



Published in final edited form as:

*Chem Commun (Camb)*. 2017 March 02; 53(19): 2878–2881. doi:10.1039/c6cc09991b.

## DNA display of folded RNA libraries enabling RNA-SELEX without reverse transcription

I. S. MacPherson<sup>a,b</sup>, J. S. Temme<sup>b</sup>, and I. J. Krauss<sup>b</sup>

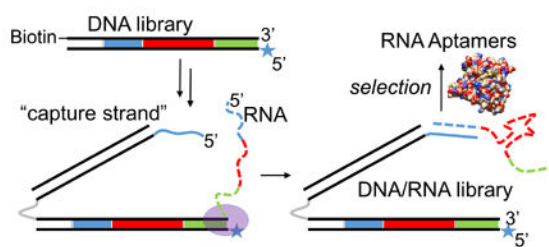
<sup>a</sup>Department of Tropical Medicine, Medical Microbiology and Pharmacology, John A. Burns School of Medicine, University of Hawai`i at Manoa, 651 Ilalo St., Biosciences Building, Suite 325, Honolulu, Hawaii 96813-5525

<sup>b</sup>Department of Chemistry, Brandeis University, 415 South St. MS 015, Waltham, MA 02454-9110, USA

### Abstract

A method for the physical attachment of folded RNA libraries to their encoding DNA is presented as a way to circumvent the reverse transcription step during systematic evolution of RNA ligands by exponential enrichment (RNA-SELEX). A DNA library is modified with one isodC base to stall T7 polymerase and a 5' "capture strand" which anneals to the nascent RNA transcript. This method is validated in a selection of RNA aptamers against human  $\alpha$ -thrombin with dissociation constants in the low nanomolar range. This method will be useful in the discovery of RNA aptamers and ribozymes containing base modifications that make them resistant to accurate reverse transcription.

### Graphical abstract



First characterized in 1990 independently by the laboratories of Szostak and Gold, oligonucleotide aptamers are DNA or RNA with high affinity for proteins or small molecules.<sup>1–4</sup> Exhibiting advantages such as affinities comparable to those of antibodies, simplicity of synthesis, and general lack of immunogenicity, aptamers have found a place in the pharmaceutical market. The first aptamer drug approved by the FDA, Pegaptanib, targets age-related macular degeneration by binding to vascular endothelial growth factor (VEGF).<sup>5–10</sup> Several other aptamers are in late-stage clinical trials, and an unknown number

are in laboratory development.<sup>5–9</sup> Aptamers are typically discovered using standard molecular biology techniques through an *in vitro* selection process termed SELEX (Systematic Evolution of Ligands by EXponential enrichment).<sup>4</sup> DNA libraries with random regions flanked by constant regions are synthesized using standard phosphoramidite chemistry and can be used directly in the selection of DNA aptamers. RNA libraries are generated by transcription of a T7 promoter-containing random DNA library using T7 RNA polymerase. Libraries are mixed with a target small molecule or protein and the bound fraction is recovered using a variety of isolation techniques. The recovered library is amplified with the polymerase chain reaction (PCR) for DNA aptamers or reverse transcription-PCR for RNA aptamers, the single-stranded DNA or RNA is regenerated from the PCR product and the selection cycle is repeated until the library is sufficiently enriched with binders.

Limitations to standard aptamer libraries (nuclease sensitivity, limited chemical diversity) have been overcome by the use of non-natural nucleotide analogs.<sup>11–17</sup> For instance, the incorporation of 2'-fluoro pyrimidines into RNA by T7 RNA polymerase variants results in significant serum nuclease resistance, a major requirement for drug development. To improve chemical diversity, base-modified DNA aptamers have been readily utilized, largely owing to the discovery that family B DNA polymerases (including Vent, KOD, and Pfu) can accommodate C5-substituted thymidine base analogs.<sup>3, 11–15, 18–22</sup> In particular, SOMAmer technology, which incorporates short hydrophobic groups at the C5 position of uridine, improved the success rate of obtaining high-affinity aptamers from 30% to 80% for hundreds of targets.<sup>20</sup> Recently, our group reported a method termed SElection of Modified Aptamers, or SELMA, allowing for the incorporation of large modifications into DNA libraries, which was successfully used to obtain multivalent glycoclusters that mimic a conserved epitope on the HIV envelope protein gp120.<sup>23–26</sup>

However, compared to DNA, successes with base-modified RNA aptamer libraries have not been as wide-spread, despite the superior serum nuclease resistance of 2'-F-modified RNA and greater folding space sampling of RNA libraries. An obstacle to development of base-modified RNA-SELEX is that it would require that two different types of enzymes (RNA polymerase and reverse transcriptase) tolerate the modified bases. Here we present the successful generation and selection of RNA libraries in which the folded RNA is physically attached to the dsDNA that encodes it. The concept of displaying a difficult-to-amplify oligonucleotide on analogous dsDNA was introduced by Szostak and coworkers,<sup>27</sup> and has been put into practice by several groups.<sup>23, 28, 29</sup> The present method circumvents the need for reverse-transcription in the amplification of RNA libraries and could be applied to base modifications for which reverse transcription is inefficient. We propose that this method could be useful in the future for introduction of alkyne-modified RNA analogues and post-transcriptional click modification in an RNA-version of SELMA.

We hypothesized that a DNA library could be designed such that each DNA duplex could capture its corresponding RNA transcript on a “capture arm” (Figure 1, F–H). Generation of this library starts with singly biotinylated, double stranded DNA (Figure 1, A), from which the coding strand is removed using streptavidin magnetic beads. A capture oligonucleotide is annealed to the non-coding strand (B, C) and extended using Bst 2.0 DNA polymerase

(DNAP) to regenerate the double stranded DNA (dsDNA) with an 18 atom hexaethyleneglycol (HEG) spacer (gray) and 124-base single-stranded DNA (ssDNA) extension 5' of the coding strand (**D**). The 5' end of the capture strand contains 21 bases (5' GCT CGT TCT CCT TCC CTC TCC, shown in blue) that complement the 5' end of the RNA strand encoded in the DNA library. An 82-base oligonucleotide is annealed to the capture strand, rigidifying it while maintaining ssDNA at the 5' end (**E**). T7 RNA polymerase (RNAP) and nucleotide triphosphates are then added to initiate transcription of the DNA library. The 5' end of the non-coding strand contains an unnatural base, isodC, which causes T7 polymerase to stall<sup>30</sup> at the end of the template (**F**, **G**), allowing enough time for the 5' ends of the capture strand and RNA transcript to anneal "intrastructurally" (**G**). The T7 RNAP dissociates and the DNA/RNA library (**H**) is amenable to selection. Amplification of the selected library with a biotinylated forward primer and 5' isodC reverse primer regenerates the dsDNA library in its original form (**A**), completing a single selection cycle. The rigidifying oligonucleotide serves to increase the likelihood that the RNA and capture strand will come into contact and anneal. Within the capture strand, the HEG spacer is critical for maintaining the ssDNA nature of the 5' capture sequence; without this spacer, the 3' end of non-coding strand would be extended in step (**C**→**D**), creating a full-length complement of the capture arm that would block its ability to anneal to RNA.

A considerable amount of effort went into the design of the sequences of the capture arm of the DNA display construct. Notably, predicted secondary structure was minimized by trial-and-error with the use of mFOLD prediction.<sup>31</sup> Minimization of secondary structure allowed for annealing of the 82-base rigidifying oligonucleotide at low temperature (50° C) despite its high predicted melting temperature ( $T_m \sim 75^\circ \text{C}$  at 1 nM concentration). This is a crucial requirement because denaturation of the double stranded DNA library at higher temperatures will destroy the integrity of the library. The capture sequence itself (GCTCGTTCTCCTTCCCTCTCC) was designed to have minimal secondary structure as well as a high  $T_m$  with RNA. A key feature for high  $T_m$  is the lack of A-U base pairs, which are known to destabilize RNA-DNA hybrids considerably.<sup>32</sup> As a result, we were able to design a capture sequence with  $T_m$  of  $\sim 70^\circ \text{C}$  (at 1 nM concentration), a strong non-covalent linkage between the RNA and encoding DNA, capable of forming at the 37 °C transcription temperature.

Figure 2 shows the various library species, as visualized in native (A) and denaturing (B) PAGE gels. For lanes 5–7, comparison of the two gels shows clearly that bands in the native gel corresponding to Figure 1 species C, D, and E each separate into two or three strands of the appropriate size in the denaturing gel. The lane 8 band highlighted with a box in the native gel was hypothesized to be the DNA-displayed RNA species. To confirm this assignment, the band was excised from the native gel, and the material extracted from the gel slice was rerun on a denaturing gel (lane 10), giving rise to four bands matching the separate component strands. Extra bands in lane 8 show that library generation results in additional species, which we hypothesized to be excess RNA transcripts and other species resulting from transcription that do not anneal to the library. This was confirmed (lanes 9) by a control transcription of the DNA library lacking a capture arm (Figure 1, form **A**), which

produced most of the bands present in Lane 8, but not the one attributed to the self-annealed RNA-DNA library, species **G** (boxed band).

To verify that the RNA transcript was being captured “intrastructurally” on its own DNA, with sufficient fidelity for selections, we performed a proof-of-principle experiment to show that a sequence with known binding activity could be enriched from a spiked library. To this end, we generated a random RNA-displayed DNA library spiked with 1:1000 of an NheI restriction site sequence. After incubation of the library with a biotinylated DNA oligo complementary to the NheI sequence, the bound fraction was isolated with streptavidin magnetic beads and binders were amplified and regenerated. After two rounds of selection, a significant proportion of the recovered DNA contained the selected sequence (as evidenced by its NheI restriction susceptibility), indicating that the RNA was correctly captured by the DNA encoding it (Figure S1). To control for the possibility that the NheI-cleavable PCR product resulted from traces of the bait sequence contaminating the sample, a parallel selection was performed with the transcription step omitted. In this negative control, more PCR cycles were required to amplify selected library, and the resulting library was not susceptible to NheI cleavage.

With this evidence that our RNA library was effectively displayed on its encoding DNA, we attempted the more rigorous test of our system by using it to select RNA sequences that could bind to a protein. As our target, we chose human  $\alpha$ -thrombin, a commonly used target in proof-of-principle studies.<sup>33</sup> The library described above was prepared without the added NheI cleavage sequence and was then panned for binding to biotinylated thrombin over 10 rounds of selection with capture on streptavidin beads. Enrichment for binders was evident after the 7th round as indicated by the decreased number of PCR cycles required to recover the library. Stringency was then increased by lowering the thrombin concentration from 10 to 1 nM and shortening incubation time from 1 hour to 5 minutes; after the 10th round of selection the library was cloned and 10 members sequenced and analyzed for thrombin binding. As Table 1 indicates, all but two of the isolated sequences showed similarity in both sequence and predicted secondary structure.

Thrombin binding affinity was measured for the corresponding RNAs of the 10 clones in nitrocellulose filter binding assays. For the eight related clones, dissociation constants (Table 1) ranged from 1.4 to 44.7 nM. Interestingly, the predicted RNA folds for these sequences, as determined by mFOLD, were strikingly similar to each other and to a previously isolated thrombin RNA aptamer for which a co-crystal structure (PDB ID 3DD2) has been obtained.<sup>34, 35</sup> A prominent feature of the aptamer fold in the crystal structure is adenosine stacking, where the aptamer contains a stem motif followed by the bulge sequence AACA opposite a single adenosine, followed by another 3-base pair stem, 8 base loop structure. Two adenosines (ACA) in the bulge sequence (red) sandwich an adenosine (red) from the loop CUGAAGUA. The fourth adenosine also stacks against Arg233 of  $\alpha$ -thrombin. In the aptamers obtained in the present study, a similar motif is found, where there is a bulge containing an AAXA opposite a lone adenosine (with the exception of clone 6, which contains AAXG opposite an adenosine), followed by a 3-base-pair stem, 6 base loop with a terminal adenosine in the otherwise variable loop. To confirm that our aptamers bind thrombin through their similar motif, we synthesized a truncated variant of clone 9,

containing only the portion highlighted in the red box (Figure S4, ESI). Nitrocellulose filter binding and biolayer interferometry (BLI) measurements confirmed tight binding to thrombin (2 and 0.4 nM, respectively). Together, these data suggest that our aptamers adopt a conformation similar to those previously reported, and bind to the same region of thrombin, exosite II.

Little or no binding was detected for the remaining two sequences, clones 5 and 7. The predicted fold of clones 5 and 7 (Figure 3) is such that the 5' capture strand complement region (green) engages in extensive base pairing, which would be predicted to inhibit the capture of the RNA by the DNA and therefore render the DNA display ineffective for these clones. Because library generation produces excess free RNA transcripts, abundant thrombin-binding free RNA present in later rounds may anneal to DNA capture strands belonging to sequences such as 5 and 7 that are unable to self-capture, thereby rescuing them. Alternative explanations for the presence of non-binding sequences are that they are simply particularly efficient PCR substrates, or that they might bind to thrombin in a manner that is dependent on the presence of their DNA tag.

In summary, we have described and verified a method for DNA display of RNA. In the future, the benefits of using RNA in selections (structural diversity, amenability to 2' modifications for nuclease resistance) can be coupled with substantial post-transcriptional modification in a SELMA-type experiment.<sup>23</sup> Applications other than SELMA may also benefit from DNA display of RNA. For instance, it will be useful in RNA-SELEX using unnatural NTPs that are accepted by T7 RNA polymerase, but then render the RNA a poor template for reverse transcriptase. It should be noted that the display method described herein uses commercially available enzymes and DNA oligonucleotide synthesis, and an entire library generation and selection cycle can be performed in a day. Further applications of this method for *in vitro* selections will be reported in due course.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

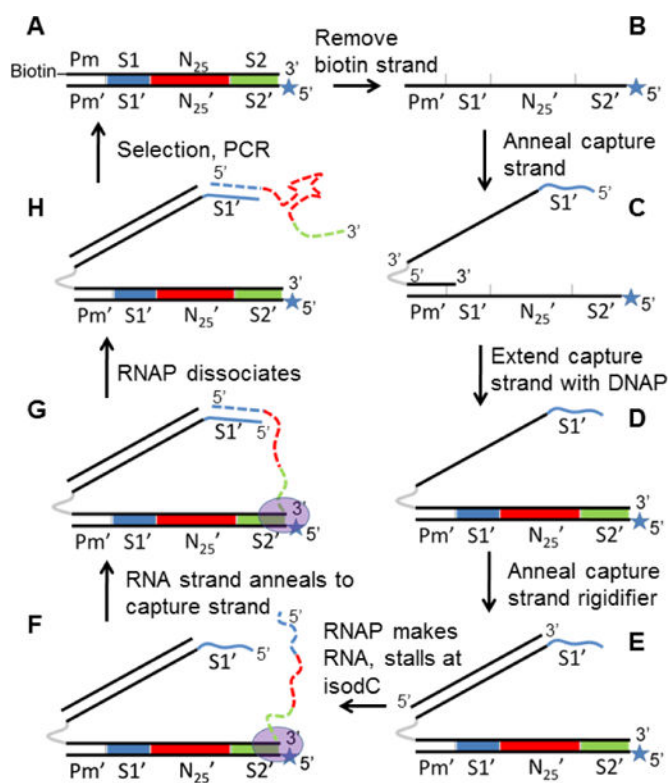
The generous support of the National Institutes of Health (NIH/NIAID R01AI090745) is gratefully acknowledged.

## Notes and references

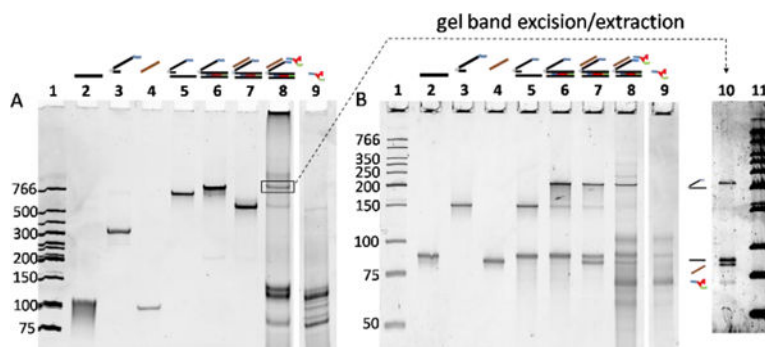
1. Ellington AD, Szostak JW. *Nature*. 1990; 346:818–822. [PubMed: 1697402]
2. Tuerk C, Gold L. *Science*. 1990; 249:505–510. [PubMed: 2200121]
3. Gold L, Janjic N, Jarvis T, Schneider D, Walker JJ, Wilcox SK, Zichi D. *Cold Spring Harb Perspect Biol*. 2012; 4
4. Stoltenburg R, Reinemann C, Strehlitz B. *Biomol Eng*. 2007; 24:381–403. [PubMed: 17627883]
5. Kanwar JR, Roy K, Maremanda NG, Subramanian K, Veedu RN, Bawa R, Kanwar RK. *Curr Med Chem*. 2015; 22:2539–2557. [PubMed: 25723512]
6. Keefe AD, Pai S, Ellington A. *Nat Rev Drug Discov*. 2010; 9:537–550. [PubMed: 20592747]
7. Sundaram P, Kurniawan H, Byrne ME, Wower J. *Eur J Pharm Sci*. 2013; 48:259–271. [PubMed: 23142634]

8. Sun H, Zhu X, Lu PY, Rosato RR, Tan W, Zu Y. *Mol Ther Nucleic Acids*. 2014; 3:e182. [PubMed: 25093706]
9. Santosh B, Yadava PK. *Biomed Res Int*. 2014; 2014:540451. [PubMed: 25050359]
10. Ruckman J, et al. *J Biol Chem*. 1998; 273:20556–20567. [PubMed: 9685413]
11. Gold L, et al. *PLoS One*. 2010; 5:e15004. [PubMed: 21165148]
12. Gupta S, et al. *J Biol Chem*. 2014; 289:8706–8719. [PubMed: 24415766]
13. Keefe AD, Cload ST. *Curr Opin Chem Biol*. 2008; 12:448–456. [PubMed: 18644461]
14. Lapa SA, Chudinov AV, Timofeev EN. *Mol Biotechnol*. 2016; 58:79–92. [PubMed: 26607475]
15. Ohsawa K, Kasamatsu T, Nagashima J, Hanawa K, Kuwahara M, Ozaki H, Sawai H. *Anal Sci*. 2008; 24:167–172. [PubMed: 18187867]
16. Shoji A, Kuwahara M, Ozaki H, Sawai H. *J Am Chem Soc*. 2007; 129:1456–1464. [PubMed: 17263432]
17. Wang RE, Wu H, Niu Y, Cai J. *Curr Med Chem*. 2011; 18:4126–4138. [PubMed: 21838692]
18. Kuwahara M, Ohsawa K, Kasamatsu T, Shoji A, Sawai H, Ozaki H. *Nucleic Acids Symp Ser (Oxf)*. 2005; :81–82.doi: 10.1093/nass/49.1.81
19. Kuwahara M, Sugimoto N. *Molecules*. 2010; 15:5423–5444. [PubMed: 20714306]
20. Rohloff JC, Gelinas AD, Jarvis TC, Ochsner UA, Schneider DJ, Gold L, Janjic N. *Mol Ther Nucleic Acids*. 2014; 3:e201. [PubMed: 25291143]
21. Kuwahara M, et al. *Nucleic Acids Res*. 2006; 34:5383–5394. [PubMed: 17012278]
22. Kuwahara M, et al. *Molecules*. 2010; 15:8229. [PubMed: 21076389]
23. MacPherson IS, Temme JS, Habeshian S, Felczak K, Pankiewicz K, Hedstrom L, Krauss IJ. *Angew Chem Int Ed Engl*. 2011; 50:11238–11242. [PubMed: 22191092]
24. Temme JS, Drzyzga MG, MacPherson IS, Krauss IJ. *Chemistry*. 2013; 19:17291–17295. [PubMed: 24227340]
25. Temme JS, Krauss IJ. *Curr Protoc Chem Biol*. 2015; 7:73–92. [PubMed: 26344234]
26. Temme JS, MacPherson IS, DeCoursey JF, Krauss IJ. *J Am Chem Soc*. 2014; 136:1726–1729. [PubMed: 24446826]
27. Ichida JK, Zou K, Horhota A, Yu B, McLaughlin LW, Szostak JW. *J Am Chem Soc*. 2005; 127:2802–2803. [PubMed: 15740086]
28. Yu H, Zhang S, Chaput JC. *Nat Chem*. 2012; 4:183–187. [PubMed: 22354431]
29. Renders M, Miller E, Hollenstein M, Perrin D. *Chemical Communications*. 2015; 51:1360–1362. [PubMed: 25487111]
30. Tanasova M, Goeldi S, Meyer F, Hanawalt PC, Spivak G, Sturla SJ. *ChemBioChem*. 2015; 16:1212–1218. [PubMed: 25881991]
31. Zuker M. *Nucleic Acids Res*. 2003; 31:3406–3415. [PubMed: 12824337]
32. Martin FH, Tinoco I Jr. *Nucleic Acids Res*. 1980; 8:2295–2299. [PubMed: 6159577]
33. Bock LC, Griffin LC, Latham JA, Vermaas EH, Toole JJ. *Nature*. 1992; 355:564–566. [PubMed: 1741036]
34. Long SB, Long MB, White RR, Sullenger BA. *RNA*. 2008; 14:2504–2512. [PubMed: 18971322]
35. White R, Rusconi C, Scardino E, Wolberg A, Lawson J, Hoffman M, Sullenger B. *Mol Ther*. 2001; 4:567–573. [PubMed: 11735341]



**Figure 1.**

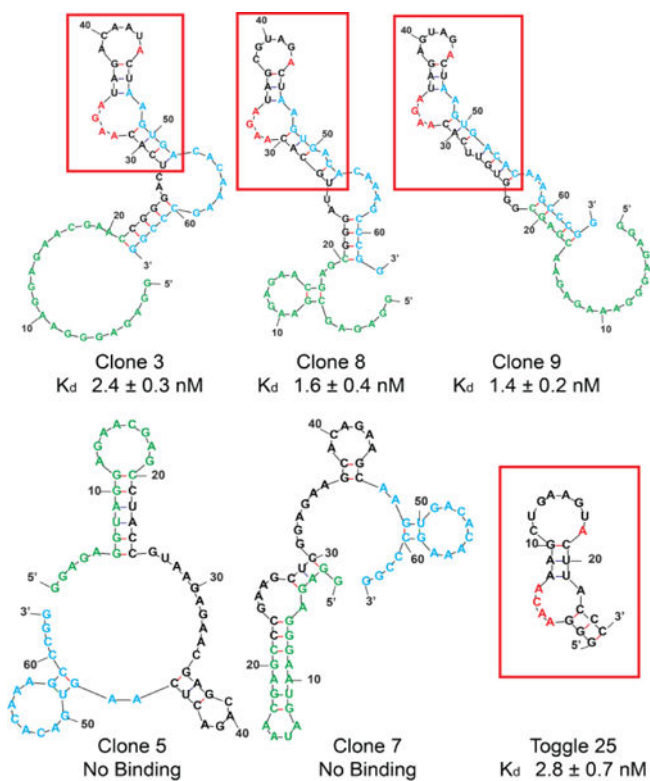
Generation and selection of a DNA-displayed RNA library. Black solid lines = DNA; dotted colored lines = RNA; grey = hexaethyleneglycol spacer; blue line = 5' terminal capture sequence complementary to S1; Pm = T7 promoter sequence; S1' = capture strand sequence; S2' = reverse primer; N<sub>25</sub> = random region; star = isodeoxycytidine; oval = T7 polymerase.



**Figure 2.**

Non-denaturing (A) and denaturing (B) gels of library generation. In both gels, lanes 2–4 are controls that show migration of the individual purchased oligos; lanes 5–8 show the migration of species produced when these components are combined and used to generate the library; lane 9 is a control showing that RNA transcription of an ordinary DNA duplex produces several additional bands observed in lane 8. Lane contents: 1) LMW ladder (NEB); 2) non-coding library strand (10 pmol); 3) capture arm (10 pmol); 4) rigidifier (10 pmol); 5) library annealed to capture arm (0.25 pmol); 6) extension product, 0.25 pmol; 7) extension product annealed to rigidifier (0.25 pmol); 8) transcribed library (0.25 pmol) 9) transcribed library without capture arm. The box indicates the final RNA-displayed library, which was excised from the gel, eluted from the gel slice and rerun on a denaturing gel (lane 10). Gel conditions: A) 10% acrylamide, 29:1 acrylamide:bis-acrylamide, stained with ethidium bromide. B) 8 M urea, 12.5% acrylamide, 29:1 acrylamide:bis-acrylamide, stained with Sybr Gold; lanes 1–9 contain the same samples in both gels.





**Figure 3.** mFold predictions of most stable secondary structures for selected clones. The red boxes and letters highlight the stem loop structure and sequence common to most binding clones, and to the “Toggle 25” aptamer previously reported by Sullenger and Long (Refs. 30, 31). Green = capture strand complement. Blue = complement of reverse primer.

**Table 1**

Sequences and thrombin binding of library clones after round 10. Dissociation constants are derived from triplicate measurement in nitrocellulose filter binding assays, and the error reported is the standard error of the curve fit. Each oligo also contains 5' and 3' sequences complementary to the capture strand and reverse primer, respectively, which are presented together with binding curves for each clone in the Supplementary Information.

Clone ID	Sequence	$K_d$ (nM)	$Fb_{max}$
1	UGUUACUCAC A Aa AUAGCGAAG ACU	$44.7 \pm 4.6$	$95.6 \pm 4.4$
2	CCGGCGUCAC A AG AUAGACAAA ACU	$3.5 \pm 0.3$	$80.6 \pm 1.6$
3	CGGGACUCAC A AG AUAGACAAU ACU	$2.4 \pm 0.3$	$65.3 \pm 1.9$
4	UGC GGUAUAC A AG AUAGCAAAG ACU	$10.0 \pm 0.9$	$84.5 \pm 2.3$
5	CUACCGU A AGAGAACGAGC AGACUC	NB	ND
6	CGGGACACU A AGGAACAUAAA AGUU	$13.8 \pm 1.9$	$74.5 \pm 3.2$
7	CCGAAGCUCGGAGAAGCACAGAAGC	NB	ND
8	GGGAUUGCAC A AG AUAGCGUAG ACU	$1.6 \pm 0.4$	$63.5 \pm 3.3$
9	GGGUGUUCAC A AG AUAGAGUAG ACU	$1.4 \pm 0.2$	$92.3 \pm 3.1$
10	GCUUUGACAC A AG AUACAAU AU AGU	$36.4 \pm 2.8$	$52.9 \pm 1.7$