Title
Multiplex Aptamer Discovery through Apta-Seq and Its Application to ATP Aptamers Derived from Human-Genomic SELEX

Permalink
https://escholarship.org/uc/item/39g6231q

Journal
ACS CHEMICAL BIOLOGY, 12(8)

ISSN
1554-8929

Authors
Abdelsayed, MM
Ho, BT
Vu, MMK
et al.

Publication Date
2017-08-01

DOI
10.1021/acschembio.7b00001

Peer reviewed
Multiplex Aptamer Discovery through Apta-Seq and Its Application to ATP Aptamers Derived from Human-Genomic SELEX

Michael M. Abdelsayed,‡,‖ Bao T. Ho,‡,‖ Michael M. K. Vu,§,‖ Julio Polanco,† Robert C. Spitale,‡,§ and Andrej Lupták*‡,†,§

‡Department of Molecular Biology and Biochemistry, University of California—Irvine, Irvine, California 92697, United States
†Department of Pharmaceutical Sciences, University of California—Irvine, Irvine, California 92697, United States
§Department of Chemistry, University of California—Irvine, Irvine, California 92697, United States

Supporting Information

ABSTRACT: Laboratory-evolved RNAs bind a wide variety of targets and serve highly diverse functions, including as diagnostic and therapeutic aptamers. The majority of aptamers have been identified using in vitro selection (SELEX), a molecular evolution technique based on selecting target-binding RNAs from highly diverse pools through serial rounds of enrichment and amplification. In vitro selection typically yields multiple distinct motifs of highly variable abundance and target-binding affinities. The discovery of new aptamers is often limited by the difficulty of characterizing the selected motifs, because testing of individual sequences tends to be a tedious process. To facilitate the discovery of new aptamers within in vitro selected pools, we developed Apta-Seq, a multiplex analysis based on quantitative, ligand-dependent 2′ acylation of solvent-accessible regions of the selected RNA pools, followed by reverse transcription (SHAPE) and deep sequencing. The method reveals, in a single sequencing experiment, the identity, structural features, and target dissociation constants for aptamers present in the selected pool. Application of Apta-Seq to a human genomic pool enriched for ATP-binding RNAs yielded three new aptamers, which together with previously identified human aptamers suggest that ligand-binding RNAs may be common in mammals.

This handicaps the description of selected RNAs and is a key hurdle to overcome for the efficient discovery of functional aptamers that may have diverse and important functions.

One application of in vitro selections seeking to characterize all enriched RNAs, including low copy number sequences, is genomic8,9 and transcriptionic SELEX10–13. These experiments reveal RNAs encoded by the genomes of predetermined species and are particularly important in the discovery of new instances of known functional RNAs, such as aptamers and ribozymes.14–16 In contrast to selections based on synthetic sequences, which are typically designed to identify a single (or a few) fittest functional RNA, genomic and transcriptionic selections are designed to map out all instances of a given function, such as binding of a target metabolite or protein. However, the relative abundance of in vitro selected RNAs is biased toward not only sequences that fulfill the selection criteria but also those that are highly amplifiable (sequences that transcribe, reverse-transcribe, and PCR amplify more efficiently than others),14,15 thus the resulting distribution of the functional RNAs in the selected pools may be strongly skewed by these two properties.

Received: January 2, 2017
Accepted: June 29, 2017
Published: June 29, 2017

DOI: 10.1021/acschembio.7b00001
ACS Chem. Biol. XXXX, XXX, XXX−XXX

ACS Publications © XXXX American Chemical Society

Articles

pubs.acs.org/acschemicalbiology

Multiplex Aptamer Discovery through Apta-Seq and Its Application to ATP Aptamers Derived from Human-Genomic SELEX

Michael M. Abdelsayed,‡,‖ Bao T. Ho,‡,‖ Michael M. K.Vu,§,‖ Julio Polanco,† Robert C. Spitale,‡,§ and Andrej Lupták*‡,†,§

‡Department of Molecular Biology and Biochemistry, University of California—Irvine, Irvine, California 92697, United States
†Department of Pharmaceutical Sciences, University of California—Irvine, Irvine, California 92697, United States
§Department of Chemistry, University of California—Irvine, Irvine, California 92697, United States

Supporting Information

ABSTRACT: Laboratory-evolved RNAs bind a wide variety of targets and serve highly diverse functions, including as diagnostic and therapeutic aptamers. The majority of aptamers have been identified using in vitro selection (SELEX), a molecular evolution technique based on selecting target-binding RNAs from highly diverse pools through serial rounds of enrichment and amplification. In vitro selection typically yields multiple distinct motifs of highly variable abundance and target-binding affinities. The discovery of new aptamers is often limited by the difficulty of characterizing the selected motifs, because testing of individual sequences tends to be a tedious process. To facilitate the discovery of new aptamers within in vitro selected pools, we developed Apta-Seq, a multiplex analysis based on quantitative, ligand-dependent 2′ acylation of solvent-accessible regions of the selected RNA pools, followed by reverse transcription (SHAPE) and deep sequencing. The method reveals, in a single sequencing experiment, the identity, structural features, and target dissociation constants for aptamers present in the selected pool. Application of Apta-Seq to a human genomic pool enriched for ATP-binding RNAs yielded three new aptamers, which together with previously identified human aptamers suggest that ligand-binding RNAs may be common in mammals.

This handicaps the description of selected RNAs and is a key hurdle to overcome for the efficient discovery of functional aptamers that may have diverse and important functions.

One application of in vitro selections seeking to characterize all enriched RNAs, including low copy number sequences, is genomic8,9 and transcriptionic SELEX10–13. These experiments reveal RNAs encoded by the genomes of predetermined species and are particularly important in the discovery of new instances of known functional RNAs, such as aptamers and ribozymes.14–16 In contrast to selections based on synthetic sequences, which are typically designed to identify a single (or a few) fittest functional RNA, genomic and transcriptionic selections are designed to map out all instances of a given function, such as binding of a target metabolite or protein. However, the relative abundance of in vitro selected RNAs is biased toward not only sequences that fulfill the selection criteria but also those that are highly amplifiable (sequences that transcribe, reverse-transcribe, and PCR amplify more efficiently than others),14,15 thus the resulting distribution of the functional RNAs in the selected pools may be strongly skewed by these two properties.

Received: January 2, 2017
Accepted: June 29, 2017
Published: June 29, 2017

DOI: 10.1021/acschembio.7b00001
ACS Chem. Biol. XXXX, XXX, XXX−XXX

ACS Publications © XXXX American Chemical Society

Articles

pubs.acs.org/acschemicalbiology
The ability to more quickly characterize selection pools for the binding function would be enabled if the structures of many aptamers and their physical interactions with their targets could be determined in a single experiment. RNA probing with chemical reagents is a robust method for analyzing the structures of RNAs and characterizing important conformational changes that can occur due to interaction with small molecules and proteins; however, such experiments are usually accomplished on single RNA molecules. Information about structure and ligand-binding sites in RNAs can be extracted from experiments based on partial hydrolysis (in-line probing) and chemical modification of RNAs, using selective 2'-hydroxyl acylation (SHAPE) or base modification (e.g., by dimethyl sulfate). The SHAPE method detects 2'-OH accessibility and reactivity to acylation, thereby reporting on which aptamer segments are more flexible and reactive in response to a ligand. Changes in structure can be probed under different environmental conditions or in the presence of varying concentrations of ligands, such as small molecules and proteins, and can be used to extract the dissociation constants for the RNA–ligand interactions. Several methods combine SHAPE with high-throughput sequencing (SHAPE-seq) to achieve single-nucleotide resolution of acylation reactivity on a diverse set of sequences simultaneously and couple the output to computational modules developed to yield genomic locations, intrinsic reverse transcriptase (RT) stops, SHAPE reactivities, and secondary structure models for each transcript. Although useful, most of these efforts have been largely descriptive with few examples of their use for novel biological discovery.

To overcome the limitations of single-sequence characterization of in vitro selected pools, we developed a multiplexed approach to couple RNA selection with structural and binding characterization of individual sequences within the selected pools. We marry selection with chemical probing of RNA structure to reveal the sequence, structural features, and ligand affinities of both dominant and minor species from the same pool. We use this technique, Apta-Seq, to discover and characterize both known and novel adenosine aptamers in the human genome. Our methodology not only increases the rate of novel aptamers discovery for our studied ligand (ATP) but also has the potential to be applied to any ligand-pool pair, thereby greatly enhancing the speed of aptamer discovery and structural characterization.

## RESULTS AND DISCUSSION

In order to establish the identity, secondary structure, and binding properties of aptamers in a single experiment, we combined in vitro selection, SHAPE-seq, and StructureFold into a pipeline (Figure 1), which provides all the information required for aptamer discovery. We applied this method to an in vitro selected pool derived from the human genome and enriched for ATP-binding aptamers, as described previously. We chose adenosine/ATP as the target of the selection, because it has been used extensively over the past two decades, and as such, it is an excellent model system for the development of new in vitro selection and analysis methods. Adenosine aptamers with a conserved motif consisting of an 11-nucleotide binding loop and an opposing bulging guanosine, flanked by two helices (the Sassanfar–Szostak motif), were initially isolated from synthetic, random pools by in vitro selections targeting ATP, nicotinamide adenine dinucleotide, S-adenosyl methionine, and S-adenosyl homocysteine. More recently, a genomic SELEX experiment revealed the same motif in two distinct loci in the human genome; the FGD3 aptamer resides in an intron of the FGD3 gene, and the ERV1 aptamer maps antisense to a junction between an ERV1 LTR repeat and its 3' insertion site. These were the only two aptamers revealed by traditional cloning; however, this approach tests only a small number of sequences and can miss low copy number aptamers, we therefore reanalyzed the pool using high-throughput sequencing and SHAPE to uncover other genomic aptamers.

To test these potential aptamers, we adapted the SHAPE-Seq analysis coupled to StructureFold to the selected human genomic RNA pool at varying ATP concentrations. The pool was transcribed and purified as during the SELEX experiment, divided into 10 fractions, and incubated with ATP at concentrations ranging from 3 μM to 10 mM, with two fractions set aside for control reactions (no-SHAPE and no-ligand SHAPE controls). The samples were acylated by 2-(azidomethyl)nicotinic acid acyl imidazole (NAI-N3), reverse transcribed, and ligated to yield circular single-stranded DNAs, in which the ligation positions correspond to the reverse transcriptase termination points. These occur either due to acylation of the RNA template on the nucleotide immediately upstream of the termination point or at boundaries of RNA structural elements, such as paired segments. The no-SHAPE control reaction reveals the natural RT stops for each sequence in the pool, whereas the SHAPE experiments reveal additional RT stops corresponding to the acylated positions in the RNA, some of which change upon ligand binding (Figure 1B). Amplification of the ssDNA by PCR and barcoding each experimental condition in the ligand titration allow one-pot sequencing of the entire population. Subsequent barcode-based demultiplexing of the sequences reveals all sequences and SHAPE-dependent RT stops for each experimental condition. Nucleotide positions of all detected sequences were analyzed for RT termination and expressed as RT-stop frequency, which is defined as a fraction of total reads (per experiment) for a given sequence (Figure 1C).

The frequency of RT stops was used to obtain SHAPE profiles in the presence of increasing concentration of ATP-Mg
makes direct contacts with the ligand through stacking, and the 2′-OH of the A60 equivalent appears partially solvent-exposed and hydrogen-bonded to the adjacent phosphate, which likely activates it for acylation. The sugar–phosphate backbone of this nucleotide in the adenosine-bound conformation is thus highly sensitive to modification, and in the case of the FGD3 aptamer, the binding loop must undergo a significant conformational change upon ligand binding, because the same position is weakly acylated in the absence of the ligand (Figure 2A, low-ATP traces, shown in shades of red).

To validate the results obtained from Apta-Seq of the pool with the reactivity of the purified FGD3 aptamer, we performed a SHAPE analysis on an isolated clone of the aptamer. One of the positions with the most prominent ATP-dependent changes in SHAPE reactivity is C72 (Figures 2A and S1A), which maps to a domain adjacent to the adenosine binding loop (Figure 2D). We confirmed this result by conventional (PAGE-based) SHAPE analysis of the aptamer clone, which also revealed a strong, ATP-dependent increase in reverse transcription termination at position C72 (Figure 2C), reaffirming that the method reveals comparable data for the same aptamer within a highly heterogeneous pool of sequences. These results demonstrate that our high-throughput sequencing approach parallels more traditional structural analysis normally reserved to single clone analyses.

For the ERV1 aptamer, we derived the SHAPE reactivity using the Reactivity Calculation module of StructureFold, with RT stops from the no-ligand SHAPE data set as the negative control for the reactivity profile of the 10 mM ATP experiment (Figure S1B), and carried out structure predictions using the RNAprobing server of the Vienna RNA Package. In the absence of ATP, the ERV1 structure is not predicted to form the ATP-binding motif, because the binding loop is sequestered within a stem flanked by two bulges. However, in the presence of ATP, the structure rearranges to form the classical ATP binding loop (Figure 2E). The binding loop showed similar acylation trends to those of the FGD3 aptamer, but shifted by one nucleotide: the third guanosine (G16; Figure 2E) of the GFG3 aptamer, the binding loop must undergo a significant conformational change upon ligand binding, because the same position is weakly acylated in the absence of the ligand (Figure 2A, low-ATP traces, shown in shades of red).

Table 1. Relative Abundance and Binding Affinity of Selected Aptamers

<table>
<thead>
<tr>
<th>aptamer</th>
<th>rel. abundance (%)</th>
<th>$K_D$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGD3</td>
<td>0.3839</td>
<td>0.7 ± 0.1</td>
</tr>
<tr>
<td>ERV1</td>
<td>0.4893</td>
<td>1.8 ± 0.6</td>
</tr>
<tr>
<td>PRR5</td>
<td>0.1742</td>
<td>1.7 ± 0.6</td>
</tr>
<tr>
<td>LIPA16</td>
<td>0.0848</td>
<td>1.3 ± 0.1</td>
</tr>
<tr>
<td>THE1B</td>
<td>0.1191</td>
<td>1.1 ± 0.4</td>
</tr>
</tbody>
</table>

*Relative abundance for each aptamer was calculated out of 1 533 526 sequences. Reported errors are standard errors of the mean of $K_D$s calculated from binding curves with $R^2 > 0.5.$
is dominated by sequences corresponding to short primer extensions of the reverse primers and are thus difficult to map to the genome uniquely.

The Apta-Seq pipeline also yielded novel, less abundant sequences not found in previous analyses. Our previously reported aptamers were discovered in an ATP column binding analysis of individual clones to find potential candidates to undergo structural analysis; Apta-Seq condenses the process to gain a large-scale analysis of the pool in one experiment in solution. Surprisingly, three new sequences also contained the Sassanfar−Szostak motif. The PRR5 aptamer maps to the second intron of the PRR5 gene or the first intron of the PRR5-ARHGAP8 fusion protein (Figure 3A). PRR5 codes for a protein that is part of the mTORC2 complex and, like the FGD3, plays important roles in pathways that regulate cell growth, but the significance of the ATP/adenosine-binding aptamers in their introns remains unknown. The PRR5 aptamer bound to ATP-agarose beads and eluted in the presence of free ATP (Figure 3C) and Apta-Seq data revealed a $K_D$ of $\sim 1.7$ mM (Figure 3B and Table 1) at a relative abundance of 0.17% among the mapped sequences. Interestingly, mutations (including a 14-nt insertion) of the aptamer sequence throughout the genomes of primates preserve the aptamer structure, suggesting that it may be a functional RNA in primates (Figure 3E and F).

The second aptamer (Figures 4A and S2) was found in a long interspersed element (LINE), L1PA16, and the third (Figures 4B and S3) maps antisense to a repeat element THE1B, which is derived from a subfamily of ERV Mammalian apparent LTR-retrotransposons (ERV-MaLR), and may be stabilized by fortuitous base-pairing with a part of the sequence derived from the forward primer of the pool (Figure 4B).

Figure 3. A novel adenosine aptamer, mapping to the PRR5 gene in primates. (A) Location of the aptamer (red line) in the PRR5 intron. The aptamer also appears in the first intron of a PRR5−ARHGAP8 fusion. (B) Trace of acylation for the PRR5 aptamer with varying ATP concentrations, ranging from 0 (dark red) to 10 mM (dark blue). The conserved adenosine-binding motif is underlined in red, and positions with ATP-dependent changes in RT stops are indicated. (C) Graph of column binding fractions of the PRR5 sequence amplified out of the selected pool. The graph displays 18% of RNA eluting with free 5 mM ATP. (D) Binding isotherms of the aptamer modeled from the change in RT stops at positions indicated in the legend, yielding an average $K_D$ of $1.7 \pm 0.6$ mM (average deviation). (E) Predicted secondary structure of the PRR5 aptamer with ATP binding loop shown in red. Boxed nucleotides indicate aptamer positions that show an increase (black outlines) or decrease (gray) in acylation with increasing ATP. Black triangle indicates sequence-insertion site for some primates. (F) Aptamer sequence conservation among primates. All mutations within the aptamer motif are conservative with respect to base-pairing interactions of the proposed structure: U49C mutation leads to a G•U wobble pair to be replaced with a canonical G•C pair in helix 5 (h5). Insertion at A54 (black triangle) extends the sequence into the loop of h5 and potentially extends the helical domain by four base-pairs, and the A81G mutation in h4 creates a G•U wobble pair from an A•U canonical base-pair.
Articles

Figure 4. Novel adenosine aptamers revealed by Apta-Seq in human retrotransposons. (A) The Sassanfar–Szostak motif mapping to the L1PA16 LINE1 element, with ATP-dependent acylation positions indicated by squares and annotated as in Figures 2 and 3. Mutation of a key binding site residue (G26A) results in no ATP binding. (B) An aptamer mapping to a THE1B retrotransposon. Part of the 5′ sequence that originated from the forward primer-binding region of the DNA pool is shown in light blue. A mutation of an essential binding site residue (G24U) abolished ATP binding. Binding isotherms of the L1PA16 (C) and THE1B (D) aptamers modeled from the change in acylation at positions indicated in the legend, yielding a \( K_D \) of 1.3 ± 0.1 mM and 1.1 ± 0.4 mM, respectively. Graph of ATP column binding of the L1PA16 (E) and THE1B (F) sequences amplified out of the selected pool, showing robust activity of the purified sequences and validating the Apta-Seq data. Full SHAPE-Seq profiles and secondary structure models of these aptamers are presented in Supporting Information Figures S2 and S3.

\( K_D \)s derived from Apta-Seq (Figure 4C and D) were 1.3 mM and 1.1 mM (Table 1); however, the SHAPE profile of the THE1B aptamer afforded only a single peak (G32; Figures 4D and S3A) that yielded a \( K_D \) model with a good fit (\( R^2 \sim 0.95 \)), whereas the L1PA16 aptamer exhibited several peaks from which the dissociation constants could be derived (Figures 4C and S2A). When transcribed from individual DNA templates and purified using PAGE, both aptamers bound to ATP-agarose beads and eluted in the presence of free ATP (Figures 4C and 3D). Mutations of key residues (Figure 4A and B) in the binding loops of these aptamers abolish binding to ATP columns (Figures S2C and S3C). These aptamers, together with the previously discovered adenosine\(^{15}\) and GTP\(^{16}\) aptamers, indicate that ligand-binding RNAs are likely common in higher eukaryotes.

To analyze these Sassanfar–Szostak motifs for potential binding in the context of the human transcriptome, we isolated total RNA from four cell culture experiments (SHSY-5Y cells incubated with DMSO or DMSO + 10 mM adenosine and OV90 and MCF7 cells) and treated it in the same manner as during the SELEX experiment. The total RNA samples were annealed in the binding buffer, introduced to the ATP beads, washed with five column volumes of the binding buffer, and eluted with 5 mM ATP-Mg. RNAs isolated from the last wash fraction and ATP elution were reverse-transcribed using primers specific for the five aptamers described above and analyzed using nested primers by qPCR. Two of the aptamers, FGD3 and THE1B, were detected at higher levels in the elution fractions than in the last washes in the SHSY-5Y cells. FGD3 aptamer exhibited more robust binding to ATP beads and the levels of the aptamer appear insensitive to the presence of adenosine in the tissue culture medium (Figure S1B). In contrast, the THE1B aptamer showed significantly higher levels in the ATP elution fraction only in RNA extracted from SHSY-5Y cells incubated with 10 mM adenosine (Figure S3D and E).

These experiments suggest that at least two of the aptamers described herein have the capacity to bind ATP within the context of their endogenous transcripts and the human transcriptome, and that their expression or activity can be modulated by exogenous adenosine.

In summary, we describe an efficient process of multiplex analysis of in vitro selected RNA pools. An in vitro selection experiment is combined with SHAPE-Seq and StructureFold analysis to efficiently and quantitatively analyze the selected pools at a single-nucleotide resolution. A number of high-throughput methods, many of which can be adapted for application in aptamer discovery, have recently been developed to analyze RNA–protein interactions;\(^{38-40}\) however, all of these techniques require immobilization of the nucleic acid on the surface of a sequencing chip and a labeled target molecule. Another recent application of high-throughput analysis of in vitro selected pools (of mRNA-displayed peptides) provides binding parameters for the target interactions, but the method requires the target molecule to be immobilized on solid support and relatively slow binding and dissociation kinetics.\(^{41}\) None of these techniques reveal information about conformational changes associated with target binding but can provide secondary structure constraints thorough covariation analysis of the active RNAs. Moreover, these techniques derive equilibrium binding constants from kinetics and presume an absence of rate-limiting conformational changes associated with target binding, whereas Apta-Seq detects target binding under equilibrium conditions. Apta-Seq makes it significantly easier to discover, sort, and characterize aptamers by measuring structural changes in RNAs in solution, providing a straightforward way to measure affinity for the target molecules. Importantly, Apta-Seq is a powerful enabling technological pipeline that is sure to expedite the transition from aptamer selection to the unraveling of their structure and binding affinity and thus their biological or biotechnological relevance. Furthermore, because Apta-Seq is performed in solution, the method can be used with a large number of targets, providing a label-free approach to studying specificity of the selected sequences. In contrast, high-throughput methods based on immobilization of the RNAs or their binding partners either require modification (immobilization or fluorescent labeling) of each target to measure the binding kinetics or only measure the rates of association, but not dissociation, when unlabeled off-target molecules are introduced in competition with labeled targets. Our results show that ATP binding by the Sassanfar–Szostak motif is in some cases coupled to significant remodeling of the RNA structure in adjacent domains. The method thus reveals not just ligand binding but also concomitant large-scale conformational changes, facilitating multiplexed experimental discovery of potential riboswitches.

**MATERIALS AND METHODS**

**Transcription.** RNAs were transcribed for 2 h at 37 °C in 400 μL of a solution containing 40 mM Tris chloride; 10% dimethyl sulfoxide (DMSO); 10 mM dithiothreitol (DTT); 2 mM spermidine; 5 mM...
each rCTP, rGTP, rUTP, and rATP; 20 mM MgCl₂; one unit of T7 RNA polymerase; and ~0.5 μM DNA template. Transcripts were purified by 7% polyacrylamide gel electrophoresis (PAGE) under denaturing conditions (7 M urea). RNA was eluted from the gel into 400 μL of 400 mM KCl and precipitated by adding 800 μL of 100% ethanol at -20 °C.

**Primer Phosphorylation.** Primer labeling was prepared in a total volume of 20 μL. Then, 20 μM of primer, 1× T4 Polynucleotide Kinase (PNK) ligase buffer (NEB), 1 unit of T4 PNK (NEB), and 0.5 μCi [γ-32P]ATP was incubated at 37 °C for 1 h then purified from denaturing PAGE.

**Synthesis of the SHAPE Reagent.** The SHAPE reagent, 2-(azidomethyl)nicotinic acid acyl imidazole, was synthesized following a previously described protocol.32

**Selective 2'-Hydroxy Acylation and Primer Extension (SHAPE).** SHAPE reactions were prepared in a total volume of 10 μL. RNAs (pools or individual aptamer sequences) were resuspended in water and heated to 70 °C for 3 min. Then, 1 μM of purified RNA was added to a buffer containing 140 mM KCl, 10 mM NaCl, 10 mM tris chloride at pH 7.5, and 5 mM MgCl₂. A dilution series of 5'-adenosine triphosphate (ATP) was prepared using a 1:1 stock of ATP·Mg₂⁺. Then, 1 μM to 10 μM ATP was aliquoted to the RNA in buffer and incubated at RT (~23 °C) for 1 min. Then, 50 mM 2-(azidomethyl)nicotinic acid acyl imidazole was added to the mixture, and the reaction was incubated for 45 min at RT. Reactions containing no SHAPE reagent were substituted with 10% DMSO. Reactions were precipitated with 10 μL 3 M KCl, 1 μL glycoal, 89 μL H₂O, and 300 μL 98% ethanol.

** Primer Extension.** The RNA pellet was reconstituted in a 10 μL reaction volume containing 0.1 μM 5'-[32P]-radiolabeled reverse transcription DNA primer, 2 μL of 5 × M-MuLV Reverse Transcriptase Reaction Buffer (New England Biolabs), 1 unit of M-MuLV enzyme, and 500 μM each of deoxyribonucleotide triphosphate. Extensions were performed at 42 °C for 15 min. 400 μM NaOH was added, and the reaction was incubated at 95 °C for 5 min to hydrolyze the RNA. Reactions were precipitated with 10 μL of 3 M KCl, 1 μL of glycoal, 89 μL of H₂O, and 300 μL of 98% ethanol. Complementary DNA (cDNA) was resolved using 12% denaturing PAGE or amplified for high-throughput analysis.

**K₀ Analysis.** RT stops were normalized by dividing peak intensities in SHAPE profiles by the total number of reads per aptamer and experimental conditions. These normalized profiles (presented as overlays in Figures 2 and 3 as well as S1 overlays in SHAPE plots) were analyzed by 7% polyacrylamide gel electrophoresis (PAGE) under denaturing conditions (7 M urea). RNA was eluted from the gel into 10 μL of 100% ethanol and heated to 70 °C for 3 min. Then, 20 μL of glycoblue, 89 μL of H₂O, and 300 μL of 98% ethanol. Complementary DNA (cDNA) was resolved using 12% denaturing PAGE or amplified for high-throughput analysis.

**Fraction Bound = (ligand) / (ligand + K₀ Baseline) **

where “Range” corresponds to the range of values above baseline that a peak in the SHAPE profile can assume. The model was fit to the data using a linear least-squares analysis and the Solver module of Microsoft Excel to extract ATP K₀ for each peak. Because some of the apparent binding constants are near the maximum of the titration and in many cases the rising SHAPE signal did not level off at the highest ATP concentration, we predominantly used the positions where the SHAPE signal decreased with ATP concentration and approached zero for K₀ modeling.

**SHAPE-Seq Library and Primer Design.** Round 6 of an in vitro selection for an ATP aptamer from a human genomic library was used as the library for Apta-Seq. Libraries were given individual barcodes based on concentration of the ligand used during SHAPE. Ten libraries in total were made using a ligand titration of 1–10 000 μM, no-ligand control, and no-SHAPE (only DMSO, as described above) control.

Apta-Seq primers contain a reverse primer sequence for the pool of interest and flanking Illumina primers in order to barcode and sequence primer extensions by Illumina Sequencing. Apta-Seq primers for reverse transcription primer extension were designed 5’ to 3’ with the following components:

<table>
<thead>
<tr>
<th>Illumina forward primer</th>
<th>No-ligand control</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-AGATCGGAGAAGCTGTACTAACG-3’</td>
<td>5'-AGATCGGAGAAGCTGTACTAACG-3’</td>
<td>5'-CAACGAGAAGGCGACCGATAGAT [barcode]</td>
</tr>
</tbody>
</table>

**Primer Extension for Apta-Seq.** Primer extension was carried out with the Apta-Seq primer as described above. cDNA was self-ligated in a 20 μL reaction using CircLigaseII Reaction Buffer (Epicerin), 2.5 mM MmC1, 50 μM ATP, and five units of CircLigase ssDNA Ligase. Reactions were precipitated with 10 μL of 3 M KCl, 1 μL of glycoal, 89 μL of H₂O, and 300 μL of 98% ethanol, and cDNA was reconstituted in 20 μL of H₂O. A polymerase chain reaction (PCR) was performed using 1 μM each of forward and reverse primers, cDNA template, and DreamTaq Master Mix (Thermo Fisher) and amplified for 16 cycles (denaturing 94 °C, 30 s, annealing 55 °C, 30 s, and elongation 72 °C, 30 s). Amplicons were sequenced on Illumina HiSeq 2500 at the UCI Genomics Facility.

**SHAPE-Seq Reactivity Mapping.** Galaxy (https://usegalaxy.org/galaxy) and the StructureFold module were used to map and determine the SHAPE reactivity of aptamers. Forward reads of libraries were used to analyze RT stops. Adapters and primers (CAATGCCTCAAG) were clipped using Clip adapter sequences on Galaxy. Default settings were used except for a minimum sequence length of 10 and an output of both clipped and nonclipped sequences. Clipped libraries were then processed using StructureFold, a series of Web-based programs to characterize RNA that have undergone SHAPE modifications. Aptamers were mapped to selected libraries using Iterative Mapping on Galaxy. Default settings for analysis were used except for a minimum read length of 12 nucleotides and three mismatches allowed (-v 3). RT stop counts were calculated using Get RT Stop Counts on Galaxy. RT stop counts were derived using mapped files from Iterative Mapping and aptamers as the reference sequences. Output from the Get RT Stop Counts module was tabulated and normalized to the total stop counts for each aptamer. The percentage of counts for each position was divided by each position by the total number of RT counts for each aptamer.

The new aptamers were mapped to the following sequences obtained from Illumina sequencing:

- L1PA6
- cctacagtGCTTACTGACCTGAGAGATCAACTTCAAGGCGACGCGTAAGTTTTGCTTCTCTGAGGAGTAGATGGAG
- THE1B
- gggcgacgtgctctactaGTTTGGGAGGGAGCACCTGGTGAGGAGGAGTTACTGGAAGATGAGGCTCTCCGTACTGTCTGACTATGAGATCTGAGCTTATAGGTTGAGTCTTTCCCTCTGCAAACTCTAg
- PRRS
- ggtGAGAGGAGGCTCTGAGCCTGCAGGTCAGGGGTGCAGGGGTTCCTACCACTGGTCTGACTCTGCTGACCTGTACTGACCTGGGACACCTCTCTACCTGGGGAGCTTCAGCTCACTCACTACAGGGCAGCAGGCCTGG
- GAGAAGACCTGCAGCGGCAGCTCATGACTGAGGCTTGTTACGTCTGCTCTGCCTGAGATCTAGGATG
Structure Prediction. The Reactivity Calculation module on Galaxy was used to obtain SHAPE reactivities by using the output from the Get RT Stop Counts module. The RT stop counts for 10 mM ligand concentration was used as the (+) library and no ligand for the (−) library, both in the presence of the SHAPE reagent. Default settings were used except for nucleotide specificity, which was changed to AUCG. To predict structure, the RNAprobing Web server was used. All default settings were used according to the Washietel et al. SHAPE method.36

Total RNA Extraction. Human cell lines OV90 and MCF7, ovary- and breast-derived adenocarcinoma, respectively, were thawed from cryo-preservation and seeded onto separate T-75 culture plates with DMEM media containing 10% Fetal Bovine Serum (FBS), 10% amphotericin B, and 10% penicillin/streptomycin. Similarly, SHSY-SY neuroblastoma cells were thawed and seeded in two separate T-75 flasks with DMEM/F12 culture media containing 10% FBS, 10% amphotericin B, and 10% penicillin/streptomycin. All cell cultures were passaged appropriately to achieve 80% confluency on the day of the experiment. In one SHSY-SY experiment, 2% DMSO with 10 mM adenosine was added to adhered confluent SHSY-SY cells 1 day prior to total RNA extraction.

Total RNA was harvested from all cell types using TRIzol Reagent (Ambion). All total RNA isolation steps were performed according to the user manual. Briefly, adhered cells were collected by washing each culture dish with 750 μL of TRIzol Reagent and collected into a fresh 1.5 mL microcentrifuge tube. A total of 200 μL of chloroform was added to each sample to allow for phase separation. Extraction of the aqueous layer was followed by RNA precipitation using 100% ethanol, pelleted by centrifugation, washed with 75% ethanol, and resuspended in a 1× H2O. Once resuspended, total RNAs were treated with DNase I to remove residual DNA contaminants before column binding and RT-qPCR.

Column Binding Assay Using Total RNA Extracts. Prior to RNA binding, 10 μL of C8-linked ATP-agarose (1 mg mL⁻¹ Sigma-Aldrich) was pre-equilibrated on a spin-filter with binding buffer (BB) containing 140 mM KCl, 10 mM NaCl, 5 mM MgCl₂, and 20 mM tris-HCl, at pH 8.0. For each tissue culture source (OV90, MCF7, SHSY-SY, and SHSY-SY incubated with 10 mM adenosine), 500 ng of total RNA was annealed in the presence of BB at 72 °C for 1 min and allowed to cool to RT over 5 min before addition onto ATP-agarose beads. A total of 11 fractions were collected starting with the flow-through, which was collected after a 30 min incubation. After flow-through collection, a total of four washes were collected over a span of 10 min using 10 μL of BB. After washing, four 30 min elution fractions were collected using an elution buffer (EB) containing 5 mM ATP, 140 mM KCl, 10 mM NaCl, 10 mM MgCl₂, and 20 mM tris-HCl, at pH 8.0. To remove any remaining RNAs attached to the column, one 10 μL denaturing wash was collected using 7 M urea following the elution fractions. Each fraction was precipitated using 300 mM KCl and 100% ethanol, pelleted by centrifugation, washed with 75% ethanol, and resuspended in ddH₂O for reverse transcription.

RT-qPCR. Pellets from the last (fourth) wash and the elution fractions were resuspended in a 1× reverse transcription master mix containing 500 nM of aptamer-specific RT primers, 5 mM dNTPs, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, and 50 mM tris-HCl at pH 8.3. Resuspended RNAs were heat-treated at 90 °C for 30 s, followed by a 50 min incubation at 50 °C after the addition of superscript III reverse transcriptase. cDNA products were diluted 10-fold with nuclease-free H₂O, and quantitative PCR was performed on a BioRad CFX Connect system using 1 μL of cDNA product, 500 nM of gene-specific primers, and BioRad iTaq supermix. The reverse primers were 3′-extended versions of the RT primers to ensure gene-specific amplification.

**REFERENCES**

SELEX) for the identification of protein-binding RNAs independent of their expression levels. Nat. Protoc. 1, 2204–2212.


