rate of RNA cleavage is orders of magnitude higher than that of DNA under conditions similar to the ones used here (32), this RNA linkage was expected to be the primary site of cleavage in the library. The ribonucleotide (an adenosine) was flanked by G-C base pairs (Figure 1). This was expected to act as a further deterrent against known RNA-cleaving deoxyribozymes such as 8–17, which requires a G.T pair 3′ of the cleavage site, and 10–23, which cleaves purine-pyrimidine but not purine-purine junctions (29). Our buffer contained potassium, magnesium, and lead. These metal ions promote nucleic acid folding by interacting with negatively charged phosphate groups in the phosphodiester backbone, and magnesium and lead often also play catalytic roles. Lead is particularly effective at promoting RNA degradation in both enzymatic and nonenzymatic systems (16–18,28,33,34), and was used in the first selection for RNA-cleaving deoxyribozymes (28). After folding and incubating the library in this buffer for 30 minutes, cleaved molecules were separated from unreacted library members by PAGE, eluted from the gel, ethanol precipitated, and amplified by PCR (Figure 2). Both the unselected and selected pools were then analyzed by high-throughput sequencing. About half (42.5%) of the 1.7×10^7 possible sequences in the library were observed in the sequencing data from the starting library, and all had read numbers of 20 or less (Figure 3). In comparison, 2890 different sequences in the selected pool had read numbers exceeding 20 (Figure 3). Based on a comparison of the number of sequences at each read number in the selected and unselected library, we estimate that 51 511 sequences were enriched in our purification, although those at the low end of this distribution could not be distinguished from the large excess of catalytically inactive sequences (Figure 3). To determine whether these enriched sequences corresponded to catalytic motifs, several were synthesized (both with and without the primer binding sites shown in Figure 1) and tested for catalytic activity in selection buffer containing potassium, magnesium, and lead. About half of the variants we tested promoted RNA cleavage, with rates of up to 0.2 min^{-1} and rate enhancements of ~5000-fold (Table 1 and Figure 3 inset; note

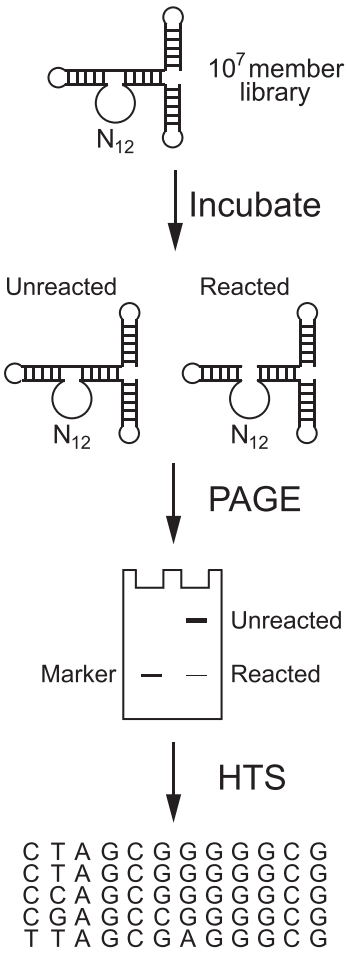

Figure 2. Single-round discovery of RNA-cleaving deoxyribozymes. PAGE, polyacrylamide gel electrophoresis; HTS = high-throughput sequencing.

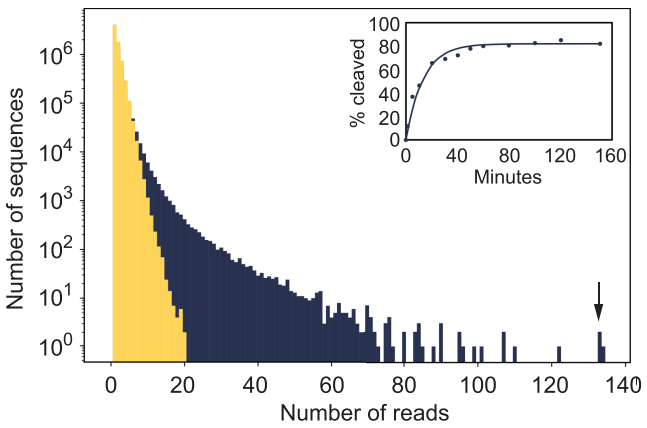

Figure 3. Identification of deoxyribozymes that cleave RNA in a single round of selection. Sequences from the unselected library are shown in yellow and those from the selected library in dark blue. The library contained 12 randomized positions and 1.7×10^7 different sequences, and was incubated in a buffer containing potassium, magnesium, and lead. A time course showing the catalytic activity of one of the sequences with the highest read number (indicated by an arrow in the graph) is shown in the inset.

that rate enhancements are defined as the rate of reaction of the deoxyribozyme divided by the rate of reaction of the starting library in the same buffer). This demonstrates that RNA-cleaving deoxyribozymes can be isolated from a $\sim 10^7$-member library in a single round of selection.

Identification of deoxyribozymes that cleave RNA under less favorable conditions

Encouraged by these results, we next investigated whether our approach could also be used to identify deoxyribozymes that cleave RNA under less favorable conditions. To address this question, we repeated our selection using a buffer that contained sodium and magnesium but not lead. Because the nonenzymatic rate of RNA cleavage is orders of magnitude slower in the absence of lead than in its presence, these conditions were expected to significantly increase the difficulty of finding new deoxyribozymes. Following a 30 minute incubation in this buffer, cleaved pool members were purified by PAGE and analyzed by high-throughput sequencing. As was the case for the selection in the presence of lead, the distribution of read numbers was significantly shifted in the selected library relative to the starting library (Figure 4A). However, the estimated number of enriched sequences (144) was 360-fold lower than that observed in the presence of lead (compare the dark blue and green distributions in Figure 4A). Rates and rate enhancements of deoxyribozymes isolated in the absence of lead were also lower: the most active variant we identified promoted RNA cleavage with a rate of 3.3×10^{-4} and a rate enhancement of $\sim$600-fold (Table 1). These experiments demonstrate the versatility of our approach by showing that it can be used to identify deoxyribozymes under conditions significantly less favorable than those used in the initial selection. They also highlight the important role that can be played by lead in deoxyribozyme-catalyzed RNA cleavage reactions.

Identification of deoxyribozymes that cleave RNA from larger libraries

The probability of finding at least one sequence in a random sequence library with a particular function is expected to increase as the number of distinct sequences in the library increases. Some types of motifs are also expected to occur more frequently in libraries containing longer stretches of randomized nucleotides (11, but also note the potential inhibitory effects of nonessential sequence discussed in reference 12). For these reasons, a library containing more randomized positions (and therefore more possible sequences) might be expected to contain a greater number of variants with a desired function. On the other hand, if the ratio of library size to read number is too high, sampling during high-throughput sequencing will not be sufficient to detect even highly enriched sequences. To better understand the tradeoff between library size and read number, a second library was synthesized that contained 15 randomized positions rather than the 12 positions present in the original library. This increased the theoretical diversity from 1.7×10^7 to 10^9 molecules, which was expected to increase the number of deoxyribozymes in the library but also to reduce sampling during high-throughput sequencing by a factor of 64.

Table 1. Properties of RNA-cleaving deoxyribozymes identified in single round selections. The sequence column indicates that of the randomized region. Rates represent the average (± standard deviation) of three independent measurements. Rates without brackets were measured for variants lacking primer binding sites (Figure 1) while rates in brackets are those of full-length variants. Rates were measured at 37°C. + Pb^{2+} buffer = 0.5 M KCl, 50 mM $MgCl_2$, 1 mM $PbCl_2$, 50 mM HEPES, pH 7.0 (or 50 mM Tris, pH 7.5 in the opt buffer). - Pb^{2+} buffer = 1 M NaCl, 10 mM $MgCl_2$, 50 mM Tris–HCl, pH 7.5

Library	Buffer	Sequence	Rank (reads)	Motif	k_{obs} (min^{-1})
N_{12}	+ Pb^{2+}	NNNNNNNNNNNN		Initial Pool	$(2 \pm 1) \times 10^{-4}$ [$(7 \pm 3) \times 10^{-6}$]
N_{15}	+ Pb^{2+}	NNNNNNNNNNNNNNN		Initial Pool	$(8 \pm 3) \times 10^{-5}$
N_{12}	- Pb^{2+}	NNNNNNNNNNNN		Initial Pool	$(5 \pm 3) \times 10^{-7}$ [$(10 \pm 7) \times 10^{-7}$]
N_{15}	- Pb^{2+}	NNNNNNNNNNNNNNN		Initial Pool	$(4 \pm 2) \times 10^{-7}$
N_{12}	+ Pb^{2+}	CTAGCGGGGGCG	2 (133)	Dvanactka	0.18 ± 0.01 [0.038 ± 0.002]
N_{12}	+ Pb^{2+}, optimized	CTAGCGGGGGCG	2 (133)	Dvanactka	0.23 ± 0.03
N_{12}	+ Pb^{2+}	ATTAGAGGGCGT	5 (110)		0.056 ± 0.008
N_{15}	+ Pb^{2+}	AGTTGGGTGGGCAGG	1 (16)		0.12 ± 0.07
N_{15}	+ Pb^{2+}	ATGTGTTGTGGGAGT	2 (14)		0.07 ± 0.02
N_{15}	+ Pb^{2+}	TAGTTAGCGACGGCG	2 (14)		0.10 ± 0.03
N_{12}	- Pb^{2+}	CCAGCCGGGCGT	2 (26)		$(3.3 \pm 0.7) \times 10^{-4}$
N_{12}	- Pb^{2+}	TTATCGGGTATG	3 (23)		$(1.3 \pm 0.7) \times 10^{-4}$
Reference	+ Pb^{2+}	CCGAGCCGGACGA		8–17 GT	0.06 ± 0.01
Reference	+ Pb^{2+}, optimized	CCGAGCCGGACGA		8–17 GT	0.10 ± 0.02
Reference	- Pb^{2+}	CCGAGCCGGACGA		8–17 GT	$(1.2 \pm 0.2) \times 10^{-4}$

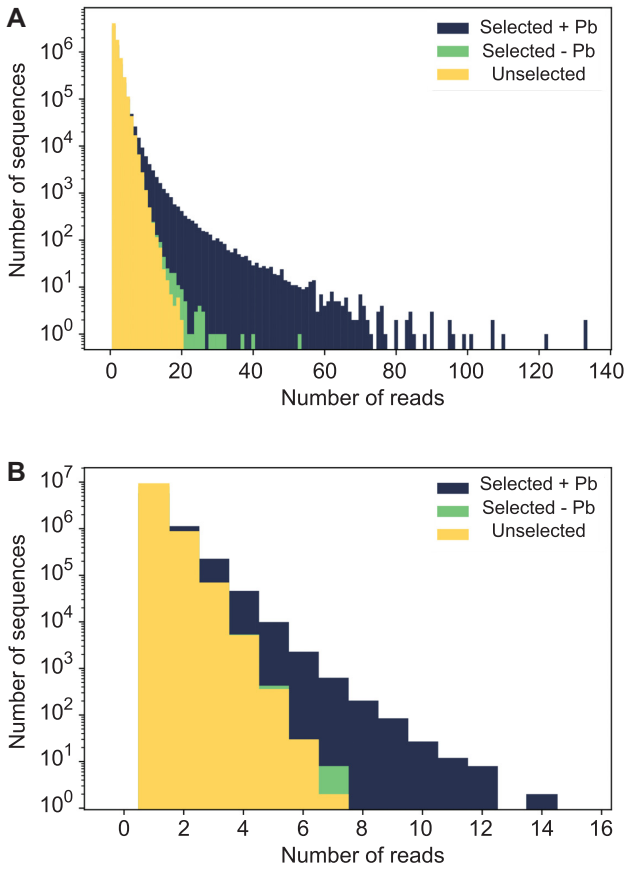

Figure 4. Effect of metal ions and library size on the number of enriched sequences in selections for RNA-cleaving deoxyribozymes. (**A**) Distribution of read numbers in an unselected N_{12} library, in an N_{12} library selected for the ability to cleave RNA in the presence of lead, and in an N_{12} library selected for the ability to cleave RNA in the absence of lead. (**B**) Same, but for an N_{15} library.

As before, the library was incubated in a buffer containing either potassium, magnesium, and lead or sodium and magnesium (but not lead). Following a 30 minute incubation, cleaved pool members were purified by PAGE and analyzed by high-throughput sequencing. As expected, the number of sequences with different read numbers in the unselected and selected populations were more similar to one another for these selections than they were for analogous selections using the N_{12} library (compare panels A and B of Figure 4). However, at least in the case of the selection performed in the presence of lead, the distribution was clearly shifted (Figure 4B). For example, 339 sequences in the selected pool had a read number ≥ 8, while none did in the unselected pool. In contrast, in the case of the selection in the absence of lead, the unselected and selected N_{15} libraries were almost indistinguishable, although a modest enrichment was seen at the highest read number detected (Figure 4B). The total number of enriched sequences was 2200-fold higher in the presence of lead than in its absence, providing additional evidence for the important role that can be played by lead in deoxyribozyme-catalyzed RNA cleavage reactions. As before, several of the most enriched sequences were confirmed to be catalytically active (Table 1), indicating that our approach can be used to isolate deoxyribozymes from libraries containing up to fifteen randomized position. A limitation of using an N_{15} pool for one-step selections is that the coverage by high-throughput sequencing is insufficient to obtain information about all catalytic sequences in the library (at least for the methods of purification used here). On the other hand, such a library makes it possible to identify motifs that cannot occur in shorter randomized regions.

Rapid identification of the sequence requirements of catalytic motifs

Deoxyribozymes identified in conventional selection experiments normally contain a conserved catalytic core flanked by nonessential sequence. A second selection experiment is typically performed to identify the catalytic core. This involves making a library by randomly mutagenizing the de-

oxyribozyme, isolating active variants by *in vitro* selection, identifying conserved parts of the motif by comparative analysis, and confirming the model by mutagenesis (3,35–37). Even when high-throughput sequencing is used to analyze the results (38), only a small fraction of sequence space is typically explored. In the case of selections performed with N_{12} libraries, however, such experiments are in principle unnecessary: each of the $\sim 10^7$ possible variants of the sequence are expected to be present in the initial library, and coverage from high-throughput sequencing should be sufficient to detect most enriched variants. On the other hand, because only a single round of selection has been performed, many inactive sequences will still be present in the library, and can potentially obscure the signal from catalytic sequences with the fold of interest. To minimize this problem, we focused our analysis on clusters of sequences one and two mutations away from those with high read numbers. Such variants occurred significantly more frequently for sequences with high read numbers than they did for randomly chosen controls (Supplementary Figures S2 and S3), suggesting that they can provide information about the sequence requirements of catalytic motifs without introducing noise from sequences with unrelated folds.

In the first step of our analysis (Supplementary Figure S4), read numbers of single-mutation variants were used to construct sequence logos showing conservation at each position in sequences of interest. Analysis of these sequence logos revealed a variety of motifs in the selected libraries (Supplementary Figures S5-S7). Some of these motifs were catalytically inactive, such as a G-rich element observed in each selection (Supplementary Figures S5–S7), but others promoted RNA-cleavage reactions (Figure 5 and Table 1). In the second step of our analysis (Supplementary Figure S4), we searched for double mutants of each sequence of interest which occurred more often than expected based on the read numbers of the corresponding single mutants. Such covariations are a powerful way to identify base pairs in the secondary structures of functional DNA and RNA motifs (3,35–37). Although not all covariations identified using this approach could be rationalized, some were consistent with base pairing. For example, in the case of Hit 2 (the sequence from the N_{12} lead selection with the second highest read number, which we named Dvanactka after the Czech word for twelve), a correlation between positions 1 and 7 suggested that these nucleotides form a C–G base pair (Supplementary Figure 6A–C). Although variants of Dvanactka in which this base pair was changed to T.G or CA had lower read numbers, the read number of the T-A mutant (which combined these two deleterious mutations) was similar to that of the original C–G variant (Figure 6D). A similar pattern was observed when these variants were tested for catalytic activity (Figure 6E). Effects of other mutations in Dvanactka were also correlated with read number (Figure 6F and Supplementary Figure S8), indicating that, in the case of this motif, read numbers provide information about how mutations affect catalytic activity. Our data also provided comprehensive information about the sequence requirements of the helix in Dvanactka. This showed that, of the 36 possible combinations of pairs in the helix, five were preferred over the others (Supplementary Figure S9).

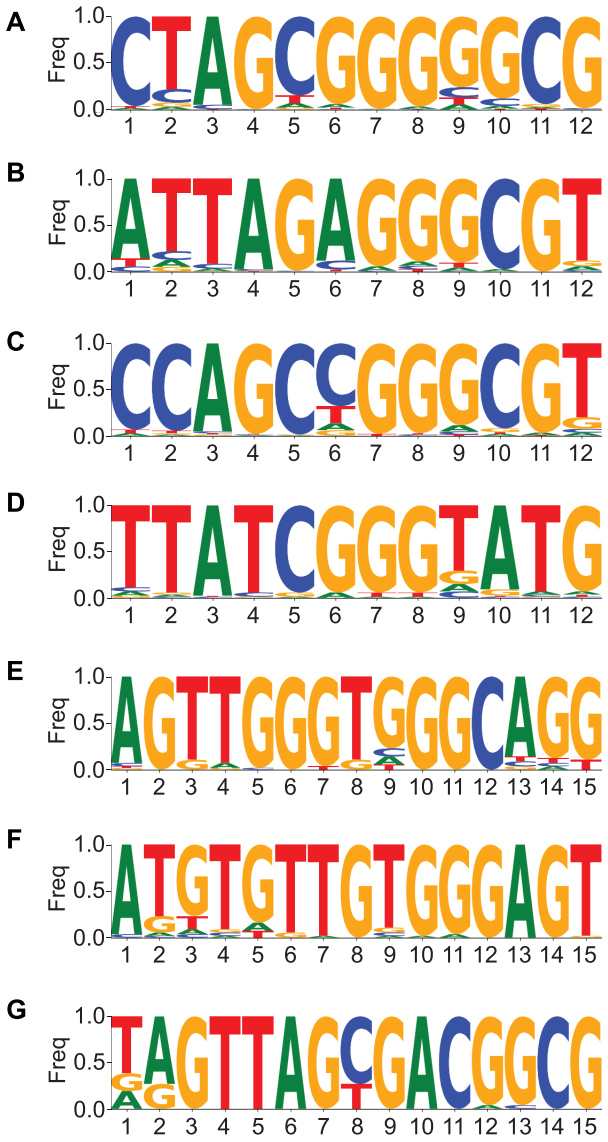

Figure 5. Diversity of catalytic motifs isolated in single-round selections. Sequence logos were derived from the read numbers of all single-mutation variants of deoxyribozymes isolated in single-round selections. Panels **A** and **B** show motifs isolated from the N_{12} library incubated in the presence of lead. Panels **C** and **D** show motifs from the N_{12} library incubated in the absence of lead. Panels **E**-**G** show motifs from the N_{15} library incubated in the presence of lead. See Table 1 for the catalytic rates of these deoxyribozymes.

Further analysis revealed that a number of other sequences with high read numbers form secondary structures similar to that of Dvanactka. For example, of the 50 most abundant sequences in the N_{12} lead library, at least 10 form this structure (Supplementary Figure S10). For most of these sequences, covariations were observed in both base pairs in the helix (Supplementary Figure S10), providing strong evidence for these secondary structure models. Weak correlations were also observed in some heat maps but not others, such as between positions 1 and 11 and between positions 4 and 11 (Supplementary Figure S10). These corre-

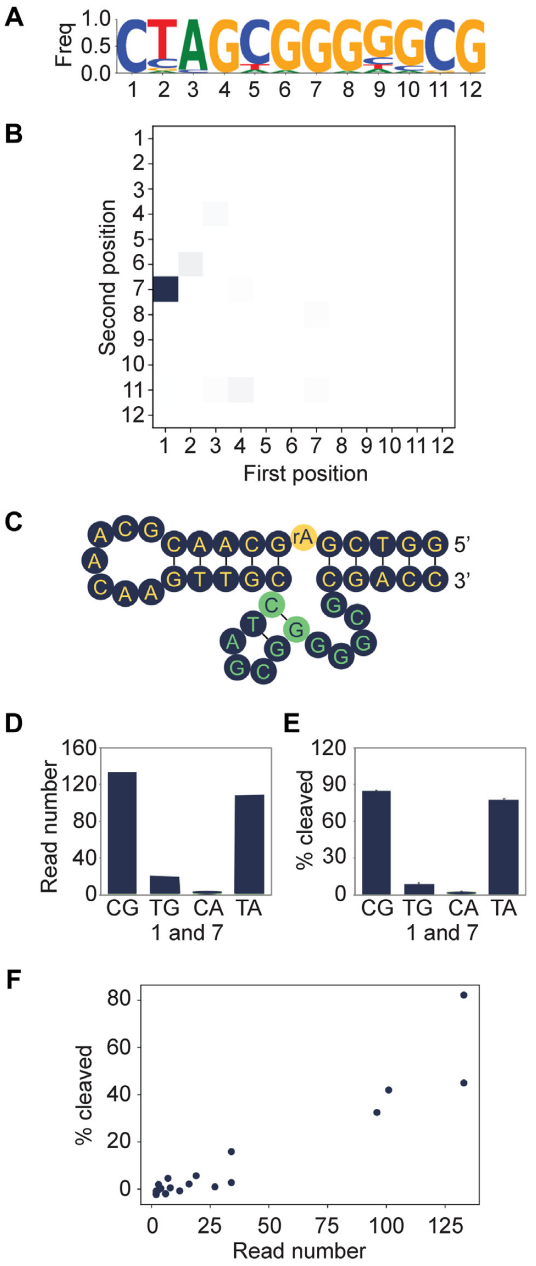

Figure 6. Rapid characterization of deoxyribozyme sequence requirements. (**A**) Sequence logo generated from read numbers of all possible single mutant variants of Dvanactka. (**B**) Pairwise correlations based on read numbers of single and double mutants of Dvanactka. Blue squares indicate pairs at which a double mutant occurred more often than expected based on the read numbers of the corresponding single mutants. (**C**) Secondary structure model of Dvanactka. Nucleotides from the randomized region are shown in green, and positions 1 and 7 with a green background. (**D**, **E**) Evidence for base pairing between positions 1 and 7 based on read numbers (panel D) and catalytic activity (panel E). Error bars indicate the standard deviation from three experiments. (**F**) Correlation between read number and catalytic activity for variants of Dvanactka. $R^2 = 0.72$. See Supplementary Table S3 for more information about these sequences.

lations likely reflect mutations that can only occur in some sequence backgrounds.

Although our dataset provided considerable information about the secondary structure of members of the Dvanactka family, some sequences with high read numbers did not exhibit covariation patterns consistent with base pairing. This could indicate that, like other RNA-cleaving deoxyribozymes such as 10.23 (29), the bulges in these motifs do not form canonical base pairs. Alternatively, these motifs could contain invariant base pairs that cannot be detected by covariation analysis. It is also possible that the random sequence regions in some motifs form base pairs with constant sequences in the scaffold. This does not appear to be the case for Dvanactka, however: inserting its N_{12} region into a different scaffold had no effect on catalytic activity (Supplementary Figure S11). Taken together, this analysis highlights the surprising catalytic diversity of RNA-cleaving deoxyribozymes in pools containing 10^7–10^9 different sequences. It also indicates that, in addition to facilitating discovery of new deoxyribozymes, our approach can also be used to rapidly elucidate their sequence requirements (and in some cases their secondary structures as well).

A variant of the 8–17 motif with a distinct substrate specificity

One of the most well-characterized deoxyribozymes is the RNA- cleaving 8–17 motif (39). It has been identified independently by at least three different groups using a wide range of conditions (29,40–42, and perhaps 28), and is one of the few deoxyribozymes for which a high-resolution structure is available (43). Dvanactka appears to be related to the 8–17 deoxyribozyme. Like 8–17, it contains an invariant AG motif in the loop of a short stem followed by an invariant CG motif (29,39,44) (Supplementary Figure S12). In addition, the correlation between positions 4 and 11 we observed in some variants of Dvanactka (Supplementary Figure S10) is consistent with the presence of a pseudoknot, which was also observed in the crystal structure of the 8–17 deoxyribozyme (43). However, Dvanactka differs from canonical versions of 8–17 in several ways. Most importantly, it contains a G-C base pair adjacent to the cleavage site rather than the G.T pair typically present in 8–17 (29,39,44) (Supplementary Figure S12). This feature was imposed by the architecture of our library, as the G–C pair was in a constant part of the scaffold and therefore occurred in all library members. Replacing the G.T pair with a G–C pair has been reported to reduce the catalytic activity of 8–17 to undetectable levels (39), which makes its presence in an 8–17 like motif surprising. To better understand the consequences of this change, we compared the ability of Dvanactka and a canonical variant of 8–17 to cleave substrates containing either a G–C or a G.T pair at the cleavage site over a range of lead concentrations and pH values. Dvanactka could only cleave the substrate with the G-C base pair, while 8–17 (as previously reported) could only cleave the substrate with the G.T pair (Supplementary Figure S13). Thus, in both deoxyribozymes the identity of the base pair 3′ of the cleavage site is critical, although the substrate specificities of these two motifs are orthogonal. When tested side-by-side in the same scaffold, the k_{obs} of Dvanactka using a

G-C substrate was about 2 times higher than that of 8–17 using a G.T substrate (Table 1 and Supplementary Figures S12-S13). Another difference is that Dvanactka contains a truncated stem with only two base pairs rather than the three which occur in 8–17 (29,39,44) (Supplementary Figure S12). Because 8–17 contains thirteen or fourteen positions in the bulge, without this truncation an 8–17 like motif could not have occurred in our N_{12} library. Although unusually short, this stem was well-supported by covariations between positions 1 and 7 and between 2 and 6 (Figure 6 and Supplementary Figure S10). Furthermore, the length of this stem appears to be functionally important: extending it to three base pairs significantly reduced catalytic activity (Supplementary Figure S14). The sequence context of the invariant CG dinucleotide at the 3′ end of the motif also differed in Dvanactka and 8–17. Instead of being flanked by W (A or T) at the 5′ end and R (A or G) at the 3′ end (29,39,44), the CG motif in Dvanactka typically occurred in contexts such as GGGCG and GCGT. It is possible that these sequence elements help to stabilize the unusually short 2 bp stem in Dvanactka or enable it to cleave substrates with G–C base pairs flanking the cleavage site. Despite these important differences, however, we hypothesize that the three-dimensional architecture of our motif is likely to be similar to that of 8–17 (43).

DISCUSSION

Here, we describe a new method to identify and characterize simple nucleic acid catalysts which is general, easy to implement, and significantly faster than conventional protocols. One key aspect of this method is the use of high-throughput sequencing technologies (21). These were not available when artificial evolution methods for nucleic acids were first developed, but have since been used by a number of groups to better characterize the sequence requirements of functional motifs (38,45–50). A second is the use of a DNA architecture which makes it possible to isolate catalytic motifs from relatively small libraries containing only 10^7–10^9 unique sequences. This architecture was inspired primarily by the structures of known aptamers, which in some cases are made up of binding pockets formed by short bulges or loops stabilized by stems which do not interact with the ligand (22–27). By embedding a relatively small number of randomized positions into such an architecture, we were able to identify an active site that promoted RNA cleavage by a factor of ∼5000 in a single round of selection. Our protocol also made it possible to elucidate the sequence requirements and secondary structures of some of these motifs without the need to perform additional experiments.

Several previous studies have investigated the possibility of isolating new aptamers in a single round of selection. This requires finding a way to distinguish aptamers from the enormous excess of inactive sequences in the library. One way this can be accomplished is to use specialized experimental setups that significantly increase the amount of enrichment that can be achieved in a single purification (51–53). A different approach is to use small libraries in combination with high-throughput sequencing (20,54). If the library size is sufficiently small (for example, 10^9), enriched sequences can be detected based on read number (20). If it is not (e.g. 10^{12}), enriched motifs corresponding to short conserved regions of aptamers can sometimes be identified (54). In the work reported here, we have shown that new catalytic motifs can also be identified in single round of selection. Our method utilizes structured libraries containing 10^7–10^9 sequences in combination with high-throughput sequencing. One significant advantage of this approach is that it is compatible with virtually any selection protocol, including those that use standard methods of purification such as PAGE and affinity chromatography. A second is that it does not require expensive equipment. And a third is it enables rapid characterization of the sequence requirements and secondary structures of newly discovered deoxyribozymes without the need for additional experiments. A potential disadvantage, however, is that it requires libraries to be 10^6 to 10^8 orders of magnitude less complex than those typically used in artificial evolution experiments (although this depends on both the enrichment that can be achieved in a single purification and the number of reads that can be obtained by high-throughput sequencing). For this reason, it might be expected to yield motifs that are less active. But how much less active? In the case of our selections that used an N_{12} pool with ∼10^7 different sequences, rates of deoxyribozymes isolated in the presence of lead were 8-fold lower than those of deoxyribozymes isolated under similar conditions from an N_{50} pool containing 10^{14} different sequences (28). In the case of deoxyribozymes isolated in the absence of lead, rates were 6 to 30-fold lower than those of deoxyribozymes isolated under similar conditions from pools containing 10^{13} to 10^{14} sequences (29,55). Similar values are obtained when estimated rate enhancements are compared. Although these conclusions only apply to one catalytic function, they indicate that the disadvantage of using a less diverse pool can be surprisingly small.

Another important question is the generality of our approach. Extremely simple RNA motifs that promote reactions such as RNA self-aminoacylation (13) and RNA cleavage (14–18) have been described, and such motifs are likely to be accessible using our method. A range of DNA and RNA aptamers containing bulges of twelve to fifteen nucleotides have also been reported (22–27), suggesting that this approach could also be well-suited for the development of nucleic acids that bind specific ligands (see references (20) and (54) for similar approaches applied to aptamers). This could be especially important for applications that require development of multiple aptamers. We note, however, that catalytic motifs such as the Class I ligase ribozyme (35) have structures that are far too complicated to identify in pools containing 12–15 randomized positions, although it is possible that simpler ligases could be generated using our approach. Identification of such complex structures will still require the use of traditional approaches and more diverse random sequence pools, although their identification and characterization can still be facilitated by high-throughput sequencing (38,45–50), perhaps in combination with novel methods of library construction such as secondary structure libraries (56).

In conclusion, we show that it is possible to identify RNA-cleaving deoxyribozymes in a single round of selection using standard purification methods. Our approach

uses structured libraries containing twelve to fifteen randomized positions, although it is possible that larger libraries can be used in combination with more powerful sequencing methods. Deoxyribozymes are identified based on their read numbers in high-throughput sequencing experiments. In addition to facilitating identification of catalytic sequences, this approach also provides information about their sequence requirements (and in some cases their secondary structures). Due to its simplicity relative to conventional methods, we anticipate that this approach will accelerate the discovery of new catalytic DNA and RNA motifs, especially those with simple structures.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank colleagues in our group and at the IOCB for useful discussions.

FUNDING

Czech Science Foundation (GACR) [19-20989S]; 'Chemical biology for drugging undruggable targets (ChemBioDrug)' [CZ.02.1.01/0.0/0.0/16_019/0000729] from the European Regional Development Fund (OP RDE). Funding for open access charge: European Regional Development Fund (OP RDE) [CZ.02.1.01/0.0/0.0/16_019/0000729].
Conflict of interest statement. None declared.

REFERENCES

1. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
2. Robertson,D.L. and Joyce,G.F. (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, **344**, 467–468.
3. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
4. Wilson,D.S. and Szostak,J.W. (1999) In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.*, **68**, 611–647.
5. Bartel,D.P. and Unrau,P.J. (1999) Constructing an RNA world. *Trends Cell Biol.*, **9**, M9–M13.
6. Joyce,G.F. (2004) Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.*, **73**, 791–836.
7. Silverman,S.K. (2016) Catalytic DNA: scope, applications, and biochemistry of deoxyribozymes. *Trends Biochem. Sci.*, **41**, 595–609.
8. Zhou,J. and Rossi,J. (2017) Aptamers as targeted therapeutics: current potential and challenges. *Nat. Rev. Drug Discov.*, **16**, 181–202.
9. Carothers,J.M., Oestreich,S.C., Davis,J.H. and Szostak,J.W. (2004) Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.*, **126**, 5130–5137.
10. Huizenga,D.E. and Szostak,J.W. (1995) A DNA aptamer that binds adenosine and ATP. *Biochemistry*, **34**, 656–665.
11. Knight,R. and Yarus,M. (2003) Finding specific RNA motifs: function in a zeptomole world? *RNA*, **9**, 218–230.
12. 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.
13. Turk,R.M., Chumachenko,N.V. and Yarus,M. (2010) Multiple translation products from a five-nucleotide ribozyme. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 4585–4589.
14. Dange,V., Van Atta,R.B. and Hecht,S.M. (1990) A Mn^{2+}-dependent ribozyme. *Science*, **248**, 585–588.
15. Kolev,N.G., Hartland,E.I. and Huber,P.W. (2008) A manganese-dependent ribozyme in the 3′-untranslated region of Xenopus Vg1 mRNA. *Nucleic Acids Res.*, **36**, 5530–5539.
16. Pan,T. and Uhlenbeck,O.C. (1992) A small metalloribozyme with a two-step mechanism. *Nature*, **358**, 560–563.
17. Hoogstraten,C.G., Legault,P. and Pardi,A. (1998) NMR solution structure of the lead-dependent ribozyme: evidence for dynamics in RNA catalysis. *J. Mol. Biol.*, **284**, 337–350.
18. Wedekind,J.E. and McKay,D.B. (1999) Crystal structure of a lead-dependent ribozyme revealing metal binding sites relevant to catalysis. *Nat. Struct. Biol.*, **6**, 261–268.
19. Curtis,E.A. and Liu,D.R. (2013) Discovery of widespread GTP-binding motifs in genomic RNA and DNA. *Chem. Biol.*, **20**, 521–532.
20. Kupakuwana,G.V., Crill,J.E. II, McPike,M.P. and Borer,P.N. (2011) Acyclic identification of aptamers for human alpha-thrombin using over-represented libraries and deep sequencing. *PLoS One*, **6**, e19395.
21. Goodwin,S., McPherson,J.D. and McCombie,W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.*, **17**, 333–351.
22. Patel,D.J. and Suri,A.K. (2000) Structure, recognition and discrimination in RNA aptamer complexes with cofactors, amino acids, drugs and aminoglycoside antibiotics. *J. Biotechnol.*, **74**, 39–60.
23. Jiang,F., Kumar,R.A., Jones,R.A. and Patel,D.J. (1996) Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature*, **382**, 183–186.
24. Dieckmann,T., Suzuki,E., Nakamura,G.K. and Feigon,J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA*, **2**, 628–640.
25. Yang,Y., Kochoyan,M., Burgstaller,P., Westhof,E. and Famulok,M. (1996) Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science*, **272**, 1343–1347.
26. Carothers,J.M., Davis,J.H., Chou,J.J. and Szostak,J.W. (2006) Solution structure of an informationally complex high-affinity RNA aptamer to GTP. *RNA*, **12**, 567–579.
27. Gelinas,A.D., Davies,D.R. and Janjic,N. (2016) Embracing proteins: structural themes in aptamer-protein complexes. *Curr. Opin. Struct. Biol.*, **36**, 122–132.
28. Breaker,R.R. and Joyce,G.F. (1994) A DNA enzyme that cleaves RNA. *Chem. Biol.*, **1**, 223–229.
29. Santoro,S.W. and Joyce,G.F. (1997) A general-purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4262–4266.
30. Davis,J.H. and Szostak,J.W. (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11616–11621.
31. Ruff,K.M., Snyder,T.M. and Liu,D.R. (2010) Enhanced functional potential of nucleic acid aptamer libraries patterned to increase secondary structure. *J. Am. Chem. Soc.*, **132**, 9453–9464.
32. Li,Y. and Breaker,R.R. (2004) Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-hydroxyl group. *J. Am. Chem. Soc.*, **121**, 5364–5372.
33. Palou-Mir,J., Barcelo-Oliver,M. and Sigel,R.K.O. (2017) The role of lead(II) in nucleic acids. *Met. Ions Life Sci.*, **17**, 403–434.
34. Brown,R.S., Hingerty,B.E., Dewan,J.C. and Klug,A. (1983) Pb(II)-catalysed cleavage of the sugar-phosphate backbone of yeast tRNAPhe - implications for lead toxicity and self-splicing RNA. *Nature*, **303**, 543–546.
35. Ekland,E.H. and Bartel,D.P. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.*, **23**, 3231–3238.
36. Curtis,E.A. and Bartel,D.P. (2013) New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.*, **12**, 994–1000.
37. 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.
38. Pitt,J.N. and Ferré-D'Amaré,A.R. (2010) Rapid construction of empirical RNA fitness landscapes. *Science*, **330**, 376–379.
39. Cepeda-Plaza,M. and Peracchi,A. (2020) Insights into DNA catalysis from structural and functional studies of the 8–17 DNAzyme. *Org. Biomol. Chem.*, **18**, 1697–1709.
40. Li,J., Zheng,W., Kwon,A.H. and Lu,Y. (2000) In vitro selection and characterization of a highly efficient Zn(II)-dependent RNA-cleaving deoxyribozyme. *Nucleic Acids Res.*, **28**, 481–488.

41. Peracchi,A. (2000) Preferential activation of the 8–17 deoxyribozyme by Ca(2+) ions. Evidence for the identify of 8–17 with the catalytic domain of the Mg5 deoxyribozyme. *J. Biol. Chem.*, **275**, 11693–11697.
42. Faulhammer,D. and Famulok,M. (1996) The Ca^{2+} ion as a cofactor for a novel RNA-cleaving deoxyribozyme. *Angew. Chem. Int. Ed.*, **35**, 2837–2841.
43. Liu,H., Yu,X., Chen,Y., Zhang,J., Wu,B., Zheng,L., Haruehanroengra,P., Wang,R., Li,S., Lin,J. *et al.* (2017) Crystal structure of an RNA-cleaving DNAzyme. *Nat. Commun.*, **8**, 2006.
44. Cruz,R.P.G., Withers,J.B. and Li,Y. (2004) Dinucleotide junction cleavage versatility of 8–17 deoxyribozyme. *Chem. Biol.*, **11**, 57–67.
45. Gotrik,M.R., Feagin,T.A., Csordas,A.T., Nakamoto,M.A. and Soh,H.T. (2016) Advancements in aptamer discovery technologies. *Acc. Chem. Res.*, **49**, 1903–1910.
46. Kobori,S. and Yokobayashi,Y. (2016) High-throughput mutational analysis of a twister ribozyme. *Angew. Chem. Int. Ed.*, **55**, 10354–10357.
47. Kobori,S., Takahashi,K. and Yokobayashi,Y. (2017) Deep sequencing analysis of aptazyme variants based on a pistol ribozyme. *ACS Synth. Biol.*, **6**, 1283–1288.
48. Dhamodharan,V., Kobori,S. and Yokobayashi,Y. (2017) Large scale mutational and kinetic analysis of a self-hydrolyzing deoxyribozyme. *ACS Chem. Biol.*, **12**, 2940–2945.
49. Blanco,C., Janzen,E., Pressman,A., Saha,R. and Chen,I.A. (2019) Molecular fitness landscapes from high-coverage sequence profiling. *Annu. Rev. Biophys.*, **48**, 1–18.
50. Yokobayashi,Y. (2020) High-throughput analysis and engineering of ribozymes and deoxyribozymes by sequencing. *Acc. Chem. Res.*, **53**, 2903–2912.
51. Berezovski,M., Musheev,M., Drabovich,A. and Krylov,S.N. (2006) Non-SELEX selection of aptamers. *J. Am. Chem. Soc.*, **128**, 1410–1411.
52. Nitsche,A., Kurth,A., Dunkhorst,A., Pänke,O., Sielaff,H., Junge,W., Muth,D., Scheller,F., Stöcklein,W., Dahmen,C. *et al.* (2007) One-step selection of Vaccinia virus-binding DNA aptamers by MonoLEX. *BMC Biotechnol.*, **7**, 48.
53. Lou,X., Qian,J., Xiao,Y., Viel,L., Gerdon,A.E., Lagally,E.T., Atzberger,P., Tarasow,T.M., Heeger,A.J. and Soh,H.T. (2009) Micromagnetic selection of aptamers in microfluidic channels. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 2989–2994.
54. Hoon,S., Zhou,B., Janda,K.D., Brenner,S. and Scolnick,J. (2011) Aptamer selection by high-throughput sequencing and informatic analysis. *BioTechniques*, **51**, 413–416.
55. Breaker,R.R. and Joyce,G.F. (1995). A DNA enzyme with Mg^{2+}-dependent RNA phosphoesterase activity. *Chem. Biol.*, **2**, 655–660.
56. Sgallov´,R. and Curtis,E.A. (2021) Secondary structure libraries for artificial evolution experiments. *Molecules*, **26**, 1671.