

METHODS MANUSCRIPT

Single-round isolation of diverse RNA aptamers from a random sequence pool

Masahiko Imashimizu,^{1,3,*} Masaki Takahashi,¹ Ryo Amano¹ and Yoshikazu Nakamura^{1,2}

¹Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo, 108-8639, Japan and ²RIBOMIC Inc., Minato-ku, Tokyo, 108-0071, Japan

³Present Address: Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology, Tokyo, 135-0064, Japan
*Correspondence address. Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology, Tokyo, 135-0064, Japan; Tel: +81-3-3599-8232; Fax: +81-3-3599-8099; E-mail: m.imashimizu@aist.go.jp

Abstract

Aptamers are oligonucleotide ligands with specific binding affinity to target molecules. Generally, RNA aptamers are selected from an RNA pool with random sequences, using the technique termed SELEX, in which the target-binding RNA molecules are repeatedly isolated and exponentially amplified. Despite several advantages, SELEX often produces uncertain results during the iterative amplifications of the rare target-binding RNA molecules. Here, we develop a non-repeated, primer-less and target immobilization-free isolation method for generating RNA aptamers, which is robust to experimental noise. Uniquely, this method focuses on finding and removal of non-aptamer sequences from the RNA pool by RNase digestion leaving target-bound aptamer molecules, and thus is independent of aptamer types. The undigested RNA sequences remaining are so few in number that they must be mixed with a large excess of a known sequence for further manipulations and this sequence is then removed by restriction digestion followed by high-throughput sequencing analysis to identify aptamers. Using this method, we generated multiple RNA aptamers targeting α -thrombin and TGF β 1 proteins, independently. This method potentially generates thousands of sequences as aptamer candidates, which may enable us to predict a common average sequence or structural property of these aptamers that is different from input RNA.

Introduction

Aptamers are single-stranded oligonucleotides that have a broad spectrum of binding functions with specificity to various targets such as small chemical compounds, purified proteins and cells [1–3]. Target-binding aptamers are generally selected using a method termed SELEX (Systematic Evolution of Ligands by EXponential Enrichment) [4, 5].

Generation of RNA aptamers by SELEX starts from preparation of an initial RNA library bearing sequence diversity of about 10^{13} to 10^{16} . Each RNA sequence typically consists of a random

sequence of about 40 bases and PCR primers flanking the random sequence. The initial RNA library is then co-fractionated with single target molecules of interest. Such a fractionation is generally promoted by a specific technique, for instance, by immobilizing the target molecules to carriers and wash out of RNA species not bound to the target molecules on the carriers. The obtained RNA species are converted into DNA by reverse transcription (RT), and are duplexed, and amplified exponentially by PCR. Using the PCR products as a template, an RNA pool is synthesized by transcription. The second round of selec-

Received: 7 March 2018; Revised: 2 April 2018; Editorial decision: 4 April 2018; Accepted: 10 April 2018

© The Author(s) 2018. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

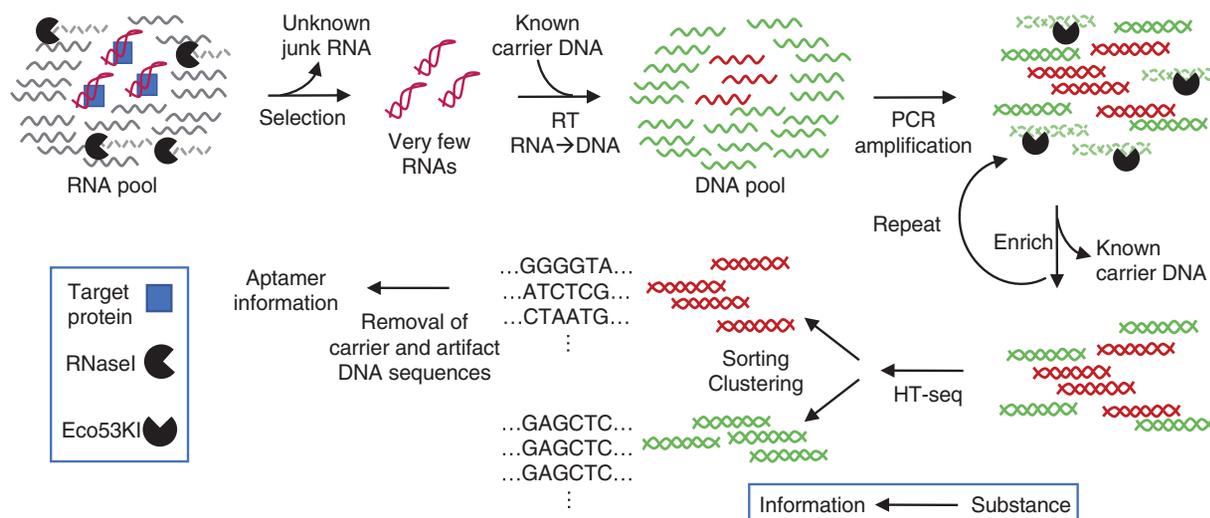


Figure 1. Scheme of the single-round RNA-aptamer isolation method. Aptamers with target-binding structures (in RNA form) are schematically represented as red lines in the RNA pool and the DNA pool, respectively. In this method, junk RNA (gray lines) represents non-aptamer RNA molecules (not binding to target proteins and degraded by RNase I). Carrier DNA (light green lines) represents added DNA molecules that we recognize by restriction sites and remove. The sequence information of the junk RNA removed by RNase I is unknown because it is designed to be random and thus consists of a massive number of sequence combinations. The sequence information of the added carrier DNA is known because it is designed to have $3 \times$ Eco53KI restriction sites. A large excess of carrier nucleic acids over aptamers allows experimental handling of aptamers as nucleic acids. Replacing the unknown junk sequence information by the known carrier sequence information is the core concept of this method. Aptamer-sequence information is obtained by subtracting information about carrier DNA and any artifact DNA (e.g. adapter sequences) from the total sequence information obtained by HT-seq.

tion and amplification is carried out in the same manner using the RNA pool that is produced in the first round. This procedure is repeated about 10 times, and the obtained nucleic acid sequences (usually DNA) are detected by a sequencing technique [6] to identify candidate sequences for specific aptamers. Binding of the candidate aptamers to the target of interest is then measured by a widely accepted technique such as surface plasmon resonance (SPR).

The essence of SELEX is stepwise selection with exponential increases in the ratio of the extremely few aptamers to the total nucleic acid sequence pool. Hence, even if one can discover aptamers from the large excess of non-aptamer nucleic acids by single-round selection, one will be faced with a challenge of detecting the rare low level of aptamers. This is a fundamental problem in aptamer generation but has not previously been well recognized. Indeed, although several studies have obtained DNA aptamers by single-round selections [7–12], to the best of our knowledge, the problem regarding the small yield has not yet been analytically resolved. On the other hand, if one persists in the repeated selections of aptamers coupled with PCR amplifications, one will continue to be faced with uncertainties related to the exponential enrichment of inevitable artifacts. Causes of such artifacts include the residual non-aptamer nucleic acids due to insufficient fractionation, influence of primer sequences on the selection process, a target molecule of inactive form due to the immobilization to a carrier, and sequence preferences of polymerases that perform PCR amplification, RT, and transcription.

To resolve this dilemma, we developed a method to thoroughly fractionate aptamers by a single selection while maintaining working amounts of nucleic acids (Fig. 1). Herein, we define any RNA not coding for aptamers against a target molecule of interest as *unknown junk RNA* and any artificial, identifiable non-aptamer coding DNA as *known carrier DNA*. The key to the success of this method is removal of unknown junk RNA occupying the initial sequence pool and preventing detection of

aptamers, and complementary addition of known carrier DNA that can protect aptamer-coding DNA as a decoy until it can be detected. Combining this approach with high-throughput DNA sequencing (HT-seq) and information analysis, the known carrier DNA can be easily recognized by its sequence information and can be removed from the sequence pool, to enable aptamer identification. We also show that the aptamer-candidates found by this method have a common sequence or structural property, which is statistically different from that of the initial library.

Materials and methods

Overview of our method developed for production of RNA aptamers is shown in Fig. 1. This method is divided into two main processes: preparing a HT-seq library and the subsequent sequence analysis. More specifically, this method consists of the sections (A) to (I) (Fig. 2). Below, we present all steps consisting of the sections (A) to (I) that we used to obtain RNA aptamers targeting α -thrombin and transforming growth factor β 1 (TGF β 1).

(A) Synthesis of multi-copy RNA library by *in vitro* transcription. We used a DNA template lacking a complete DNA duplex for T7 RNA polymerase in order to minimize the process needed for generating an initial RNA library. The DNA template is single-stranded and only the promoter region is double-stranded (Table 1A). Such a DNA template is known to yield RNA transcripts as much as the double-stranded DNA template [13].

1. Partially double-stranded template DNA was prepared by mixing 10 μ l of 100 μ M random DNA (the starting diversity of the DNA pool is $\sim 6 \times 10^{14}$) and 10 μ l of 100 μ M T7 promoter DNA in the buffer containing 5 mM HEPES-NaOH (pH 8.0) and 2 mM MgCl₂. Note that the random DNA has the complementary T7 promoter sequence at its 3'

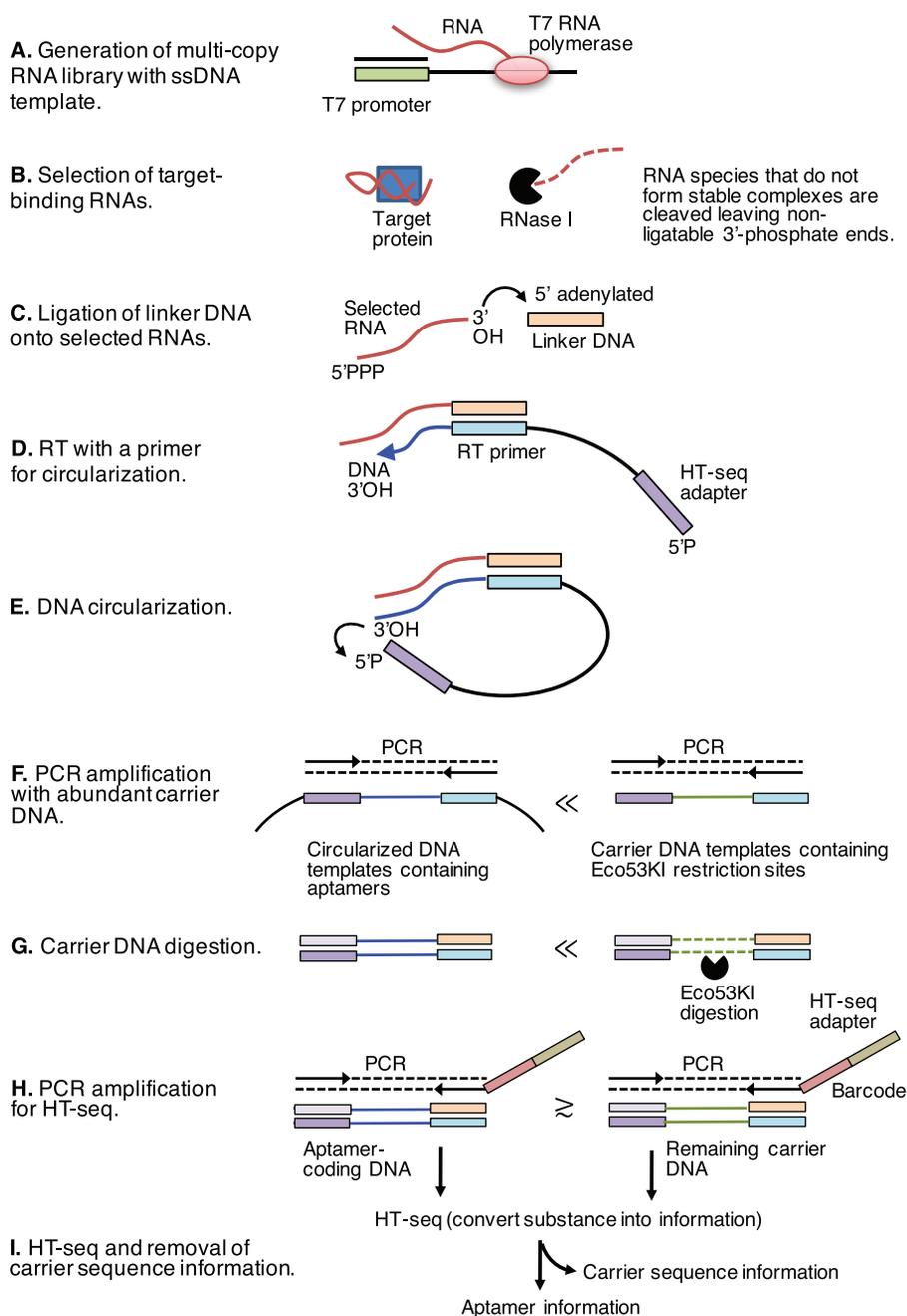


Figure 2. Processes for generating, fishing and identifying RNA aptamers.

terminus. The mixture was heated for 3 min at 95°C to denature and then the temperature was gradually decreased to 25°C.

- In vitro* transcription was performed by Y639F mutant of T7 RNA polymerase. We used 2' hydroxyl (OH) purine and 2' fluoro (F)-pyrimidine nucleotides as the substrate in order to obtain RNA aptamers that are stable to chemical hydrolysis. The mutant enzyme is known to readily incorporate the 2' F-nucleotides into the nascent RNA chain [14]. We purified the enzyme as described in Ref. [15]. Note that in this method, a combination of the wild-type T7 RNA polymerase and 2'-OH nucleotides is also available for producing an initial library. The reagents and template DNA were mixed as follows:

20 µl	Template DNA (GeneDesign, see Table 1A)
20 µl	10 × Transcription buffer [400 mM HEPES-NaOH (pH 8.0), 200 mM NaCl, 200 mM MgCl ₂ , and 30 mM spermidine]
20 µl	NTP mix [25 mM each of ATP(2' OH), GTP(2' OH), CTP(2' F) and UTP(2' F)]
20 µl	100 mM DTT
1 µL	100 mM MnCl ₂
1 µL	40 U/µl RNase Inhibitor (Takara Bio)
1 µL	100 U/ml Yeast Inorganic Pyrophosphatase (NEB)
10 µL	0.4 mg/ml T7 RNA polymerase, Y639F mutant
107 µL	Water
200 µl	Total

and SYBR gold staining of the gel as described in the step 8 of section A.

- We verified that the band corresponding to 40-nt RNA is invisible (i.e. is the 40-nt RNA amount below ~10 pg, a detection limit of oligo RNA stained by SYBR gold). As a control, we loaded the initial RNA library into the next lane of the sample.
- Fifty Microliters of 1 × SELEX buffer was added to the sample (~95 μl in total), and then pre-heated 100 μl PCI (70°C) was added to the sample, vortexed, and was centrifuged for 10 min at 20 000 × g at room temperature. The supernatant was collected for the next step.
- According to Ref. [16], the RNA sample was ethanol precipitated with a co-precipitating agent. The obtained RNA pellet was resuspended with 5.2 μl water.

(C) Ligation of linker DNA onto the selected RNA species. The linker DNA sequence is complementary to the RT primer, and has a 5' adenylation (promotes ligation) and a 3' amination (blocks self-ligation) (Table 1B).

- The RNA sample was heated for 2 min at 80°C to denature, and then was immediately placed on ice.
- Ligation reagents and the sample were mixed on ice as follows:

5.2 μl	Denatured RNA sample
2 μl	50% PEG8000
1 μl	10 × buffer
0.5 μl	0.5 μg/μl 5' adenylylated DNA linker (NEB, Table 1B)
0.3 μl	40 U/μl RNase Inhibitor (Takara Bio)
1 μl	T4 DNA ligase 2, truncated (NEB) or T4 RNA Ligase 2, Deletion Mutant (Epicentre)
10 μl	Total

- The temperature of the ligation mixture was immediately shifted to 37°C and incubated for 3 h.

(D) RT with a primer for circularization. In the 5'-end, the RT primer has the adapter sequence with phosphorylation that is needed for HT-seq by Ion PGM system (Thermo Fisher Scientific) (Table 1C). The adapter sequence can be changed depending on HT-seq platform used. By replacing the adapter sequence with a T7 promoter, repeated selections (i.e. SELEX version of this method) are also possible. RT reaction was carried out with PrimeScript RTase (Takara Bio). We have verified that the RT enzyme has high fidelity and is robust against RNA secondary structure [18]. We have obtained similar results using PrimeScript II RTase (Takara Bio).

- The ligation product (10 μl) was heated for 2 min at 80°C to denature, and then was immediately placed on ice.
- RT reagents and the ligation product were mixed on ice as follows:

10 μl	Denatured ligation product
4.4 μl	5 × PS (or PSII) RTase buffer
4.4 μl	2.5 mM dNTP mixture
1 μl	20 μM RT primer (Table 1C)
0.3 μl	40 U/μl RNase Inhibitor (Takara Bio)
1.9 μl	PrimeScript (or PrimeScript II) RTase (Takara Bio)
22 μl	Total

- The temperature of the RT sample was immediately shifted to 42°C and incubated for 45 min, and then for 15 min at 50°C.
- A total of 0.5 μl of 60 U/μl RNaseH (Takara Bio) was added to the RT sample, which was incubated for 30 min at 37°C.

(E) DNA circularization. In this reaction, we used an ATP-dependent ligase CircLigase ssDNA Ligase (Epicentre), which catalyzes the intramolecular ligation reaction (circularization) of a single-stranded DNA having a 5' phosphate group and a 3' hydroxyl group, respectively. We designed the sections C, D, and E, using the protocol of NET-seq library production developed by Churchman and Weissman [16] as a reference.

- Circularization reagent was added to 8.6 μl aliquot of the RT product as follows (the remaining 11.4 μl aliquot was stored at -80°C):

8.6 μl	RT product
1.2 μl	10× buffer
0.6 μl	1 mM ATP
0.6 μl	50 mM MnCl ₂
1 μl	CircLigase ssDNA Ligase (Epicentre)
12 μl	Total

- The sample was incubated for 60 min at 60°C, and then for 10 min at 80°C. The sample was stored at ≤ -20°C until proceeding with PCR amplification.

(F) PCR amplification with carrier DNA. PCR performed here must be done with high fidelity. We used a PCR enzyme PrimeSTAR Max DNA polymerase (Takara Bio), since we have verified the high-fidelity PCR by this enzyme (an average error rate is in the order of 10⁻⁵ per base or less) [18]. In this PCR, we added a large excess of single-stranded linear carrier DNA (10 or 100 pg) over the circularized template DNA that is derived from the RNA obtained in section B. The carrier DNA contains three Eco53KI restriction sites 5' GAGCTC 3' and has flanking sequences to which PCR primers bind (Table 1D). We designed the length of the PCR product of carrier DNA to be 80 bp (Table 1D), which is the same as that of the circularized template DNA.

- PCR reagents were added to a 6 μl aliquot of the circularized product as follows (the remaining 6 μl aliquot was stored at -80°C):

6 μl	Circularized product
1 μl	10 pg or 100 pg carrier DNA (Table 1D)
1.6 μl	10 μM 1 st PCR primers (Table 2)
11.4 μl	Water
20 μl	2 × PrimeSTAR Max Premix (Takara Bio)
40 μl	Total

- 1st PCR was performed as follows:

1 cycle	15 s	98°C
12 cycles (100 pg carrier DNA) or 18 cycles (10 pg carrier DNA)	10 s	98°C
	5 s	55°C
	5 s	72°C
1 cycle	Hold	4°C

Table 2. PCR primers used in this study

Purpose	Name	Sequence
1st PCR	Anti-linker F	5' -ATTGATGGTGCCTACAG-3'
	P1-full R	5' -CCTCTCTATGGGCAGTCGGTGAT-3'
2nd PCR	Anti-linker + adapter + barcode F	Adapter for Ion PGM HT-seq / Barcode / Anti-linker(variable) 5' -CCATCTCATCCCTGCGTGTCTCCGACTCAG/ CTAAGGTAAC/ATTGATGGTGCCTACAG-3'
	P1-full R	5' -CCTCTCTATGGGCAGTCGGTGAT-3'
	A30 F	5' -CCATCTCATCCCTGCGT-3'
3rd PCR	P1 R	5' -CCTCTCTATGGGCAGTC-3'

3. A 4 μ l of 10 \times DNA loading dye was added to each PCR sample. The sample was loaded into 8% TBE native polyacrylamide gel. PAGE and the following SYBR gold staining of the gel were performed as described in the step 8 of section A. The band corresponding to 80-bp PCR product was visualized by UV irradiation and was excised from the entire gel. The DNA sample was extracted from the gel piece, according to the protocol (16).

(G) Digestion of carrier DNA sequences followed by (H) PCR amplification for HT-seq. We digested the carrier DNA by Eco53KI in order to enrich aptamer sequences in the library. We chose Eco53KI because the enzyme works optimally in the low salt PCR buffer and produces blunt-ended DNA which causes less PCR artifacts compared to sticky-ended DNA. We repeated the PCR amplifications three times in total (including the 1st PCR performed in section F); alternatively, this iterative process can be removed and complemented by increasing the number of sequencing reads obtained by HT-seq (i.e. by expanding the search space in the sequence library of use). During the 2nd PCR, the adapter and barcode sequences necessary for HT-seq were attached into the library, so the length of the 2nd PCR product becomes 120 bp.

1. The ethanol precipitated 1st PCR product was resuspended with 8 μ l of water, to which 1 μ l of 10 \times cut smart buffer (NEB) and 1 μ l of 10 U/ μ l Eco53KI (NEB) were added in this order.
2. The sample was incubated for 60 min at 37°C.
3. Second PCR mixture was added to 5 μ l aliquot of the Eco53KI-digested product (the remaining 5 μ l aliquot was stored at $\leq -20^\circ\text{C}$, which can be used for optimizing the cycle number of the 2nd PCR if necessary).

5 μ l	Eco53KI digested product
0.8 μ l	10 μ M 2nd PCR primers (Table 2)
4.2 μ l	Water
10 μ l	2 \times PrimeSTAR Max Premix (Takara Bio)
20 μ l	Total

4. Second PCR was performed at 9 cycles (cycle program is the same as 1st PCR).
5. A 2 μ l of 10 \times DNA loading dye was added to each PCR sample. The sample was loaded into 8% TBE native polyacrylamide gel. PAGE and the following SYBR gold staining of the gel were performed as described in the step 8 of section A. The band corresponding to 120-bp PCR product was visualized by UV irradiation and was excised from the entire

gel. The DNA was extracted from the gel piece, according to the protocol [16].

6. Eco53KI digestion was performed as described in the step 1 of this section.
7. Third PCR mixture was added to 7.5 μ l aliquot of Eco53KI digested product (the remaining 2.5 μ l aliquot was stored at $\leq -20^\circ\text{C}$, which can be used for optimizing the cycle number of the third PCR if necessary):

7.5 μ l	2nd Eco53KI digested product
2.4 μ l	10 μ M 3rd PCR primers (Table 2)
20.1 μ l	Water
30 μ l	2 \times PrimeSTAR Max Premix (Takara Bio)
60 μ l	Total

8. Three PCR tubes were prepared and 20 μ l each of 3rd PCR product was added to the three PCR tubes. Third PCR was performed at 6 cycles for one tube, and at 9 cycles for the other two tubes (cycle program is the same as 1st PCR).
9. A 0.5 μ l of Eco53KI was added to one of the two tubes that have the 9-cycle PCR products. The sample was incubated for 30 min at 37°C.
10. PAGE and the following DNA extraction were performed as described in the step 4 of this section. One of the three 3rd PCR products with the highest homogeneity and with an optimal amount was chosen and used for HT-seq analysis. In this study, the 9-cycle PCR product with Eco53KI digestion (with the step 9) was chosen for the HT-seq analysis of thrombin aptamers, and the 9-cycle PCR product (without the step 9) was chosen for that of TGF β 1 aptamers.
11. DNA concentration was quantified using absorption. Typical concentration of the extracted DNA per sample in 200 μ l water was 1–10 nM. The sample was used for qPCR measurement and HT-seq, without concentration by Ethanol precipitation. The remaining sample was stored at -80°C until use.

(I) HT-seq analysis and removal of carrier sequence information. Sequencing reads were obtained by Ion PGM system using the Ion 318 chip (Thermo Fisher Scientific) and preprocessed using FASTX-Toolkit programs (http://hannonlab.cshl.edu/fastx_toolkit/). Next, the number of unique sequences in the total reads was counted, normalized, ranked, sorted and clustered into sequence families based on Levenshtein edit distance, using FASTAptamer programs [19]. Carrier DNA sequences were removed from the clustered sequences to reveal RNA aptamers. Not only Eco53KI-restriction sites, any artifacts that were generated during the library production can be removed in this step. Those artifact sequences were categorized into four groups with different origins: (i) Derivatives of the original Eco53KI-restriction sites GAGCTC, which were introduced during PCR (e.g. GAGTCT, GAGCCC and GAGCCT). (ii) Artifacts derived from *in vitro* transcription, such as highly homopolymeric A/T and T7 promoter sequence (Table 1A). (iii) Linker DNA sequence (Table 1B). (iv) Adapter sequences (Table 1C and D). Typically in our experiments, (iii) and (iv) were detected as the major artifact DNA. Since we digested the carrier DNA comprised of Eco53KI-restriction sites intensively, we detected more such additional artifact DNA than the carrier DNA in the sequencing reads.

Measurement of aptamer-target-protein interactions by SPR. Biacore 2000 (GE healthcare) was used as the SPR instrument. α -thrombin or TGF β 1 protein was immobilized on a CM5 sensor

chip by amine coupling, according to the manufacturer instructions. Briefly, CM5 dextran matrix was activated by 70 μ l of the crosslinking agent, using running buffer [10 mM Hepes-KOH (pH 7.5) and 100 mM KCl] at flow rate of 10 μ l/min. A 1.4 μ g each of α -thrombin or TGF β 1 protein in 70 μ l of 10 mM sodium acetate (pH 4.0 for α -thrombin and pH 5.0 for TGF β 1) was injected into a flow cell (Fc2) of CM5 sensor chip, and then was inactivated by 70 μ l of 50% ethanol ammine with the running buffer, and was further washed by 70 μ l of regeneration buffer [10 mM HEPES-KOH (pH 7.5), 1 M KCl, and 20 mM EDTA]. Similarly, 1 μ g BSA was immobilized on a reference flow cell (Fc1). RNA samples, synthesized by *in vitro* transcription, were injected into both Fc2 and Fc1. We used a reference mode (Fc2 – Fc1) for measuring the difference of SPR response. Kinetic analysis was performed with five RNA samples of different concentrations, the above-mentioned running/regeneration buffers, flow rate of 30 μ l/min, dissociation time of 2.5 min, and injection time of 2 min. In order to obtain kinetic parameters, data were globally fitted by curves according to a reaction model of 1: 1 binding, by using BIAevaluation software.

Results and discussions

How to pull aptamers out of junk sequence pool

We first prepared a random-sequence library comprised of multi-copy RNA species (estimated as \sim 10 copies or more) by *in vitro* transcription (Fig. 2A). Single-copy RNA species are unlikely to be detected even if successfully co-fractionated with target proteins. We introduced 2' F modification in any pyrimidine nucleotides of such RNA species in order to increase the chemical stability of aptamers obtained (see 'Materials and Methods' section).

Next, using this random-sequence library, RNA species that cannot form stable complexes with target proteins are eliminated by RNase I digestion coupled with ultrafiltration (Fig. 2B). RNase I cleaves any single-stranded regions of normal RNA and 2' F-modified RNA without base specificity, but does not cleave any double-stranded region [20]. Thus, any loop regions of RNA will also be susceptible to the cleavage, while structural studies reveal that loop regions play key roles for interacting with proteins [21–23]. Although the loss of true aptamers by the cleavage of the functional loops is inevitable in this method, this issue will not be problematic because of the multi-copy presence of RNA species and the fact that any RNA loops that are strongly bound to proteins can be subject to resistance to any RNase cleavage [17, 24, 25]. Furthermore, the concentration of target-bound RNA is considerably smaller, relative to the free RNA concentration, thus, even when cleavage occurs, the reaction for the target-bound RNA will be significantly delayed compared to that of free RNA, according to the definition of chemical kinetics (rate = rate constant \times concentration). Hence, in principle, by stopping the reaction after a certain incubation time, cleavage of the target-bound RNA can be negligible relative to the free RNA cleavage. Although we did not precisely specify the optimal incubation time in the step for isolating aptamers by the RNase I cleavage, it turned out that the time point used in this study (40 min) is practically feasible for finding multiple aptamers.

Both free RNA and partially-cleaved RNA were also substantially removed from the RNA pool by using a low-binding ultrafiltration spin column with a pore size retaining an RNA-protein complex but not the free RNA. Importantly, any remaining partially-cleaved RNA has a phosphate group at the 3'-end,

and is not ligated onto the linker DNA in the next step, thus can be removed from the subsequent steps for the library generation.

During the fractionation step, the target protein is fully intact and is not immobilized. In addition, RT and PCR primer sequences, usually embedded into the ends of RNA library in conventional SELEX methods, are omitted from our RNA library, reducing noise that interferes with the identification of true aptamers. Such primer sequences were added after the fractionation step, using efficient intermolecular and intramolecular ligation reactions (Fig. 2C–E) [16].

A successful fractionation of target-binding RNA from the initial RNA pool means a marked decrease in the amount of RNA to atto-gram orders or so [7]. Because the fractionated RNA pool would include RNA that poorly bound to the target protein as well as residual junk RNA, the actual amount of the true aptamers will be much smaller than the estimation, making it impossible to handle and detect with typical biochemical approaches like HT-seq. Hence, after converting the fractionated RNA into DNA by RT, we added carrier DNA (10 pg or 100 pg), consisting of triplicated Eco53KI restriction sites, to the aptamer-derived DNA (Fig. 2F). We then amplified not only the aptamer-derived DNA but also carrier DNA by PCR, and only the carrier DNA was then digestible by Eco53KI (Fig. 2G). The digested product was further amplified by PCR with primers containing HT-seq adapters, completing HT-seq library (Fig. 2H). Using HT-seq, we identified sequence information for aptamers, by eliminating the carrier DNA sequences that remained in the library (Fig. 2I). Note that instead of using carrier DNA, carrier 'RNA' having a genetically identical sequence to the carrier DNA is also possible to be used for the above-mentioned purpose, with caution in the reduction of ligation efficiency of the rare fractionated RNA. The carrier RNA can be subtracted after conversion to DNA by the same Eco53KI digestion.

This selection method based on RNase I protection can efficiently generate a quantitative difference between RNase-susceptible (i.e. non-aptamer) and RNase-insusceptible (i.e. target-bound aptamer) molecules, which is readily recognizable by HT-seq analysis, as described below. Indeed, we showed that the aptamers isolated by this method have much smaller rate constant for RNase I cleavage than a non-aptamer RNA and the initial RNA pool in the presence of their target proteins (Supplementary Fig. S1).

Generation of aptamers targeting thrombin

Based on the single-round RNA-aptamer isolation method, we screened RNA aptamers targeting human α -thrombin, a serine protease involved in blood coagulation. We chose this protein for the initial test of this method, since multiple 2'F-modified RNA and DNA aptamers for the protein have been already generated by SELEX [26–29]. In particular, we obtained \sim 58 thousand sequencing reads, from which we extracted unique sequences with the copy number information (Table 3). We carried out clustering analysis, removed known carrier sequences (see 'Introduction' section for the definition) as well as artifact sequences that were introduced during the HT-seq library generation [see the section (I) of 'Materials and Methods' section], and finally identified 54 sequences as the candidates for thrombin aptamers (Table 3). Note that the number of the analyzed sequencing reads here is quite small relative to typical HT-seq analysis dealing with more than one million reads. However, our method allowed identifying \sim 0.1% of the total reads as the aptamer candidates.

Table 3. Search for α -thrombin aptamers by HT-seq analysis

No. of sequences	%	Contents
57 878	100	Total sequencing reads (quality filter passed and barcode identified).
21 001	36	Unique sequences of the total sequencing reads
3001	5.2	Clustered sequences (edit distance ≤ 6 , copy number ≥ 2).
54	0.093	Aptamer candidates (clustered sequences \neq carrier/artifact sequences, copy number ≥ 3 , sequence lengths = 33–46 bases).

10 pg carrier DNA was used for the library generation

Among those 54 candidates, we chose 12 candidates with the highest frequencies of occurrence and with the length closest to 40 bases (Supplementary Table S1), in order to investigate their binding to α -thrombin by SPR measurements. The α -thrombin has two cationic patches (anion binding sites) termed exosites, to which heparin and nucleic acids nonspecifically bind [30]. DNA aptamer (G-quadruplex) [31] targeting exosite I, and RNA aptamer (stem-loop with an internal bulge) targeting exosite II have also been isolated [26]. Thus, we injected 1 $\mu\text{g}/\mu\text{l}$ (~ 50 – $100 \mu\text{M}$) heparin into the thrombin-immobilized SPR sensor chip prior to injecting each RNA aptamer candidate (Fig. 3A). In this experimental condition, binding of heparin to thrombin should be saturated, since an equilibrium dissociation constant (K_D) of heparin for thrombin binding has been reported as 6 μM [32]. This experimental setup enabled us to screen aptamers that (i) specifically bind to any thrombin sites with higher affinity than heparin, and/or (ii) fully novel aptamers targeting the thrombin sites other than the cationic patches.

We synthesized the 12 RNA species by *in vitro* transcription, which were partly purified by ultrafiltration. We injected each of the 12 RNA species into the thrombin-immobilized SPR sensor chip at a concentration of $\sim 100 \text{ nM}$. For monitoring the specific interaction of the RNA with thrombin, the SPR response resulting from the non-specific binding of RNA to BSA was subtracted from that of the RNA binding to thrombin. Although those RNA transcripts contained byproducts, we successfully identified the two RNA species as the aptamers with obvious capabilities for binding to thrombin in the presence of saturated heparin (Fig. 3A, Tbn#3 and Tbn#6). Next, we isolated homogeneous RNA transcripts for Tbn#3 and Tbn#6, respectively, by PAGE purification, and evaluated kinetics for the interactions with thrombin (Fig. 3B). Finally, we revealed that the two aptamers bind to thrombin with K_D of $\sim 1 \text{ nM}$ (Fig. 3B and C).

Despite having been assigned as candidates, the remaining 10 RNA species were not identified as aptamers by the SPR measurements. This false recognition may be due to (i) residual junk RNA exempt from RNase I cleavage, (ii) unknown artifacts introduced during HT-library production by this method, or (iii) thrombin-binding affinity of the RNA species comparable to that of heparin. Note that as represented by Tbn#8, the candidates for thrombin aptamer mostly had the SPR responses slightly higher than that of the initial RNA library (Fig. 3A), suggesting that the third case is the most likely cause. The initial RNA library that was tested for the SPR measurement was prepared as described in the section (A) of ‘Materials and Methods’ section. Sequences of the Tbn#3 and Tbn#6 RNA aptamers, and

those of all candidates obtained in this study are shown in Fig. 3D and in Supplementary Table S1, respectively.

Generation of aptamers targeting TGF β 1

We have also used the single-round RNA-aptamer isolation method to search for RNA aptamers targeting human TGF β 1. In this experiment, we employed two different amounts of carrier DNA (10 pg and 100 pg) for generating the HT-seq library, and compared those two results: We obtained ~ 6 – 7 thousand aptamer candidates (1.1%) out of ~ 500 – 600 thousand sequencing reads using either carrier DNA amount (Tables 4 and 5). We also showed that (i) carrier DNA amounts lower than ~ 10 pg did not sufficiently amplify the aptamer-derived DNA by PCR and (ii) carrier DNA amounts higher than ~ 100 pg increased the ratio of carrier DNA among the total sequencing reads so as to reduce the number of final aptamer candidates found. Therefore, we concluded that carrier DNA of 10–100 pg is optimal for obtaining the maximal number of aptamer candidates. These results also indicate that there is no linearity between the amount of carrier DNA and the final number of aptamer candidates.

In order to test TGF β 1-binding capability of the aptamer candidates obtained, we chose six with the highest frequencies of occurrence and with the length closest to 40 bases, from among each of two candidate libraries (Supplementary Tables S2 and S3). Similarly to the SPR measurement for thrombin aptamers described above, we used partly purified RNA transcripts for the initial test. TGF β 1 and BSA were immobilized on different flow cells, and upon injecting each RNA transcript, the difference of SPR response between TGF β 1 and BSA was monitored. We detected a larger response to TGF β 1 than BSA in all RNA transcripts analyzed, and the three with the highest affinity to TGF β 1 were arbitrarily assigned as aptamers (Fig. 3D). We purified the RNA transcripts for the three aptamers by PAGE, and used them for kinetic analysis. The result showed that all three aptamers have K_D of about 10 nM for TGF β 1 binding (Fig. 3C and Supplementary Fig. S2); however, interestingly the association rate (defined by k_{ON} value) and dissociation rate (defined by k_{OFF} value) were up to two orders of magnitude different one to the other (Fig. 3C, shown as $k_{\text{on}}-k_{\text{off}}-K_D$ map [33]). This difference may stem from different folding transitions of secondary structures for the aptamers (Fig. 3E) [34]. Importantly, three TGF β 1 aptamers with different kinetic properties were obtained, yet we tested only 12 candidates by SPR, suggesting that this method acquired aptamers with diverse affinity properties in a single selection.

Comparison of the products: SELEX and the single-round isolation method

SELEX experiments sometimes allow us to reproducibly find the aptamers with a conserved motif(s), as reported for SELEX using human immunoglobulin G as a target protein [22]. This fact prompted us to compare all the candidate sequences obtained in this study, with sequences of already-known RNA aptamers targeting thrombin [26, 27, 29] and TGF β 1 [35–37], respectively. In the candidates for thrombin aptamers, we found three different sequences that were 18-to-20-bases (55–61%) identical to the corresponding sequences of the already-known aptamers (Supplementary Fig. S3). These similar sequence pairs also appeared to be responsible for the similar configurations of the secondary structures in the pairs (Supplementary Fig. S3).

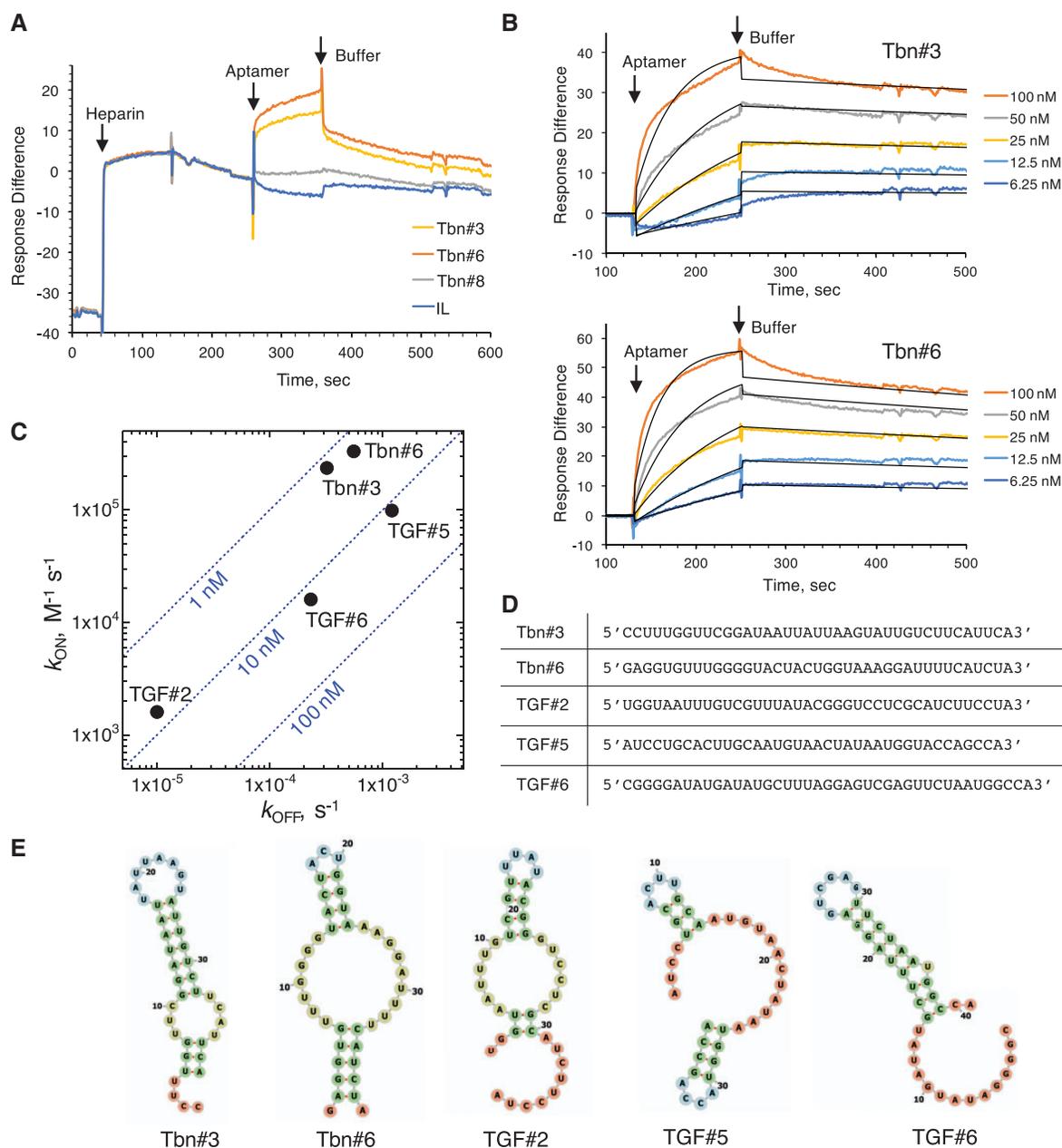


Figure 3. Characterizations of RNA aptamers targeting α -thrombin and TGF β 1, respectively. (A) Identification of aptamers targeting thrombin. SPR responses of RNA-thrombin binding by subtracting those of RNA-BSA binding are shown. Heparin, RNA aptamer candidates, and buffer were injected at the time points indicated by arrows, respectively. Tbn#3 and Tbn#6 are the aptamers identified. Tbn#8 represents an example of the candidates whose binding to thrombin was not detected, and IL represents an initial library used in this experiment. (B) Kinetic analysis of Tbn#3 and Tbn#6. The response differences between the thrombin and BSA upon aptamer injections at five different concentrations were globally fitted to obtain K_D values of 1.4 nM for Tbn#3 and 1.7 nM for Tbn#6, respectively (see details for main text). (C) A k_{on} - k_{off} - K_D map [33] depicting association (y axis, k_{on}) and dissociation (x axis, k_{off}) rate constants and affinities (dashed lines, K_D) of the α -thrombin (Tbn) and TGF β 1 (TGF) aptamers obtained in this study. (D) RNA sequences for α -thrombin and TGF β 1 aptamers, respectively. (E) RNA secondary structures predicted by using Vienna RNA Websuite [55], and visualized by Forna [56] with a color code: stems (Green), interior Loops (Yellow), hairpin loops (Blue), and 5' and 3' unpaired region (Orange).

We did not find any conserved motif within the candidate sequences for thrombin-aptamer obtained in this study.

In the candidates for TGF β 1 aptamers, we found two different conserved sequence motifs: one is a pyrimidine-rich sequence motif and the other is a G-rich sequence motif (Supplementary Fig. S4). The latter appeared to be too short to form a G-quadruplex. These two motifs existed in both candidate libraries that were made with 10 pg and 100 pg of carrier DNA. Interestingly, we also found another pyrimidine-rich

motif and a longer G-rich motif (the latter can form a G-quadruplex) in the sequences of already-known TGF β 1 aptamers (Supplementary Fig. S4). Since the detailed sequences and lengths in the respective motifs were quite different between our candidates and the already-known aptamers, it was uncertain whether the structural origins of the two motifs observed in our candidates and the already-known aptamers are related.

As far as assuming that the initial sequence pool is nearly random, the possibility of extracting similar aptamer

Table 4. Search for TGFβ1 aptamers by HT-seq analysis

No. of sequences	%	Contents
641 483	100	Total sequencing reads (quality filter passed and barcode identified).
521 411	81	Unique sequences of the total sequencing reads.
19 133	3.0	Clustered sequences (edit distance ≤ 6, copy number ≥ 2).
7215	1.1	Aptamer candidates (clustered sequences ≠ carrier/artifact sequences, copy number ≥ 3, sequence lengths = 33–45 bases).

10 pg carrier DNA was used for the library generation.

sequences out of different sequence pools is generally very low, even though the same target protein is used. In each experiment, we are limited to $\sim 1 \times 10^{14}$ unique sequences out of a possible $\sim 1 \times 10^{24}$ (i.e. the total number of the sequence combinations theoretically obtained from random 40-mer RNA). However, the result of the comparison between thrombin/TGFβ1 aptamers made by standard SELEX and those made by our method suggests that regardless of the method, finding aptamers with similar structural motifs or backbones out of different sequence pools might not be unlikely when the target protein is the same. This is because it is likely to be able to find ≤ 20 -bases identical sequence pairs among $\sim 1 \times 10^{14}$ unique sequences, it is also likely that high sequence similarity is not even necessary for constructing the same structural motif.

Compared to our single-selection method, SELEX can provide more chances for the structural optimization of aptamers for having higher target-binding affinity, since each of the repeated selections is coupled with an increased sequence diversity due to the accumulation of PCR errors. Hence, it is reasonable that the convergent point or advantage of SELEX products is different from that of our single-selection products. In terms of searching for the highest affinity aptamers, SELEX would be more beneficial, while our method would be more beneficial to find aptamers with diverse functions (see Fig. 3C as the example). Our method can also be used as a part of SELEX (see ‘Materials and Methods’ section).

Advantages of the single-round isolation method

The single-round isolation method (Figs 1 and 2) enabled us to generate multiple RNA aptamers targeting two different proteins, respectively (i.e. human α -thrombin and human TGFβ1), with different molecular sizes and physical properties (Fig. 3). The α -thrombin (~ 60 kDa) has two cationic patches, which can interact with nucleic acids nonspecifically. TGFβ1 homodimer (~ 25 kDa) does not have a nucleic-acid binding domain. In the former case, nonspecific RNA-protein binding may hinder efficient fractionation of aptamers. In the latter case, lack of a strong ionic interaction as well as small protein size may cause incomplete protection against RNase I digestion, via forming unstable complexes with the RNA. Nevertheless, the fact that we acquired aptamers with high and specific binding affinities to both proteins, respectively, implies the robustness of this method against physical chemical properties of target proteins. We envision future applications of this method for acquiring

Table 5. Search for TGFβ1 aptamers by HT-seq analysis

No. of sequences	%	Contents
543 803	100	Total sequencing reads (quality filter passed and barcode identified).
446 788	82	Unique sequences of the total sequencing reads.
15 287	2.8	Clustered sequences (edit distance ≤ 6, copy number ≥ 2).
6119	1.1	Aptamer candidates (clustered sequences ≠ carrier/artifact sequences, copy number ≥ 3, sequence lengths = 33–45 bases).

100 pg carrier DNA was used for the library generation.

aptamers targeting membrane proteins, which have been believed to be difficult due to the molecular heterogeneity of a membrane surface and the flexibility of the protein regions protruded from the membrane [38].

To date, several single-round selection methods for obtaining DNA aptamers have been reported, which utilize the techniques/instruments of nonequilibrium capillary electrophoresis [8], microfluidic selection [9], affinity chromatography [10], AFM/fluorescence system [11], amylose resin-mediated selection coupled with HT-seq detection [12] and DNase I digestion [7, 38, 39]. Notably, our method using known DNA carrier enabled overcoming the problem of isolating very small number of molecules, which has continued to be a problem for those pre-existing methods. For example, using the DNase I-mediated single-round selection, Liu et al. [(7)] has selected ~ 500 DNA species (~ 20 atto grams) that bind to Hepatitis B virus core protein out of 10^{16} DNA species that initially exist in a library used. Although they succeeded in obtaining multiple aptamers with high and specific affinity to the protein, their detection of aptamers must have been limited by the quality of PCR amplification of the selected library.

Our method has several advantages: (i) one can generate both DNA and RNA aptamers based on this single method. To the best of our knowledge, single-round selection has not been achieved for RNA aptamers [38] except a HT-seq study using human genomic RNA pool as the initial library [40], because RNA is not the direct template for PCR and thus is not effectively amplified. (ii) One can increase the sensitivity to detect aptamers by simply increasing the number of sequencing reads, i.e. by enlarging the searching space in the selected library. (iii) One does not need specific instruments enhancing fractionations, specific assumptions and specific knowledge of bioinformatics.

Statistical evaluation of the structural properties of aptamers

Using the sequence data obtained in this study, we investigated a statistical difference between the sequences in the unselected and selected pools. In particular, we attempted to find the sequence–structure relationships specifically existing in the RNA pool for the aptamer candidates. We sampled all 54 sequences from the thrombin aptamer pool (group 1), and 108 sequences from the two kinds of TGFβ1 aptamer pools whose libraries were made with 10 pg and 100 pg of carrier DNA (group 2 and 3), respectively. These sequences sampled have the highest frequencies of occurrence. We also randomly sampled 324

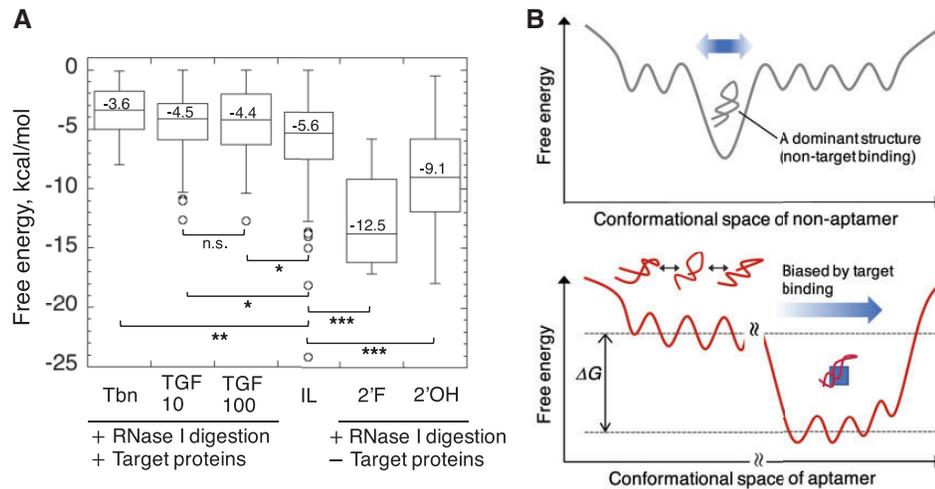


Figure 4. Average structural property of the RNA pool selected for the target binding. (A) Aptamers have higher calculated MFEs of the secondary structures compared to non-aptamers. Box plots of MFEs are shown with the mean values. Data for aptamers are obtained from selected sequence pools that were digested by RNase I in the presence of thrombin (Tbn, $n = 54$) or TGF β 1 [two different libraries using 10 pg (TGF10, $n = 108$) and 100 pg carrier DNA (TGF100, $n = 108$) are shown]. Data for control groups are obtained from unselected sequence pools that were digested by RNase I in the absence of any proteins (2' F-modified RNA, $n = 9$; 2' OH-unmodified RNA, $n = 143$) and the initial library (IL, $n = 324$). The sequences of initial library are randomly chosen from the three non-selected libraries ($n = 108$ for each) that are independently sequenced as described in 'Materials and Methods' section. Outliers are also plotted as individual points. P-values ($P < 1 \times 10^{-3}$, $**P < 1 \times 10^{-5}$, $***P < 1 \times 10^{-9}$) of two-tailed t-test are shown for pairs with statistically significant differences and for pairs with non-significant (n.s.) difference. (B) Model illustrating the structural difference between an average non-aptamer (top) and an average aptamer (bottom). In an average non-aptamer RNA (gray line), the conformation ensemble is dominated by a single stable structure that has no capability for the target binding, therefore, the ensemble is not affected by the target protein. In an average aptamer (red line), the conformation ensemble is not dominated by a single stable structure but can be biased toward a stable mode by binding to the target protein (blue rectangle). Without the target protein, the shift from the unstable to stable mode of an aptamer is likely hindered by a high activation energy (data not shown). The free energy difference (ΔG) in the presence or absence of the target protein is responsible for the stability of the complex.

sequences from the initial library as a control (group 4). Each of these four groups is comprised of highly diverse sequences (Supplementary Tables S4–S7). Using the RNA sequences in all groups, we computed the minimum free energies (MFEs) of the secondary structures and the related structural factors by the RNAfold program (Supplementary Tables S4–S7) [41]. Interestingly, we observed that the average MFEs for the aptamer groups are significantly higher than that for the initial library (Fig. 4A, $P < 0.001$).

Similar to the groups 1, 2, and 3, we also sampled 74 sequences from the pool of thrombin aptamer candidates and 125 sequences from the pools of two kinds of TGF β 1 aptamer candidates, all of which were selected by the RNase I protection but were not assigned as aptamers due to the lower frequencies of occurrence of sequencing reads (Supplementary Tables S8–S10, all these sequences have only one copy of sequencing read). Similarly, we observed significantly higher average MFEs for the selected groups than for the unselected initial library (Supplementary Fig. S5, $P < 0.01$). Therefore, the higher average MFE of the secondary structure is a common feature of the RNA sequence pools selected by this method.

Note that there is no significant difference in the average MFEs between the two TGF β 1 aptamer groups 2 and 3 (Fig. 4A, $P = 0.7$), even though no two share the same sequence. This implies that the average MFEs will turn out to be the statistically indistinguishable when the sequence (or structural) properties of the two groups are the same.

These results suggest that the higher average MFE represents the primary structural signature of the averaged aptamer candidate or otherwise it represents the presence of RNase I digestion-resistant sequences. To distinguish between these two possibilities, we digested random RNA sequence pools (with or without 2' F modification) using RNase I in the absence

of any proteins as we did for selecting target-binding RNA sequences. This experiment should provide sequence pools that are simply resistant to RNase I digestion. Interestingly, the sequence pools obtained showed significantly lower average MFEs compared to those of the initial library and any sequence pools selected in the presence of target proteins (Fig. 4A, $P < 1 \times 10^{-9}$, Supplementary Tables S11–S14). Since a decrease in MFE of the secondary structure can often be achieved by forming a long stable stem-loop, this result is also consistent with the fact that any RNA molecules with stably base-paired structures are essentially insensitive to RNase I digestion [20]. Therefore, we support the former possibility and presume that aptamers selected for binding to target proteins tend to have the increased diversity of a conformational ensemble in thermodynamic equilibrium due to the higher MFE (i.e. the lack of the single thermodynamically stable state).

Although the above-mentioned hypothesis was made based on a limited number of samples $< 10^3$, a strikingly similar result has been reported by NMR measurements that used much larger sequence pools of the protein-targeting aptamer candidates [42]. In particular, NMR imino proton signals of guanosine and uridine residues contain information about base-pairing in RNA molecules [43]. Hence, the imino proton signals are proportional to the number of stably base-paired RNA molecules in an RNA pool tested. Amano et al. [42] measured one-dimensional imino proton NMR spectra of RNA pools during SELEX. They compared imino proton signals of the RNA pools selected by 2–4 SELEX rounds with that of the initial RNA pool and found significant reductions of the signals (at the chemical shift of 10–12 ppm) in the selected pools. In such selected pools, overall target-binding affinity was increased but the high sequence diversity remained. Therefore, the result indicates that these stably base-paired structures that dominated the initial pool were

significantly reduced once selected for target-binding. The NMR study also showed that the imino proton signal was increased by the presence of the target protein in the selected RNA pool [42], indicating the overall RNA structures in the pool were biased and stabilized by binding to the target protein.

Taking this information into consideration, we speculate that the dynamic conformational equilibrium itself as well as the bias of the conformational equilibrium toward stable states upon binding to the target protein favors the RNA–aptamer–protein complex, thereby increasing the binding affinity of the aptamer (Fig. 4B): in a typical RNA molecule, the conformational ensemble appears to be dominated by a single low free energy state, depriving it of the chance to be shifted to the target-binding states by thermal fluctuations. In the typical selected aptamer, the conformational ensemble appears not to be significantly constrained by such a thermodynamically stable state, and thus has more chances to be shifted to a state(s) to bind the target. Upon binding to the target protein, the aptamer conformations can be biased to be lower overall free energy, stabilizing the complex.

This statistical view of aptamer–target interactions is not assigned to either induced fit or conformational selection model that has been traditionally assumed, but may be described by the hybrid of the two paradigmatic models as has been argued before by McCluskey and Penedo [44]. This notion is also consistent with an entropy dominated mechanism regulating DNA–protein interactions, where the stability of DNA–protein complex is determined by the overall free energy of the complex rather than by the binding energy of a unique static state [45]. More specifically, the conformational (thermal) fluctuations of DNA–protein complexes biased by DNA sequence repetitive elements [46] or by small molecule compounds [47] can significantly affect the binding affinity, underlying the regulatory mechanisms of transcription regulation [45, 48]. The effect of the conformational fluctuations on the affinity has also been discussed in RNA–protein interactions [49, 50].

Insight into binding affinity of aptamers

In general, aptamers can be created based on the concept of ‘binding affinity’ of nucleic acids for target molecules. While recent progress of structural studies have revealed individual pictures of aptamer–protein interactions [1, 23, 51], the molecular fundamentals necessary for designing protein-targeting aptamers remain poorly understood due to the presence of uncertainties. Such obligatory uncertainties stem from the facts that (i) unknown technical artifacts remained in these methods used to select and identify aptamers, (ii) binding of aptamers to the target molecules can in principle be stochastic rather than deterministic, when the molecular copy number of each aptamer is very small [52], (iii) a secondary structure of an aptamer has a probability distribution [53], (iv) the sizes of aptamers and the complexes with the target molecules ranges approximately from 1 to 10 nm, so fluctuations of the structures (i.e. conformational heterogeneity) are not negligible but are essential for interactive functions [45]. (iii) and (iv) are related or essentially the same. The role of conformational heterogeneity on interaction of an RNA aptamer with Interleukin-17 heterodimer has been suggested [54].

In order to design aptamers, the statistical evaluation of a large number of aptamers would be a possible approach to address such uncertainties. Although more detailed investigation of effects of the RNase I digestion on the selected sequence pool would be needed, we believe that this method has a potential to

enable such statistical evaluation. Therefore, our method would deepen the understanding of nucleic-acid–protein interactions, opening a window for an analytical design of protein-targeting aptamers with desirable functions. Such an analytical approach would also enhance the academic and industrial potentials inherent in single-stranded nucleic-acids.

Supplementary data

Supplementary data is available at BIOMAP online.

Conflict of interest statement. Y.N. is an employee of Ribomic Inc., and holds equity of Ribomic Inc.

Acknowledgements

We thank Donald L. Court for critical reading of the article. We are also grateful to Yoshifumi Hashimoto and Kazuteru Aoki for helpful discussions and for sharing their SELEX data.

Funding

This work was supported by JSPS KAKENHI (number 16H06692) to M.I. and by AMED (number 17ak0101054h0002) to M.T.

Authors' contributions

M.I., M.T., and Y.N. designed research; M.I. and R.A. performed research; M.I. analyzed data; M.I., wrote the article.

References

1. Nakamura Y, Ishiguro A, Miyakawa S. RNA plasticity and selectivity applicable to therapeutics and novel biosensor development. *Genes Cells* 2012;17:344–64.
2. Groher F, Suess B. In vitro selection of antibiotic-binding aptamers. *Methods* 2016;106:42–50.
3. Vinkenborg JL, Karnowski N, Famulok M. Aptamers for allosteric regulation. *Nat Chem Biol* 2011;7:519–27.
4. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: rRNA ligands to bacteriophage T4 DNA polymerase. *Science* 1990;249:505–10.
5. Ellington AD, Szostak JW. In vitro selection of RNA molecules that bind specific ligands. *Nature* 1990;346:818–22.
6. Nguyen Quang N, Perret G, Duconge F. Applications of high-throughput sequencing for in vitro selection and characterization of aptamers. *Pharmaceuticals* 2016;9:76.
7. Liu Y, Wang C, Li F et al. DNase-mediated single-cycle selection of aptamers for proteins blotted on a membrane. *Anal Chem* 2012;84:7603–6.
8. Berezovski M, Drabovich A, Krylova SM et al. Nonequilibrium capillary electrophoresis of equilibrium mixtures: a universal tool for development of aptamers. *J Am Chem Soc* 2005;127:3165–71.
9. Lou X, Qian J, Xiao Y et al. Micromagnetic selection of aptamers in microfluidic channels. *Proc Natl Acad Sci USA* 2009;106:2989–94.
10. Nitsche A, Kurth A, Dunkhorst A et al. One-step selection of vaccinia virus-binding DNA aptamers by MonoLEX. *BMC Biotechnol* 2007;7:48.
11. Peng L, Stephens BJ, Bonin K et al. A combined atomic force/fluorescence microscopy technique to select aptamers in a

- single cycle from a small pool of random oligonucleotides. *Microsc Res Tech* 2007;**70**:372–81.
12. Zykovich A, Korf I, Segal DJ. Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. *Nucleic Acids Res* 2009;**37**:e151.
 13. Arnaud-Barbe N, Cheynet-Sauvion V, Oriol G et al. Transcription of RNA templates by T7 RNA polymerase. *Nucleic Acids Res* 1998;**26**:3550–4.
 14. Sousa R, Padilla R. A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J* 1995;**14**:4609–21.
 15. Bonner G, Patra D, Lafer EM, Sousa R. Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. *EMBO J* 1993;**11**:3767–75.
 16. Churchman LS, Weissman JS. Native elongating transcript sequencing (NET-seq). *Curr Protoc Mol Biol* 2012; **4**, Unit 4.14.1–17.
 17. Imashimizu M, Takahashi H, Oshima T et al. Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol* 2015;**16**:98.
 18. Imashimizu M, Oshima T, Lubkowska L, Kashlev M. Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res* 2013;**41**:9090–104.
 19. 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.
 20. Shen V, Schlessinger D. (1982) *RNase I of Escherichia coli*. Academic Press, New York.
 21. Nomura Y, Sugiyama S, Sakamoto T et al. Conformational plasticity of RNA for target recognition as revealed by the 2.15 Å crystal structure of a human IgG-aptamer complex. *Nucleic Acids Res* 2010;**38**:7822–9.
 22. Miyakawa S, Nomura Y, Sakamoto T et al. Structural and molecular basis for hyperspecificity of RNA aptamer to human immunoglobulin G. *RNA* 2008;**14**:1154–63.
 23. Bjerregaard N, Andreasen PA, Dupont DM. Expected and unexpected features of protein-binding RNA aptamers. *Wires Rna* 2016;**7**:744–57.
 24. Glatt S, Zabel R, Kolaj-Robin O et al. Structural basis for tRNA modification by Elp3 from *Dehalococcoides mccartyi*. *Nat Struct Mol Biol* 2016;**23**:794–802.
 25. Komissarova N, Kashlev M. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc Natl Acad Sci USA* 1998;**95**:14699–704.
 26. Long SB, Long MB, White RR, Sullenger BA. Crystal structure of an RNA aptamer bound to thrombin. *RNA* 2008;**14**:2504–12.
 27. Bompiani KM, Monroe DM, Church FC, Sullenger BA. A high affinity, antidote-controllable prothrombin and thrombin-binding RNA aptamer inhibits thrombin generation and thrombin activity. *J Thromb Haemost* 2012;**10**:870–80.
 28. Jeter ML, Ly LV, Fortenberry YM et al. RNA aptamer to thrombin binds anion-binding exosite-2 and alters protease inhibition by heparin-binding serpins. *FEBS Lett* 2004; **568**:10–4.
 29. Liu X, Cao G, Ding H et al. Screening of functional antidotes of RNA aptamers against bovine thrombin. *FEBS Lett* 2004;**562**:125–8.
 30. Carter WJ, Cama E, Huntington JA. Crystal structure of thrombin bound to heparin. *J Biol Chem* 2005;**280**:2745–9.
 31. Macaya RF, Schultze P, Smith FW et al. Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc Natl Acad Sci USA* 1993;**90**:3745–9.
 32. Olson ST, Halvorson HR, Bjork I. Quantitative characterization of the thrombin-heparin interaction. Discrimination between specific and nonspecific binding models. *J Biol Chem* 1991;**266**:6342–52.
 33. Markgren PO, Schaal W, Hamalainen M et al. Relationships between structure and interaction kinetics for HIV-1 protease inhibitors. *J Med Chem* 2002;**45**:5430–9.
 34. Treiber DK, Williamson JR. Exposing the kinetic traps in RNA folding. *Curr Opin Struct Biol* 1999;**9**:339–45.
 35. Gold L, Pagratis N. (2000) Patent US006124449A.
 36. Gold L, Janjic N, Pagratis N. (2001) US6207816B1; PCT/US96/08014.
 37. Toothman PJ, Ringquist S, Gold L. (1998) US5731424A.
 38. Takahashi M. (2017) Aptamers targeting cell surface proteins. *Biochimie* 2018;**145**:63–72.
 39. Wu YX, Kwon YJ. Aptamers: the 'evolution' of SELEX. *Methods* 2016;**106**:21–8.
 40. 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.
 41. Lorenz R, Bernhart SH, Honer Zu Siederdisen C et al. ViennaRNA Package 2.0. *Algorithms Mol Biol* 2011;**6**:26.
 42. Amano R, Aoki K, Miyakawa S et al. NMR monitoring of the SELEX process to confirm enrichment of structured RNA. *Sci Rep* 2017;**7**:283.
 43. Reid BR. NMR studies on RNA structure and dynamics. *Annu Rev Biochem* 1981;**50**:969–96.
 44. McCluskey K, Carlos Penedo J. An integrated perspective on RNA aptamer ligand-recognition models: clearing muddy waters. *Phys Chem Chem Phys* 2017;**19**:6921–32.
 45. Imashimizu M, Lukatsky DB. Transcription pausing: biological significance of thermal fluctuations biased by repetitive genomic sequences. *Transcription* 2017;**1**:8.
 46. Afek A, Schipper JL, Horton J et al. Protein-DNA binding in the absence of specific base-pair recognition. *Proc Natl Acad Sci USA* 2014;**111**:17140–5.
 47. Takeuchi K, Imai M, Shimada I. Dynamic equilibrium on DNA defines transcriptional regulation of a multidrug binding transcriptional repressor, LmrR. *Sci Rep* 2017;**7**:267.
 48. Imashimizu M, Afek A, Takahashi H et al. Control of transcriptional pausing by biased thermal fluctuations on repetitive genomic sequences. *Proc Natl Acad Sci USA* 2016;**113**: E7409–17.
 49. Helder S, Blythe AJ, Bond CS, Mackay JP. Determinants of affinity and specificity in RNA-binding proteins. *Curr Opin Struct Biol* 2016;**38**:83–91.
 50. Hayashi T, Oshima H, Mashima T et al. Binding of an RNA aptamer and a partial peptide of a prion protein: crucial importance of water entropy in molecular recognition. *Nucleic Acids Res* 2014;**42**:6861–75.
 51. Gelinis AD, Davies DR, Janjic N. Embracing proteins: structural themes in aptamer-protein complexes. *Curr Opin Struct Biol* 2016;**36**:122–32.
 52. Spill F, Weinstein ZB, Irani Shemirani A et al. Controlling uncertainty in aptamer selection. *Proc Natl Acad Sci USA* 2016; **113**:12076–81.
 53. Hamada M. Fighting against uncertainty: an essential issue in bioinformatics. *Brief Bioinformatics* 2014;**15**:748–67.
 54. Adachi H, Ishiguro A, Hamada M et al. Antagonistic RNA aptamer specific to a heterodimeric form of human interleukin-17A/F. *Biochimie* 2011;**93**:1081–8.
 55. Gruber AR, Lorenz R, Bernhart SH et al. The Vienna RNA web-suite. *Nucleic Acids Res* 2008;**36**:W70–4.
 56. Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics* 2015;**31**:3377–9.