Nucleic Acids Research, 2025, **53**, gkaf307
https://doi.org/10.1093/nar/gkaf307
Methods

Rapid discovery of functional RNA domains

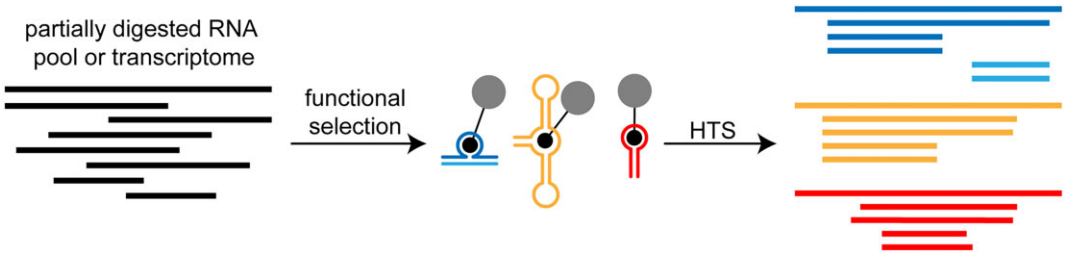
Brandon Latifi [1], Kyle H. Cole [2], Michael M.K. Vu [3], Andrej Lupták [1,2,3,*]

[1]Department of Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92697, United States
[2]Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA 92697, United States
[3]Department of Chemistry, University of California, Irvine, Irvine, CA 92697, United States
[*]To whom correspondence should be addressed. Email: aluptak@uci.edu

Abstract

Many strategies have been implemented to enrich an RNA population for a selectable function, but demarcation of the optimal functional motifs or minimal structures within longer libraries remains a lengthy and tedious process. To overcome this problem, we have developed a technique that isolates minimal active segments from complex heterogeneous pools of RNAs. This method allows for truncations to occur at both 5′ and 3′ ends of functional domains and introduces independent primer-binding sequences, thereby removing sequence and structure bias introduced by constant-sequence regions. We show examples of minimization for genomic and synthetic aptamers and demonstrate that the method can directly reveal an active RNA assembled from multiple strands, facilitating the development of heterodimeric structures used in cellular sensors. This approach provides a pipeline to experimentally define the boundaries of active domains and accelerate the discovery of functional RNAs.

Graphical abstract



Introduction

The versatility of RNAs as functional macromolecules endows them with a broad range of activities, including ligand binding and catalysis. One approach for discovering and engineering RNAs for a specific function is based on *in vitro* selection experiments, which identify RNAs with a specific biochemical activity (as aptamers, ribozymes, and more complex RNAs, such as riboswitches) [1–4]. However, the entire sequence of a functional RNA isolated from *in vitro* selection may not participate in the desired function, especially in selections using longer libraries, which are more likely to yield RNAs with complex folds [5]. Demarcation of the boundaries of active domains becomes a significant challenge during the discovery of novel functional motifs because identification of the minimal functional motif is typically achieved by manual removal of various segments [6] followed by biochemical validation of function. Rational truncations of an RNA usually require a model of the secondary structure, often aided by sequence-motif discovery, phylogenetic and sequence covariation analysis, and chemical probing. Identification of the minimal motif by this process can be tedious and especially elusive when secondary structure data are unavailable or difficult to decipher, as in the case of tertiary structures that dominate the overall fold (e.g. noncanonical quadruplex structures), and often difficult to apply to an entire *in vitro* selected pool of sequences.

Previous strategies to identify the minimum functional domains have relied on recombination events of its DNA templates or biochemical assessment of single sequences [6–9]. This was further showcased by reselection and design of a degenerate pool based on a ligase ribozyme to uncover its minimal functional domain for activity. Additionally, identification of a ribozyme core and nonessential sequences (internal deletions) has been shown to effectively identify the catalytic core of a ribozyme, but this approach was based on a single sequence and not a heterogeneous pool [10, 11]. Recently, a combinatorial approach to identify truncated ribozymes was developed [12], but this method only allowed for 3′ truncations to occur while retaining the 5′ constant region from the original selection pool. Similarly, DNA recombination strategies have been limited to pools derived from single sequences, reducing the throughput of discovery of functional sequences. This limitation can be mitigated using *in vitro* selection

Received: May 23, 2024. **Revised:** March 25, 2025. **Editorial Decision:** March 31, 2025. **Accepted:** April 4, 2025

platforms, which can accommodate much higher diversity, with a practical limit of $\sim 10^{16}$ molecules [13]. A pool of sequences enriched for functional RNAs can be further diversified and reselected to characterize multiple sequences at once; however, selections based on genomic or transcriptomic sequences are not amenable to demarcation of minimal motifs by recombination because recombined sequences would likely lose their biological relevance.

To mitigate these issues, we developed a high-throughput pipeline that experimentally demarcates functional RNA domains within a diverse heterogeneous pool and eliminates the need to individually test single truncated RNAs. This method is particularly powerful for revealing functional domains in genomic SELEX [14] because it removes artificially introduced primer-binding sequences. Using this pipeline, we not only identify the minimum functional domain(s) necessary for binding, but also establish whether constant-sequence regions are necessary for the formation of the active domain. With this approach, we showcase the discovery of functional domains in aptamers and a riboswitch. We experimentally minimized the human *FGD3* adenosine-binding aptamer, identified its minimized functional domain, and directly revealed how the aptamer can be assembled into a functional heterodimer from nonfunctional strands. We also identified the aptamer domain of the Werewolf-1 (Were-1) photoriboswitch and discovered a new Ni-NTA-binding aptamer. We anticipate that this method will increase the throughput of discovery of new functional RNAs domains and facilitate their engineering into downstream applications (e.g. aptazymes, riboswitches, and biosensors) [15–20].

Materials and methods

Synthesis of DNA oligonucleotides

All DNA sequences with or without modification(s) were purchased from Integrated DNA technologies (IDT). All oligonucleotide sequences are listed in Supplementary Table S1.

General information about the *in vitro* selection pools

As previously described, the selections for the ATP [21, 22], Were-1 [22], and Ni-NTA [23] were used to identify the functional domains for this manuscript. Rounds 13, 10, and 12 for the ATP, Were-1, and Ni-NTA selections, respectively, were used to validate this pipeline. For all the minimized aptamers presented in this manuscript, only one round of selection was performed when using the T4 splint hairpin approach. For the CircLigase approach, four rounds of selection with the ATP aptamer were performed after base hydrolysis. The Were-1 selection was performed using Dynabeads MyOne Carboxylic Acid beads (Thermo Fisher Scientific) that were coupled with a-*t*SS using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC). The Ni-NTA and *FGD3* aptamer selections were performed with HisPur Ni-NTA resin (Thermo Fisher Scientific) and adenosine-5′-triphosphate-agarose (Sigma–Aldrich), respectively.

Generation of RNA fragments using alkaline hydrolysis

Transcription reactions were performed at 37°C for 2 h in 1× transcription buffer [40 mM Tris–HCl (pH 7.5) and 2 mM spermidine], 16 mM $MgCl_2$, 2 mM each rNTP, 10 mM dithiothreitol (DTT), and one unit of T7 RNA polymerase (house preparation). After, 4 units of DNase I (New England Biolabs) and a 1× final concentration of DNase I reaction buffer (New England Biolabs) were added to each reaction, incubated for 1 h at 37°C, and purified twice using the RNA Clean and Concentrator kit (Zymo Research). RNA was then hydrolyzed under alkaline conditions (5 mM KOH, pH 10) at 90°C. Aliquots were withdrawn during hydrolysis and then quenched in 100 mM Tris–HCl (pH 7.5). Aliquots were assessed using a 3% agarose gel electrophoresis to determine RNA hydrolysis progression (Supplementary Fig. S1). Hydrolyzed RNA was then buffer-exchanged using the RNA Clean and Concentrator Kit and eluted in TET buffer [10 mM Tris–HCl (pH 7.5), 0.1 mM EDTA, and 0.001% Triton X-100].

Kit-purified RNAs were then treated with T4 polynucleotide kinase (PNK, New England Biolabs) and a 1× final T4 PNK reaction buffer for the removal of 2′,3′-cyclic phosphates. After, RNA was again purified and eluted in TET buffer.

In vitro selection

Kit-purified RNAs were resuspended in a 1× binding buffer [140 mM KCl, 10 mM NaCl, 10 mM Tris–HCl (pH 7.5), and 5 mM $MgCl_2$] and subsequently refolded by heating to 60°C for 5 min and then cooled to 25°C over the next 10 min. The folded RNA was then incubated with ligand-conjugated beads in a Spin-X tube (Corning, NY, USA) for 20 min. RNA was then centrifuged for 1 min at 3000 *g* to remove unbound RNAs, and beads were washed four times with 1× binding buffer by incubating for 5 min and centrifuging to remove nonspecific binders. Finally, the RNA was eluted with free ligand [e.g. 100 μM of amino *trans*-stiff stilbene (a-*t*SS) or 5 mM ATP-$MgCl_2$ in 1× binding buffer for the selections of a-*t*SS or adenosine aptamers, respectively] by incubating for 30 min at room temperature. The Ni-NTA aptamer was eluted in 7 M urea and 40 mM EDTA for 5 min. Elutions were collected and either extracted with phenol and chloroform (a-*t*SS aptamers) or kit purified prior to precipitation with 300 mM KCl, 1 μl of GlycoBlue (Invitrogen, MA, USA), and 2.5 volumes of cold 100% ethanol at −80°C. RNA was pelleted and then resuspended in TET buffer.

Addition of universal 3′ adapter to RNAs

To ligate a 3′ adapter to the truncated RNAs, the adapter oligo was first adenylylated in a reaction containing 1× T4 RNA ligase reaction buffer (New England Biolabs), 25 μM rATP, 20% PEG8000, 10 μM primer containing a 5′ monophosphate, and 20 U of T4 RNA ligase 1 (New England Biolabs) for 12 h at 25°C [24–26]. The reaction was buffer exchanged into TET buffer using 10 times the reaction volume of G10 Sephadex beads (Sigma–Aldrich) with centrifugation at 2000 *g* and stored at −80°C. The adenylylated oligo had the following primary structure, in which 5′-App-A and 3′SpC3 represent the 5′–5′ diphosphate-linked adenosines and a 3′ carbon spacer designed to prevent concatemerization of the oligo, respectively. 3′ adapter: 5′-App-AGATCGGAAGAGCACAC-3′SpC3

The hydrolyzed RNA was resuspended in TET buffer and added to a reaction containing 5 nM of the adenylylated 3′ adapter, 15% PEG8000, 1× T4 RNA ligase reaction buffer (New England Biolabs), and T4 RNA ligase 2, truncated KQ

(New England Biolabs). The reaction was incubated at 25°C for 18 h and 4°C for 2 h prior to purification using the RNA Clean and Concentrator Kit. RNA was then precipitated in 300 mM KCl, 1 µl of GlycoBlue (Invitrogen), and 2.5 volumes of cold 100% ethanol and pelleted by centrifugation.

Reverse transcription (RT)

RNA pellets were resuspended in TET buffer. Subsequently, 0.18 mM dNTPs, 0.02 mM biotinylated dUTP and dATP (Jena Bioscience), and 5 nM of the reverse primer were added to each reaction. The RNA and primer were annealed by heating at 90°C for 1 min and cooled on ice for 3 min before the addition of 10 mM DTT, 1× Protoscript buffer (New England Biolabs), 150 U of Protoscript II reverse transcriptase (New England Biolabs), 15 U of WarmStart RTx reverse transcriptase (New England Biolabs), and 4 U of *Bst* 3.0 DNA polymerase (New England Biolabs) to each reaction. The reaction was initiated at 25°C for 10 min, and then the temperature was ramped up to 35°C for 10 min, 42°C for 30 min, 55°C for 15 min, and 65°C for 30 min. Finally, the enzymes were inactivated at 70°C for 10 min. RT primer: 5′-GTGTGCTCTTCCGATCT-3′

Binding to streptavidin beads

5 U of RNaseH (New England Biolabs) was added to each RT reaction and incubated at 37°C for 1 h. After, the complementary DNA (cDNA) was buffer-exchanged using G10 Sephadex beads with TET buffer by centrifugation at 2000 *g* to remove free biotinylated dATP and dUTP. The cDNA was then incubated with magnetic streptavidin beads (Dynabeads™ MyOne™ C1, Invitrogen) that were pre-equilibrated with streptavidin wash buffer (SWB; 60 mM Tris, 50 mM KCl, and 0.001% Triton X-100) and incubated for 20 min at room temperature. After, five washes were performed with SWB to remove excess RT primer.

Addition of universal 5′ PCR adapter to 3′ of cDNA

Ligation was performed on the magnetic beads following the RT step. The 3′ end of the cDNA was ligated with a 5′ phosphorylated splint hairpin containing a 3′ carbon spacer to attach a constant-sequence region (T7 RNA polymerase A1 promoter sequence) to minimized sequences. The ligation was performed with a final concentration of 1× T4 DNA ligase reaction buffer (New England Biolabs) and 400 U of T4 DNA ligase (New England Biolabs), 1 mM rATP, 0.5 M betaine, 20% PEG8000, and 5 nM of splint hairpin. The reaction was incubated at 16°C for 6 h, 30°C for 6 h, and inactivated at 65°C for 15 min. Splint hairpin for first round of *in vitro* selection: 5′-P-CCTATAGTGAGTCGTATTAAAAAAATATAGG NNNNNNN- 3′SpC3

The DNA oligos contained a 5′ phosphate to facilitate ligation by T4 DNA ligase and a 3′ carbon spacer to prevent self-ligation. Both oligos were predicted to fold into a hairpin secondary structure with a random sequence (NNNNNNN) acting as a nonsequence-specific splint (Supplementary Fig. S2).

The ligated cDNA was washed five times with SWB and incubated for 5 min on an orbital shaker to remove excess splint hairpin oligo. Two washes with 200 mM KOH were performed to remove cDNA from the magnetic streptavidin beads. After, elutions were pooled and precipitated with 1 µl of GlycoBlue (Invitrogen, MA, USA) and 2.5 volumes of cold 100% ethanol, and pelleted by centrifugation. The pellet was resuspended in TET buffer.

Streamlined RNA ligation

To streamline the 5′-adenylylated oligo ligation, hydrolyzed RNAs were treated with or without T4 PNK. After, RNA was buffer exchanged using the RNA Clean and Concentrator Kit and eluted in TET buffer and added to a reaction containing 5 nM of the adenylylated 3′ adapter, 15% PEG8000, 1× RNA T4 ligase reaction buffer (New England Biolabs), and T4 RNA ligase 2, truncated KQ (New England Biolabs). The reaction was incubated at 25°C for 18 h and 4°C for 2 h prior to purification using the RNA Clean and Concentrator Kit. RNA was then precipitated in 300 mM KCl, 1 µl of GlycoBlue (Invitrogen), and 2.5 volumes of cold 100% ethanol and pelleted by centrifugation.

RNA pellets were resuspended in TET buffer. RT and cDNA ligations were performed as described above. PCR amplification was performed on the cDNA as described below and analyzed using 3% agarose gel electrophoresis to identify any gross biases (Supplementary Fig. S3A).

Library amplification by PCR

Polymerase chain reaction (PCR) amplification was performed on cDNA using a final concentration of a 1× Standard *Taq* reaction buffer (New England Biolabs), 0.2 mM dNTPs, 0.5 µM forward and reverse primer, and 1.25 U of DNA *Taq* polymerase (New England Biolabs). The reaction was amplified for 16 to 28 cycles (95°C for 30 s, 50°C for 30 s, and 72°C for 30 s) to determine the optimal amplification. Forward primer (first selection round): 5′-CCTATATT TTTTTAATACGACTCACTATAGG-3′ and reverse primer: 5′-GTGTGCTCTTCCGATCT-3′

HTS sequencing, processing, and mapping

DNA was prepared for sequencing by first amplifying the DNA library with an Illumina forward adapter and reverse primer (above) to incorporate the Illumina adapter during PCR. The reaction was amplified for 4 to 8 cycles (95°C for 30 s, 55°C for 30 s, and 72°C for 30 s). After, the Illumina barcodes were added in a second PCR reaction and amplified for 1 cycle (95°C for 30 s, 55°C for 30 s, and 72°C for 30 s) and then 7 cycles (95°C for 30 s, 65°C for 30 s, and 72°C for 30 s). Both PCR reactions contained a final concentration of a 1× Standard *Taq* reaction buffer (New England Biolabs), 0.2 mM dNTPs, 0.5 µM forward and reverse primer, and 1.25 U of DNA *Taq* polymerase (New England Biolabs). All PCR products were visualized on a 3% agarose gel prior to purification with a DNA Clean and Concentrator Kit (Zymo Research). The DNA libraries were then submitted for Illumina sequencing. Illumina forward adapter: 5′-ACGACGCTCTTCCGAT CTCCTATATTTTTTTAATACGACTCACTATAGG-3′

Sequenced reads were obtained using an Illumina MiSeq. The following discusses how each target of interest was processed to map sequences.

Mapping of reads

***FGD3* adenosine aptamer:** Analysis of sequences generated using CircLigase

Reads were merged with PEAR [27] using default settings. Adapters were clipped with cutadapt:

-a CAATGCGTCAAGCTCAG -g GATCTGTAATACGACT CACTATAGGGCAGACGTGCCTCACTAC -m 40
Bowtie2 [28] was used to map the reads to reference sequences, which corresponded to the human *FGD3* adenosine aptamer genomic sequence (Genbank JX863074). The parameters -N 1 and –local were used. For more permissive mapping, no minimum length was used during clipping and Bowtie1 scoring options were adjusted to -min_score G, 40, 8

Identification of unmapped reads

Sequences from high-throughput sequencing (HTS) were analyzed using FastAptamer [29, 30] count and cluster options. Clustering was performed with an edit distance of 10 with no filter flag. The top read per cluster was retrieved via grep -A 1 '[0–9]-[0–9]\{1,\}-1–0' and aligned manually in Jalview.

Mapping of reads: T4 DNA ligase

***FGD3* adenosine aptamer**: Analysis of sequences generated using T4 DNA ligase
Mapping of reads:
Reads were merged with PEAR using default settings. Adapters were clipped with cutadapt [31]:
-g TATAGTTTTTAATACGACTCACTATA
Bowtie2:
Alignment to FGD3: bowtie2 -x FGD3 -N 1 –local -q filename.fastq -S filename.sam
Samtools:
samtools view -b -S filename.sam > filename.bam
samtools sort filename.bam -o filename.sorted.bam
samtools index filename.sorted.bam
samtools fasta -F4 filename.sorted.bam > filename.sorted.fasta
Bedtools genomecov:
Depth: bedtools genomecov -d -ibam filename.sorted.bam
5′ truncations: bedtools genomecov -d -strand + -5 -ibam filename.sorted.bam
Reverse reads for 5′ truncations: bedtools genomecov -d -strand - -3 -ibam filename.sorted.bam
3′ truncations: bedtools genomecov -d -strand + -3 -ibam filename.sorted.bam
Reverse reads for 3′ truncations: bedtools genomecov -d -strand - -5 -ibam filename.sorted.bam
Were-1 riboswitch:
Mapping of reads:
Reads were merged with PEAR using default settings. Adapters were clipped with cutadapt [31]:
-g TATAGTTTTTAATACGACTCACTATA
Bowtie2:
Alignment to Were-1: bowtie2 -x Were-1 –end-to-end -q filename.fastq -S filename.sam
Samtools:
Sam files were converted to bam files as previously described.
Bedtools genomecov:
Bedtools genomecov files were processed as previously described.
Ni-NTA aptamer:
Mapping of reads:
Reads were merged with PEAR using default settings. Adapters were clipped with cutadapt [31]:
-g TATAGTTTTTAATACGACTCACTATA
Bowtie2:
Alignment to Ni-NTA: bowtie2 -x Ni-NTA -N 1 –local -q filename.fastq -S filename.sam
Samtools:
Sam files were converted to bam files as previously described.
Bedtools genomecov:
Depth: bedtools genomecov -d -strand + -ibam filename.sorted.bam
5′ truncations: bedtools genomecov -d -strand + -5 -ibam filename.sorted.bam
3′ truncations: bedtools genomecov -d -strand + -3 -ibam filename.sorted.bam
Bowtie2 was used first to remove any reads matching the PhiX genome and then to map sequences to the desired reference sequence. The parameter – N 1 and –local or –End-to-end were used when aligning the aptamer to their respective sequences. After, samtools was used to sort and index mapped files, and sequence alignments were visualized either using Jalview or in PRISM. Bedtools genomecov was used to determine the most common 5′ start sites and 3′ truncations and this helped to dictate design of minimized functional domains.

Column affinity chromatography of minimal RNAs

The transcription reactions of individual sequences were performed at 37°C for 2 h in 1× transcription buffer, 16 mM $MgCl_2$, 2 mM each rGTP, rUTP, rCTP, and 0.2 mM rATP, 3 μCi [α-^{32}P] ATP (Perkin Elmer), 10 mM DTT, and one unit of T7 RNA polymerase. After, 4 U of DNase I (New England Biolabs) and a 1× final concentration of DNase I reaction buffer (New England Biolabs) were added to each reaction; the reaction was incubated for 1 h at 37°C. RNAs were kit-purified using an RNA Clean and Concentrator and then refolded at 60°C for 5 min in 1× binding buffer. The RNA was cooled to 25°C for 10 min before binding to ligand-conjugated beads. The beads were subsequently washed for (5 min per wash) using 1× binding buffer and collected by centrifugation at 2000 *g* for 30 s. Ligand of interest (respective to RNA) was supplemented into 1× binding buffer and incubated for 20 min prior to collecting, and each fraction was analyzed using a liquid scintillation counter (Beckman Coulter LS6500).

T1 nuclease probing

The hhr1-Were-1 minimized RNA construct (25–103, Supplementary Fig. S4) [32] was transcribed using conditions described above and the reaction was buffer-exchanged using G25 Sephadex beads with 7 M urea, 40 mM EDTA, and then purified using a 10% denaturing PAGE. RNA was excised and eluted from the gel into 300 μl of 300 mM KCl and precipitated by adding 700 μl of cold 100% ethanol.

5′-labeled RNA was prepared in a reaction buffer (70 mM Tris–HCl, pH 7.6, 10 mM $MgCl_2$, and 5 mM DTT) using 3 pmol of 5′-dephosphorylated RNA, 0.9 μM [γ-^{32}P] ATP (Perkin Elmer), and 20 units of T4 PNK (New England Biolabs). The reaction was incubated at 37°C for 1 h prior to buffer exchange using G25 Sephadex beads with 7 M urea, 40 mM EDTA, and subsequently purified by 10% denaturing PAGE.

Purified 5′ radiolabeled RNA was divided into multiple reactions containing 1× binding buffer and varying concentrations of amino-*t*SS or no ligand (as a control). Samples were then refolded at 60°C for 5 min and subsequently cooled at room temperature for 1 h. Next, T1 nuclease was added

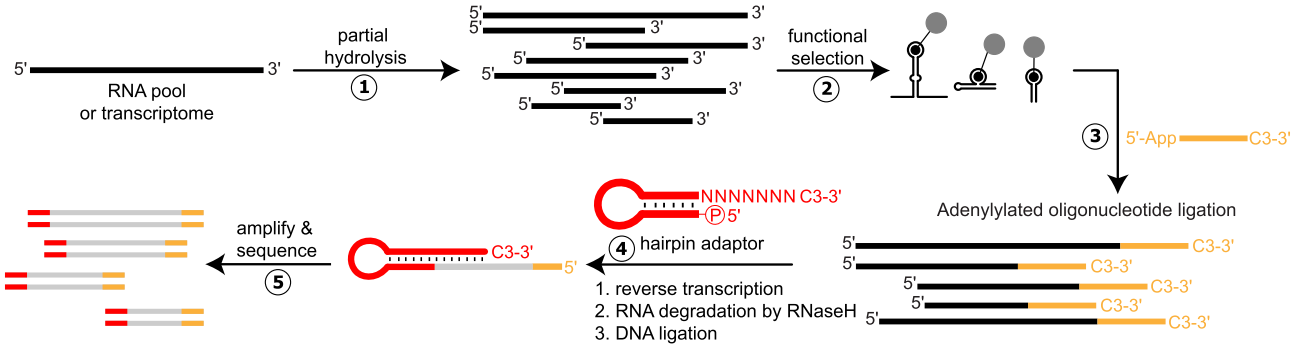


Figure 1. A schematic of high-throughput identification of minimized RNA domains. RNA (black) is first hydrolyzed to produce truncated variants of the RNA population (step 1). A selection is performed to isolate a function of interest (step 2) and a DNA adaptor is ligated to the 3′ end of the RNA (step 3) to serve as a primer-binding region for reverse transcription (RT) (step 4). After RT, a DNA hairpin adaptor is ligated to the 3′ end of the newly generated cDNA (gray, step 4) to serve as a primer-binding site for PCR. Finally, the cDNA is amplified with the newly appended primer-binding sites (step 5) and subsequently analyzed by HTS.

to each reaction (0.0007 U, Thermo Fisher Scientific), and samples were incubated at room temperature for 1 min and quenched with 7 M urea and 40 mm EDTA. The G-sequencing ladder was prepared using the same labeled RNA, in 250 mM sodium citrate, pH 7 and 0.013 U of RNase T1, incubated for 5 min at 55°C, and quenched with 7 M urea and 40 mM EDTA. The partial hydrolysis ladder (OH) was prepared in 50 mM $NaHCO_3$, 1 mM EDTA, pH 10, and incubated at 90°C for 5 min and quenched in 7 M urea and 40 mM EDTA. RNA was added to an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and mixed. Samples were centrifuged for 3 min at 8000 RPM, and the aqueous phase was collected and transferred to a new tube. A second extraction with chloroform:isoamyl alcohol was repeated and transferred to a new tube. Samples were directly loaded onto a 12% denaturing polyacrylamide gel electrophoresis (PAGE) gel. A phosphor imaging screen was exposed for a minimum of 24 h prior to imaging on a GE Typhoon 9410.

Results

Workflow for minimizing RNA domains

We developed a high-throughput method to delineate functional RNA domains in heterogeneous populations by hydrolyzing the RNA and subjecting it to a functional selection (Fig. 1). First, the RNA is partially hydrolyzed under alkaline conditions to promote scission of the phosphodiester backbone (step 1). This approach enables truncations at either 5′, 3′, or both termini of the RNA, resulting in a population of RNAs of varying length. This process is monitored by gel electrophoresis (Supplementary Fig. S1) to determine the ideal distribution of RNA lengths in the hydrolyzed pooled RNA. This step preserves a small fraction of full-length RNAs, but primarily yields truncated sequences, such as those lacking the primer-binding regions introduced into *in vitro* selection pools. Deletion of constant-sequence segments can be important when performing *in vitro* selections from genomic pools (genomic SELEX [14]), in which the primer regions do not originate from the genome and can modify the RNA structure. In contrast, there are some RNAs, in which the primer-binding regions may be essential for the formation of the functional RNA secondary structure [33], and the deletion of these constant regions may yield few, if any, functional RNAs. A method for experimental demarcation of active domains in such short pools would likely reveal whether the primer-binding sequences are essential for a function.

We next performed an *in vitro* selection step to enrich functional RNAs that specifically bind our target molecule of interest (step 2). This selection step enriches for RNAs that remain functional after truncations. The selected RNAs may not contain a consensus sequence for primer binding; we therefore introduced a new constant region by ligating an adenylylated oligo (5′-App) [11, 24, 26, 34] to the 3′ termini of the RNAs that had been treated with T4 PNK to facilitate reverse transcription (RT; steps 3 and 4) (Fig. 1 and Supplementary Fig. S2A). During the course of this study we found that treatment of the hydrolyzed RNA with PNK to remove 2′–3′ cyclic phosphate or monophosphate from the 2′ or 3′ positions of the terminal ribose was not necessary, because amplification of DNA derived from these hydrolyzed and ligated RNAs was equally efficient whether PNK was used or not (Supplementary Fig. S3A). We speculate that this result was likely due to the base treatment of the RNA, yielding either dephosphorylated 3′ termini or that the ligation proceeded in presence of 2′ or 3′ phosphate. T4 RNA ligase has been known to accommodate modified hydroxyls at the 2′ site [35] and recent work on T4 RNA ligase 1 has shown weak ligation in the presence of a 2′ phosphate [36]. Given that T4 RNA ligase 1 and T4 RNA ligase 2 have similar active sites with an unhindered pocket near the 2′ hydroxyl [37], it is likely that the ligation by the T4 RNA ligase 2 could occur without T4 PNK treatment by ligating a 2′ phosphorylated RNA to the adenylylated adapter oligo.

RT was performed in the presence of biotinylated dNTPs to help capture newly synthesized cDNAs. This step was particularly important to prevent amplification of undesired ligation products, such as concatemerized ligation primers resulting from an excess of RT primer during the RT reaction [38]. Subsequently, RNaseH was added to degrade the RNA and the biotinylated cDNA was then captured on streptavidin magnetic beads to remove excess RT primer. DNA ligation was then performed to append a 5′-phosphorylated splint DNA hairpin to the 3′ termini of the cDNA (corresponding to the 5′ termini of the selected RNAs) using T4 DNA ligase (step 4) [38, 39] (Fig. 1 and Supplementary Fig. S2B). This hairpin serves as a new primer-binding sequence for the DNA pool regardless of whether the previous 5′ primer region is present or not. Once the ligation was completed on streptavidin beads, the reaction was washed to remove excess hairpin prior to amplification

of the newly selected pool. Enriched sequences were then analyzed by HTS to determine the minimum domain(s) of functional RNAs (step 5) or used in another round of selection. Comparison of DNA ligated and amplified immediately after RNA hydrolysis (without selection) with DNA amplified after selection revealed significant shortening of the selected sequences (Supplementary Fig. S3A and B). Similarly, a comparison of the transcribed RNA from the selected population with the hydrolyzed starting sample of a single sequence (Were-1 photoriboswitch, see below) showed dramatic shortening after selection, providing experimental evidence for likely minimization of the functional domains (Supplementary Fig. S3C).

Human *FGD3* adenosine aptamer

The first functional RNA we analyzed was the adenosine-binding aptamer previously identified in an intron of the human *FGD3* gene [21]. A genomic aptamer pool is an ideal model for validation of our approach in demarcating functional domains because it typically contains nongenomic (artificial) primer-binding sites at the 5′ and 3′ ends. These sequences provide constant regions for amplification [14] and serve as promoters for *in vitro* transcription [40–43]. While these regions are necessary for *in vitro* selections, they do not originate from within the genome and can modulate the structure of the genomic sequences [14, 44]. Therefore, a method that assists in mapping functional genomic RNAs lacking the synthetic primer-binding sequences is desirable. The human *FGD3* aptamer structure was previously discovered using a genomic SELEX experiment, but the functional domain was defined based on the similarity of the secondary structure to previously described adenosine aptamers [21, 45]. A secondary structure-based truncation of the aptamer was attempted, but the presumed minimized sequence only bound ATP beads cotranscriptionally, suggesting that it misfolded after purification [21]. To reveal any stable functional domains, we applied the experimental minimization to the genomic pool containing the *FGD3* aptamer (Fig. 2A).

The *FGD3* sequences were first analyzed using bedtools genomecov [46] to assess both the 5′ and 3′ termini of the aptamer. Three major populations were identified with the 5′ termini at positions 1, 57, and 69–76. Another group of sequences had 5′ termini around positions 110–140, which were suggestive of forming a heterodimeric complex because they lacked the large adenosine-binding loop (based on previously deduced models of these aptamers) and did not bind ATP beads alone (Supplementary Fig. S5). The 3′ termini were clustered around positions 165, 173, and 181 (Fig. 2B). After manually aligning the 40 most abundant reads, we noticed a distinct truncation at the 5′ end of the previously defined functional aptamer (Fig. 2C) [21, 45]. Together these data suggested that the first ∼60 nucleotides of the *in vitro* selected sequence are nonessential for binding. The 3′ ends of the selected sequences were more heterogeneous, but a clear minimal sequence could be discerned outside of the primer-binding sequences. A construct, *FGD3* (76–165), consisting of a sequence ranging from the beginning of one of the most prevalent 5′ truncations to the end of the innermost major 3′ truncation was tested in a standard affinity-based experiment (Fig. 2D). We observed that ∼60% of RNA eluted off the beads in the presence of excess ATP, suggesting that a robust adenosine-binding aptamer was identified and minimized.

We also tested a second approach for ligating the cDNA termini using CircLigase, rather than random-sequence splint ligation using T4 DNA ligase (Supplementary Fig. S6). In this selection scheme, the *FGD3* aptamer was hydrolyzed as described above, but the RT step was performed with a primer containing the constant region from the original aptamer pool with both the HTS forward and reverse primers. Subsequently, CircLigase was used to circularize the cDNA and then amplification was performed to further enrich for new aptamers. After four rounds of selection, random priming was performed in the RT step to increase the sequence coverage and one more round of *in vitro* selection was performed prior to submission to HTS. HTS sequences were again mapped to the full-length aptamer sequence and analyzed for minimal sequences (Supplementary Fig. S7). While a sequence similar to the minimized *FGD3* aptamer was identified, two other striking results that were revealed: first, an internal deletion [*FGD3* (Del)] was observed within the *FGD3* aptamer. This internal deletion exhibited high ligand affinity in ATP column binding, with ∼60% of the population eluting in the presence of ATP, similarly to the *FGD3* (76–165) aptamer (Fig. 2E). Second, we noticed many nonoverlapping segments of the aptamer and hypothesized that a split aptamer in the form of a heterodimer consisting of two strands of the minimized *FGD3* aptamer formed from the hydrolyzed RNAs. We confirmed that the individual strands did not bind the ATP column (Supplementary Fig. S5), but when both sequences were transcribed simultaneously, they formed a functional heterodimer and ∼3% of the complex was bound and eluted using free ATP (Fig. 2F). These results show that the method not only demarcates the functional domain but also provides a facile approach for directly discovering split aptamers. While the results obtained from the CircLigase minimization scheme yielded information about the *FGD3* aptamer, the high-throughput pipeline using the 5′-App oligo ligation (Fig. 1) was used in subsequent experiments due to limited sequence bias, more straightforward workflow, and cost-effectiveness.

Identification of Were-1 photoriboswitch aptamer domain

Next, we applied the minimization to a heterogeneous pool of synthetic riboswitches. A photoriboswitch, Werewolf-1 (Were-1), was previously found to bind a photoactive molecular switch [amino *trans* stiff-stilbene (a-*t*SS)] with a K_D of ∼1.5 μM [22]. However, efforts aimed at identifying the minimal aptamer domain of Were-1 through chemical and enzymatic probing have been confounded by the conformational changes of the riboswitch upon either binding of the ligand or subsequent rearrangements of its expression platform.

To demarcate the ligand-binding domain of Were-1, the full-length construct (Fig. 3A) was doped (10%) into a selected riboswitch pool and hydrolyzed under alkaline conditions prior to selection on magnetic beads conjugated with a-*t*SS. The RNA captured on beads was eluted in the presence of a-*t*SS and sequenced after one round of *in vitro* selection. Sequences obtained from HTS were mapped to the Were-1 parent sequence and analyzed for minimal and nonoverlapping (split-aptamer) sequences, as described above. First, we measured the overall frequency of reads terminated at each position to discern which segments were enriched by the binding assay (Fig. 3B). Positions 25–67 of the riboswitch had

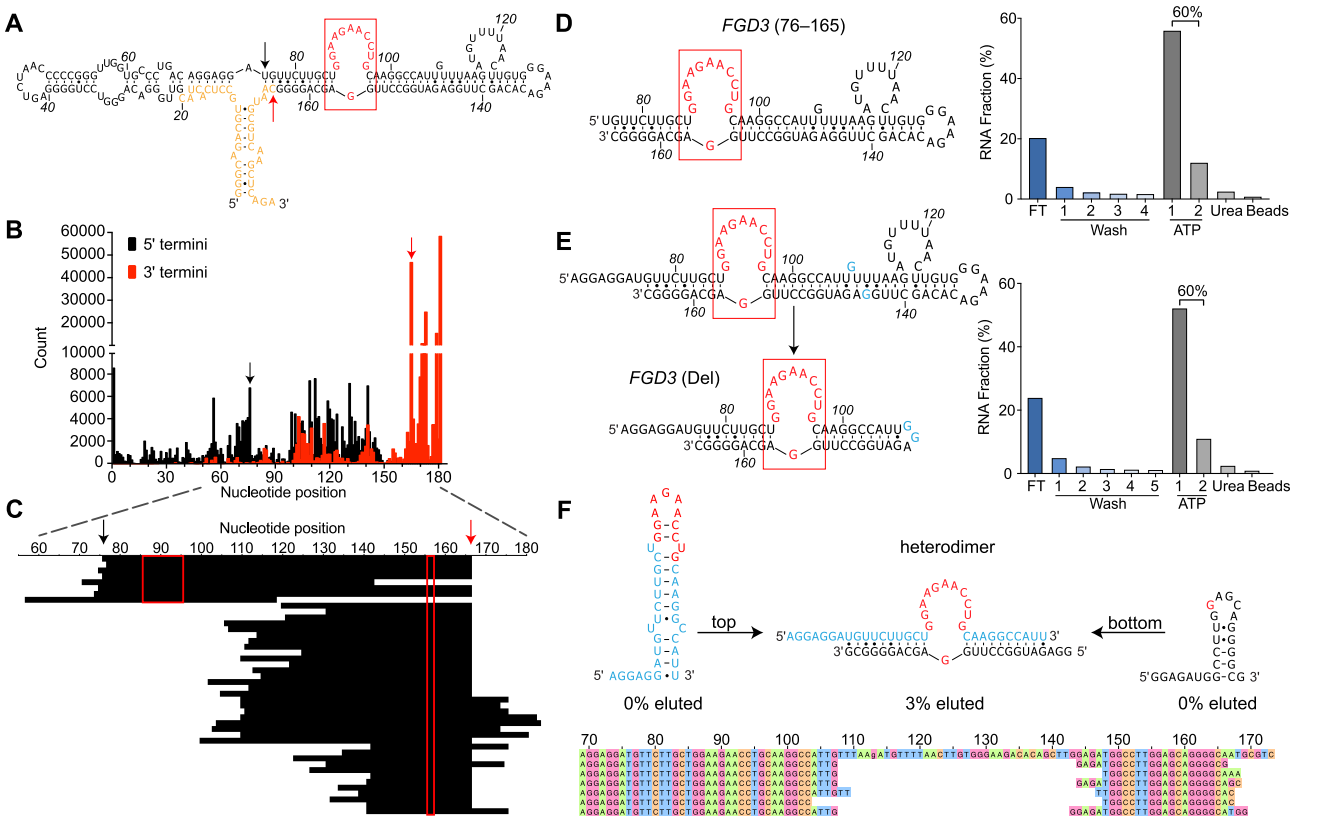


Figure 2. Experimental demarcation of the human *FGD3* adenosine aptamer domain. (**A**) A secondary structure model of the aptamer identified by genomic SELEX [21]. The adenosine-binding loop is outlined with a red box; yellow nucleotides represent the primer-binding regions derived from the *in vitro* selection pool. Arrows mark truncations that commonly occurred at the 5′ (black) and 3′ (red) ends of the reselected truncated pool. (**B**) Mapping of minimized sequences to the full-length genomic *FGD3* aptamer. The graph displays the distribution of the 5′ (black) and 3′ (red) termini. (**C**) Alignment of the top 40 most common sequences from the minimization experiment. Red boxes show the location of the adenosine-binding loop. (**D**) *FGD3* (76–165) aptamer secondary structure and its binding to ATP-agarose beads. FT: flow-through; ATP: competitive elution with free ATP (5 mM). (**E**) An internal deletion observed in the *FGD3* aptamer. The nucleotides highlighted in blue (107 and 144) show the positions of the deletion within the sequence. ATP column binding profile of the *FGD3* (Del) shows binding and elution profile almost identical to the truncated sequence shown in panel (D). (**F**) A heterodimeric minimized domain revealed by modeling of the cofolding of nonoverlapping sequences shown from the CircLigase experiment. The binding of individual strands (shown as % competitively eluted) was compared with ATP column binding by the heterodimer (see Supplementary Fig. S5 for full binding profiles). The blue (top strand) and black (bottom strand) nucleotides within the split heterodimer represent the sequences from the 5′ and 3′ segments, respectively. The alignment shows the shortest nonoverlapping sequences predicted to form the minimal split-aptamer domain. All minimized secondary structure models were based on computational predictions using ViennaRNA [59].

the highest frequency, and the regions closer to the 5′ and 3′ termini were lower in positional abundance. Five positions, at nucleotides 1, 25, 38, 46, and 109, were identified to be among the most common 5′ termini, whereas the most frequent 3′ truncations occurred at nucleotides 67, 77, 103, 107, and 147 (Fig. 3C). Manual alignment of the 20 most common sequences represented by these truncations could be categorized into three populations (Fig. 3D): the first population aligned to the middle of the riboswitch, largely overlapping with the random-sequence region of the starting pool, whereas the other two populations aligned to either end of the riboswitch, suggesting that they were parts of functional split-aptamer heterodimers. To decide on which sequence to assess for activity, the top represented sequences within the presumed *cis* construct population were aligned to initially identify those that do not contain the primer-binding regions. Sequences supported by the 5′ and 3′ truncation trends observed from this experiment, overlapping with segments of the riboswitch sensitive to T1 nuclease probing in presence of a-*t*SS [22], revealed one of the most abundant presumed *cis* constructs, which spanned positions 25–103 (Fig. 3C). We anticipate that for structure-switching RNAs, such as riboswitches, this analysis coupled with secondary structure modeling may be necessary to distinguish between *cis* and *trans* constructs. For example, a segment spanning positions ~47–87 is strongly represented among the selected sequences but corresponds to a sequence that starts and ends in a loop of the presumed stem–loop structures of the riboswitch and was therefore deemed to be part of a putative *trans* complex.

To determine whether the Were-1 (25–103) construct had a similar affinity to a-*t*SS as the full-length riboswitch, the accessible regions of Were-1 (25–103) were probed using RNase T1 footprinting [22]. We observed a ligand-dependent reduction in RNase T1 digestion within Were-1 (25–103) as a-*t*SS concentration increased (Fig. 3E). An apparent dissociation constant (K_D) value was derived from four distinct sites (G42, G46, G49, and G77), yielding a K_D of 5.3 μM and revealing a similar affinity to the full-length Were-1. A secondary structure model based on the open segments suggested that the minimized Were-1 (25–103) requires a flanking region to form a structure that binds a-*t*SS (Fig. 3F).

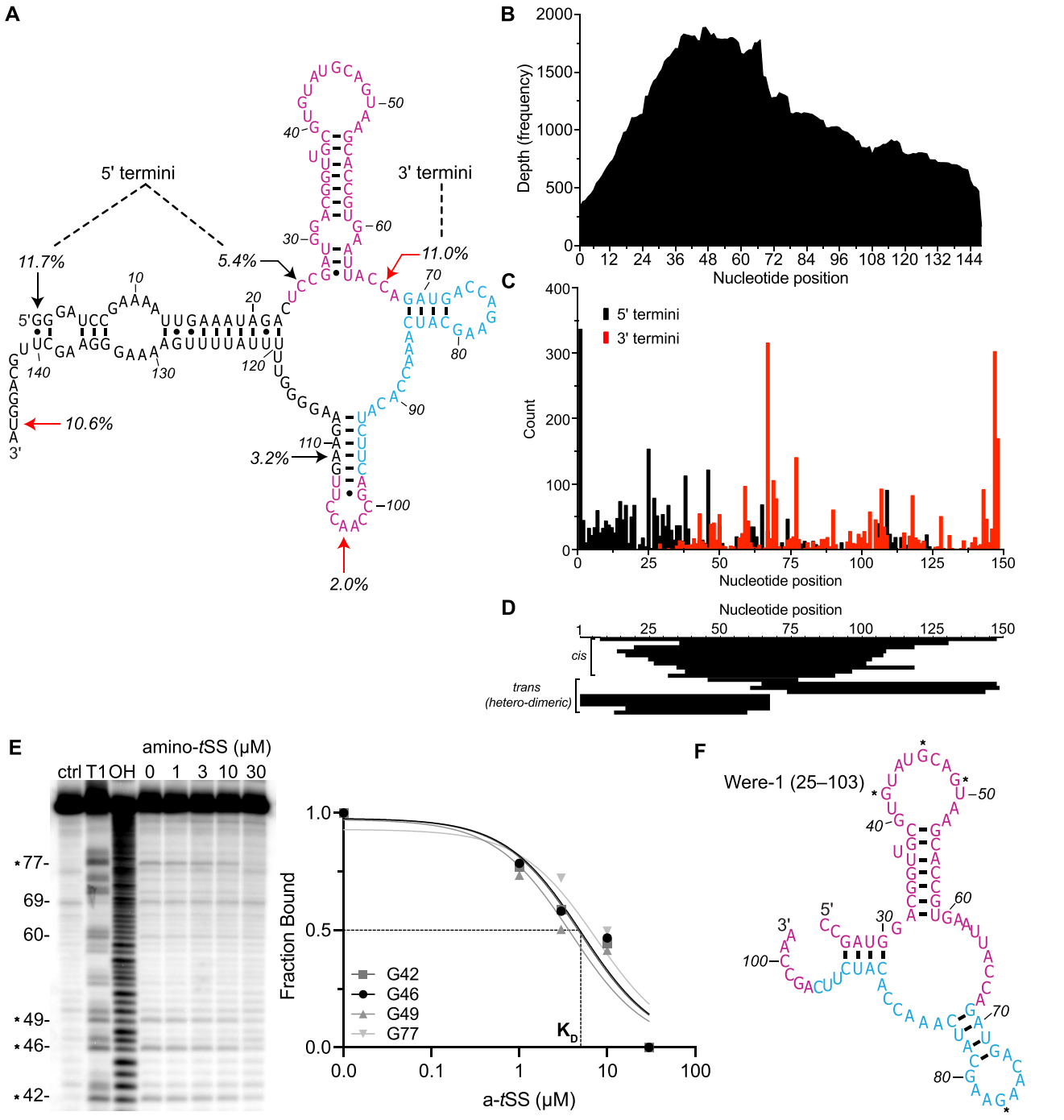


Figure 3. Demarcation of the minimized Were-1 photoriboswitch aptamer domain. (**A**) Mapping of minimized sequences to the Were-1 riboswitch. The original model of the riboswitch secondary structure, with the pink letters representing positions derived from random regions of the starting *in vitro* selection pool and blue representing partially randomized positions. Nucleotides in black represent primer-binding regions of the pool. (**B**) Graph depicting nucleotide positions with the highest representation among the truncated and reselected sequences mapped to the Were-1 riboswitch. (**C**) Most common 5′ and 3′ termini among truncated sequences after selection for a-*t*SS binding. The three most frequent start and end positions, labeled in black and red respectively, are indicated in the secondary structure model in panel (A). (**D**) Top 20 most abundant truncated sequences aligned to full-length Were-1. Three populations were observed: sequences aligned to the middle of the sequence, likely forming a *cis* aptamer, and two nonoverlapping populations extending to the termini of the riboswitch that likely form *trans* (heterodimeric) aptamers. (**E**) Native RNase T1 probing of Were-1 (25–103). Lanes from left to right: undigested RNA (ctrl), digested G-sequencing (T1), partial hydrolysis ladder (OH), and partial T1 digestion in the presence of increasing a-*t*SS concentrations, with ligand concentrations indicated above the PAGE gel image. Plot of the estimated fraction of RNA bound versus a-*t*SS concentration. Fraction bound values were calculated by measuring the band intensities at four sites (G42, G46, G49, and G77) and fitting to a simple K_D model. Band intensities were normalized to positions G60 and G69. A lower contrast version of the gel image is shown in Supplementary Fig. S12. (**F**) Secondary structure model of the Were-1 (25–103) aptamer based on the native RNase T1 probing gel depicted in panel (E). Asterisks indicate positions that were less susceptible to native RNase T1 cleavage in the presence of ligand.

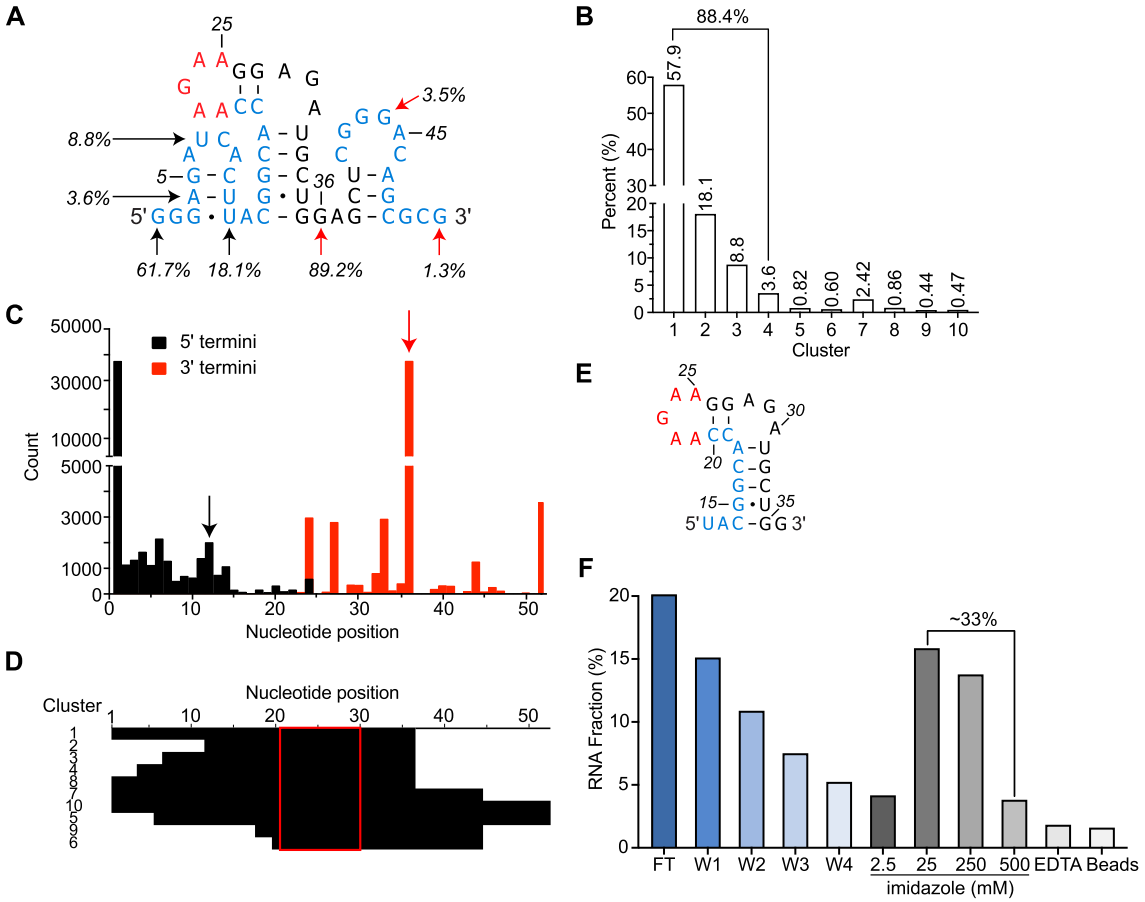


Figure 4. Identification and minimization of a novel Ni-NTA aptamer. (**A**) Full-length sequence and proposed secondary structure of a novel Ni-NTA–binding aptamer as predicted by ViennaRNA [59]. Regions in blue represent the primer-binding regions from the *in vitro* selection pool. Nucleotides in black and red are from the random regions of the pool. Arrows indicate 5′ (black) and 3′ truncations (red) identified in the minimization experiment. (**B**) The top ten clusters were plotted to indicate the relative abundance within the mapped populations. (**C**) HTS read depth of the 5′ start sites (black) and 3′ truncation sites (red) mapped to the parent sequence. (**D**) Top 10 clusters aligned to the parent sequence (A). The red box outlines the common region among the sequences. (**E**) Predicted secondary structure model of the minimized Ni-NTA aptamer represented by cluster 2. (**F**) The minimized Ni-NTA aptamer binding to the Ni-NTA resin and eluted with imidazole at indicated concentrations.

We then compared the minimization of Were-1 doped within a heterogeneous pool with a stand-alone Were-1 construct. A functional selection was performed as described above to map the boundary of the aptamer domain and reveal any differences in the samples. Positional frequency of the Were-1 segments showed that performing the minimization on either a single construct or within a highly heterogeneous pool led to a similar distribution of both the 5′ and 3′ termini (Supplementary Fig. S8), and suggested that our method may reveal multiple families of minimized aptamers within the selected pool of sequences. To test this hypothesis, we analyzed three additional putative aptamers contained within the same heterogeneous pool as Were-1, revealing well defined boundaries for their presumed aptamer domains (Supplementary Figs S9–S11), and showcasing the high-throughput nature of this workflow.

Identification of a Ni-NTA aptamer

To test our minimization method without *a priori* knowledge of the identity of the target RNA, we used an *in vitro* selected pool containing previously identified Ni-NTA aptamers to uncover new examples of this class of functional RNAs [23]. Analysis of the HTS data revealed an abundant sequence, which was aligned to identify minimal functional variants (Fig. 4A). To assess the subpopulations within the pool, we used FastAptamer [30] to count and cluster the sequences and compared the top 10 clusters; the top 4 clusters represented ~88% of the sequences within the pool (Fig. 4B). Mapping of the 5′ termini showed nucleotide positions 1, 4, 7, and 11 had the greatest abundance, while a dominant 3′ truncation was apparent at nucleotide position 36 with an abundance of 89.2% (Fig. 4C). Sequence alignment of the top 10 clusters revealed the putative minimal domain (Fig. 4D). The 10 most abundant clusters of the novel Ni-NTA aptamer contain a purine-rich segment (positions 21–30), as have previously identified Ni-NTA binding motifs [23, 47]. The predicted secondary structure of the suggested Ni-NTA aptamer indicated that a portion of the 5′ constant region (blue) played an important role in the structure formation. A construct that started at the beginning of one of the most prevalent 5′ truncations (position 11) that was predicted to form the same secondary structure as the full-length sequence, combined with the most common 3′ truncation, was tested for Ni-NTA affinity (Fig. 4E). We observed that ~33% of RNA bound to the Ni-NTA column and eluted in the presence of imidazole at concentrations higher than 25 mM, confirming the sequence as a new Ni-NTA aptamer (Fig. 4F). In addition, we used RNArobo

[48] to find many examples of the minimized domains of previously characterized Ni-NTA aptamer embedded within many different parent sequences (Supplementary Fig. S13) [23]. This aptamer has recently been used to isolate mRNA–protein complexes from cell extracts, demonstrating robust activity of the minimized domain in the context of a large ribonucleoprotein structure [49].

Discussion

The discovery of functional RNAs has led to many advances in biology and biotechnology. One limitation in increasing the output of these innovations is the identification of the minimal motifs necessary for the given function. *In vitro* selection is one approach commonly used to discover new functional RNAs, which typically do not require their entire sequences to fulfill their biochemical role. Demarcation of the functional domains' boundaries is necessary for downstream applications. Conversely, functional RNAs may also be isolated from pools that are too short to yield functional RNAs solely from their heterogeneous (random) regions [33] or are structurally and informationally complex, requiring primer-binding regions to form the active motifs [50]. This situation can be particularly difficult to interpret if the 3′ constant region participates in the functional motif and the RNA is characterized using chemical probing methods that require reverse transcription for readout, for example, in approaches that analyze solvent accessible segments of the RNA using alkylating and acylating reagents [51–53]. All these scenarios would greatly benefit from a high-throughput approach that yields information about the minimal sequence necessary for the chosen biochemical function.

Here, we developed a high-throughput platform that identifies functional RNA domains within heterogeneous RNA pools. This approach not only revealed multiple functional RNAs from both synthetic and genome-derived pools but also led to the first direct discovery of split fragments of a functional aptamer. This platform will also enable removal of sequence bias associated with the constant regions introduced during the *in vitro* selection experiment [54]. By creating a system to interchange these sequences, we increase the opportunity to identify functional regions within an RNA without losing its biological relevance. Furthermore, the installation of these new primer regions allows for the development of new workflows to be adapted in biological systems, where these sequences serve to promote transcription or translation. Some caution should be noted, because solvent accessible positions in functional RNAs, such as dynamic or unstructured loops, are likely more susceptible to hydrolysis, whereas positions engaged in secondary and tertiary structure may be less hydrolyzed even at elevated pH and temperatures. Furthermore, highly structured constructs may be somewhat suppressed in the HTS data because the conditions of the enzymatic ligation step, performed at neutral pH and in presence of Mg^{2+}, likely modulate the availability of the 3′ termini for ligation. Nevertheless, we anticipate that this method will greatly increase the throughput of discovery of new functional RNAs by reducing the tedious workflow of biochemically assessing individual constructs identified from *in vitro* selection pools.

The ability to directly discover split aptamers can further expand the applications of functional RNAs in multiple settings, such as in design of biosensors or split ribozymes used in cells to regulate gene expression. One exciting application may be the development of split light up aptamers to use in study of RNA localization and dynamics [55–57]. We anticipate that the method will enable fast identification of diverse functional RNAs, which will not only increase discovery of novel targets but help to engineer RNAs for multiple applications. While we focused on the discovery of demarcating functional RNAs within synthetic and genomic RNA pools, we envision this platform accommodating the discovery of functional RNAs within a transcriptome as well [58].

Acknowledgements

We would like to thank the staff at the Genomic High Throughput Facility (GHTF) Shared Resource of the Cancer Center Support Grant (P30CA-062203) at the University of California, Irvine and NIH shared instrumentation grants 1S10OD010794-01 and 1S10OD021718-01.

Author contributions: Conceptualization: M.M.K.V, B.L., K.H.C., and A.L.; Methodology: B.L., K.H.C., M.M.K.V, and A.L.; Investigation: B.L. and M.M.K.V.; Formal analysis: B.L. and M.M.K.V.; Writing—original draft: B.L. and A.L.; Writing—review and editing: B.L., M.M.K.V., K.H.C., and A.L.; Visualization: B.L.; Funding acquisition: A.L.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by NASA Exobiology [80NSSC21K0488 to A.L.]; W.M. Keck Foundation [to A.L.]; NASA ICAR [80NSSC21K0596 to A.L.]; NSF CBET Biophotonics [1804220 to A.L.]. Funding to pay the Open Access publication charges for this article was provided by NASA ICAR.

Data availability

Sequencing data were deposited on SAR (Sequence Read Archive [https://www.ncbi.nlm.nih.gov/sra]) and can be accessed with SUB14463409. The bioproject accession number is PRJNA1113781.

References

1. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 1990;**249**:505–10. https://doi.org/10.1126/science.2200121
2. Ellington AD, Szostak JW. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* 1990;**346**:818–22. https://doi.org/10.1038/346818a0
3. Martini L, Meyer AJ, Ellefson JW *et al. In vitro* selection for small-molecule-triggered strand displacement and riboswitch activity. *ACS Synth Biol* 2015;**4**:1144–50. https://doi.org/10.1021/acssynbio.5b00054

4. Harbaugh SV, Martin JA, Weinstein J *et al.* Screening and selection of artificial riboswitches. *Methods* 2018;**143**:77–89. https://doi.org/10.1016/j.ymeth.2018.05.012
5. Sabeti PC, Unrau PJ, Bartel DP. Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem Biol* 1997;**4**:767–74. https://doi.org/10.1016/S1074-5521(97)90315-X
6. He X, Guo L, He J *et al.* Stepping library-based post-SELEX strategy approaching to the minimized aptamer in SPR. *Anal Chem* 2017;**89**:6559–66. https://doi.org/10.1021/acs.analchem.7b00700
7. Bittker JA, Le BV, Liu DR. Nucleic acid evolution and minimization by nonhomologous random recombination. *Nat Biotechnol* 2002;**20**:1024–9. https://doi.org/10.1038/nbt736
8. Liu JM, Liu DR. Discovery of a mRNA mitochondrial localization element in *Saccharomyces cerevisiae* by nonhomologous random recombination and *in vivo* selection. *Nucleic Acids Res* 2007;**35**:6750–61. https://doi.org/10.1093/nar/gkm777
9. Wang QS, Unrau PJ. Ribozyme motif structure mapped using random recombination and selection. *RNA* 2005;**11**:404–11. https://doi.org/10.1261/rna.7238705
10. Ekland EH, Bartel DP. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucl Acids Res* 1995;**23**:3231–8. https://doi.org/10.1093/nar/23.16.3231
11. Chapple KE, Bartel DP, Unrau PJ. Combinatorial minimization and secondary structure determination of a nucleotide synthase ribozyme. *RNA* 2003;**9**:1208–20. https://doi.org/10.1261/rna.5500603
12. Arriola JT, Müller UF. A combinatorial method to isolate short ribozymes from complex ribozyme libraries. *Nucleic Acids Res* 2020;**48**:e116. https://doi.org/10.1093/nar/gkaa834
13. Wilson DS, Szostak JW. *In vitro* selection of functional nucleic acids. *Annu Rev Biochem* 1999;**68**:611–47. https://doi.org/10.1146/annurev.biochem.68.1.611
14. Singer BS, Shtatiand T, Brown D *et al.* Libraries for genomic SELEX. *Nucleic Acids Res* 1997;**25**:781–6. https://doi.org/10.1093/nar/25.4.781
15. Peng H, Latifi B, Müller S *et al.* Self-cleaving ribozymes: substrate specificity and synthetic biology applications. *RSC Chem Biol* 2021;**2**:1370–83. https://doi.org/10.1039/D0CB00207K
16. Dey SK, Filonov GS, Olarerin-George AO *et al.* Repurposing an adenine riboswitch into a fluorogenic imaging and sensing tag. *Nat Chem Biol* 2022;**18**:180–90. https://doi.org/10.1038/s41589-021-00925-0
17. Mohsen MG, Midy MK, Balaji A *et al.* Exploiting natural riboswitches for aptamer engineering and validation. *Nucleic Acids Res* 2023;**51**:966–81. https://doi.org/10.1093/nar/gkac1218
18. Porter EB, Polaski JT, Morck MM *et al.* Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat Chem Biol* 2017;**13**:295–301. https://doi.org/10.1038/nchembio.2278
19. Rehm C, Klauser B, Hartig JS. In: Ponchon L (ed.), *Engineering Aptazyme Switches for Conditional Gene Expression in Mammalian Cells Utilizing an in Vivo Screening Approach BT - RNA Scaffolds: Methods and Protocols*. New York, NY: Springer, 2015, 127–40.
20. Yokobayashi Y. Applications of high-throughput sequencing to analyze and engineer ribozymes. *Methods* 2019;**161**:41–5. https://doi.org/10.1016/j.ymeth.2019.02.001
21. Vu MMK, Jameson NE, Masuda SJ *et al.* Convergent evolution of adenosine aptamers spanning bacterial, human, and random sequences revealed by structure-based bioinformatics and genomic SELEX. *Chem Biol* 2012;**19**:1247–54. https://doi.org/10.1016/j.chembiol.2012.08.010
22. Rotstan KA, Abdelsayed MM, Passalacqua LFM *et al.* Regulation of mRNA translation by a photoriboswitch. *eLife* 2020;**9**:e51737. https://doi.org/10.7554/eLife.51737
23. Chizzolini F, Passalacqua LFM, Oumais M *et al.* Large phenotypic enhancement of structured random RNA pools. *J Am Chem Soc* 2020;**142**:1941–51. https://doi.org/10.1021/jacs.9b11396
24. Silverman SK. Practical and general synthesis of 5′-adenylated RNA (5′-AppRNA). *RNA* 2004;**10**:731–46. https://doi.org/10.1261/rna.5247704
25. Wang Y, Silverman SK. Efficient RNA 5′-adenylation by T4 DNA ligase to facilitate practical applications. *RNA* 2006;**12**:1142–6. https://doi.org/10.1261/rna.33106
26. Duzdevich D, Carr CE, Szostak JW. Deep sequencing of non-enzymatic RNA primer extension. *Nucleic Acids Res* 2020;**48**:e70. https://doi.org/10.1093/nar/gkaa400
27. Zhang J, Kobert K, Flouri T *et al.* PEAR: a fast and accurate Illumina Paired-end reAd mergeR. *Bioinformatics* 2014;**30**:614–20. https://doi.org/10.1093/bioinformatics/btt593
28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–9. https://doi.org/10.1038/nmeth.1923
29. Alam KK, Chang JL, Burke DH. FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol Ther Nucleic Acids* 2015;**4**:e230. https://doi.org/10.1038/mtna.2015.4
30. Kramer ST, Gruenke PR, Alam KK *et al.* FASTAptameR 2.0: a web tool for combinatorial sequence selections. *Mol Ther Nucleic Acids* 2022;**29**:862–70. https://doi.org/10.1016/j.omtn.2022.08.030
31. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j* 2011;**17**:10–2. https://doi.org/10.14806/ej.17.1.200
32. Jimenez RM, Delwart E, Lupták A. Structure-based search reveals hammerhead ribozymes in the human microbiome. *J Biol Chem* 2011;**286**:7737–43. https://doi.org/10.1074/jbc.C110.209288
33. Pressman AD, Liu Z, Janzen E *et al.* Mapping a systematic ribozyme fitness landscape reveals a frustrated evolutionary network for self-aminoacylating RNA. *J Am Chem Soc* 2019;**141**:6213–23. https://doi.org/10.1021/jacs.8b13298
34. Chen YR, Zheng Y, Liu B *et al.* A cost-effective method for Illumina small RNA-seq library preparation using T4 RNA ligase 1 adenylated adapters. *Plant Methods* 2012;**8**:41. https://doi.org/10.1186/1746-4811-8-41
35. Ebhardt HA, Thi EP, Wang M-B *et al.* Extensive 3 modification of plant small RNAs is modulated by helper component-proteinase expression. *Proc Natl Acad Sci USA* 2005;**102**:13398–403. https://doi.org/10.1073/pnas.0506597102
36. Hu Y, Lopez VA, Xu H *et al.* Biochemical and structural insights into a 5' to 3' RNA ligase reveal a potential role in tRNA ligation. *Proc Natl Acad Sci USA* 2024;**121**:e2408249121. https://doi.org/10.1073/pnas.2408249121
37. Nandakumar J, Shuman S, Lima CD. RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. *Cell* 2006;**127**:71–84. https://doi.org/10.1016/j.cell.2006.08.038
38. Ritchey LE, Su Z, Tang Y *et al.* Structure-seq2: sensitive and accurate genome-wide profiling of RNA structure in vivo. *Nucleic Acids Res* 2017;**45**:e135. https://doi.org/10.1093/nar/gkx533
39. Kwok CK, Ding Y, Sherlock ME *et al.* A hybridization-based approach for quantitative and low-bias single-stranded DNA ligation. *Anal Biochem* 2013;**435**:181–6. https://doi.org/10.1016/j.ab.2013.01.008
40. Salehi-Ashtiani K, Lupták A, Litovchick A *et al.* A genomewide search for ribozymes reveals an HDV-like sequence in the Human CPEB3 gene. *Science* 2006;**313**:1788–92.
41. Abdelsayed MM, Ho BT, Vu MMK *et al.* Multiplex aptamer discovery through Apta-Seq and its application to ATP aptamers derived from Human-genomic SELEX. *ACS Chem Biol* 2017;**12**:2149–56. https://doi.org/10.1021/acschembio.7b00001
42. Sedlyarova N, Rescheneder P, Magán A *et al.* Natural RNA polymerase aptamers regulate transcription in E. coli. *Mol Cell* 2017;**67**:30–43. https://doi.org/10.1016/j.molcel.2017.05.025
43. Gold L, Brown D, He Y *et al.* From oligonucleotide shapes to genomic SELEX: novel biological regulatory loops. *Proc Natl Acad Sci USA* 1997;**94**:59–64. https://doi.org/10.1073/pnas.94.1.59

44. Shtatland T, Gill SC, Javornik BE *et al.* Interactions of *Escherichia coli* RNA with bacteriophage MS2 coat protein: genomic SELEX. *Nucleic Acid Res* 2000;**21**:e93. https://doi.org/10.1093/nar/28.21.e93
45. Sassanfar M, Szostak JW. An RNA motif that binds ATP. *Nature* 1993;**364**:550–3. https://doi.org/10.1038/364550a0
46. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2. https://doi.org/10.1093/bioinformatics/btq033
47. Hofmann HP, Limmer S, Hornung V *et al.* Ni^{2+}-binding RNA motifs with an asymmetric purine-rich internal loop and a G-A base pair. *RNA* 1998;**3**:1289–300.
48. Rampášek L, Jimenez RM, Lupták A *et al.* RNA motif search with data-driven element ordering. *BMC Bioinf* 2016;**17**:216. https://doi.org/10.1186/s12859-016-1074-x
49. Halbers LP, Cole KH, Ng KK *et al.* A modular platform for bioluminescent RNA tracking. *Nat Commun* 2024;**15**:9992. https://doi.org/10.1038/s41467-024-54263-5
50. Carothers JM, Oestreich SC, Davis JH *et al.* Informational complexity and functional activity of RNA structures. *J Am Chem Soc* 2004;**126**:5130–7. https://doi.org/10.1021/ja031504a
51. Wells SE, Hughes JMX, Haller Igel A *et al.* Use of dimethyl sulfate to probe RNA structure *in vivo*. In: *Methods in Enzymology*. Vol. **318**. Academic Press, 2000, 479–93. https://doi.org/10.1016/S0076-6879(00)18071-1
52. Spitale RC, Incarnato D. Probing the dynamic RNA structurome and its functions. *Nat Rev Genet* 2023;**24**:178–96. https://doi.org/10.1038/s41576-022-00546-w
53. Merino EJ, Wilkinson KA, Coughlan JL *et al.* RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* 2005;**127**:4223–31. https://doi.org/10.1021/ja043822v
54. Pan W, Clawson GA. The shorter the better: reducing fixed primer regions of oligonucleotide libraries for aptamer selection. *Molecules* 2009;**14**:1353–69. https://doi.org/10.3390/molecules14041353
55. Wang R, Zhang Q, Zhang Y *et al.* Unconventional split aptamers cleaved at functionally essential sites preserve biorecognition capability. *Anal Chem* 2019;**91**:4223–31. https://doi.org/10.1021/acs.analchem.9b04115
56. Debiais M, Lelievre A, Smietana M *et al.* Splitting aptamers and nucleic acid enzymes for the development of advanced biosensors. *Nucleic Acids Res* 2020;**48**:15811–7. https://doi.org/10.1093/nar/gkaa132
57. Gambill L, Staubus A, Mo KW *et al.* A split ribozyme that links detection of a native RNA to orthogonal protein outputs. *Nat Commun* 2023;**14**:543 https://doi.org/10.1038/s41467-023-36073-3
58. Terasaka N, Futai K, Katoh T *et al.* A human microRNA precursor binding to folic acid discovered by small RNA transcriptomic SELEX. *RNA* 2016;**22**:1918–28. https://doi.org/10.1261/rna.057737.116
59. Lorenz R, Bernhart SH, Höner Zu Siederdissen C *et al.* ViennaRNA package 2.0. *Algorithms Mol Biol* 2011;**6**:26. https://doi.org/10.1186/1748-7188-6-26

Received: May 23, 2024. **Revised:** March 25, 2025. **Editorial Decision:** March 31, 2025. **Accepted:** April 4, 2025