NAR Breakthrough Article

Directed evolution of orthogonal RNA–RBP pairs through library-vs-library *in vitro* selection

Keisuke Fukunaga [ID] and Yohei Yokobayashi [ID]*

Nucleic Acid Chemistry and Engineering Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904 0495, Japan

ABSTRACT

RNA-binding proteins (RBPs) and their RNA ligands play many critical roles in gene regulation and RNA processing in cells. They are also useful for various applications in cell biology and synthetic biology. However, re-engineering novel and orthogonal RNA–RBP pairs from natural components remains challenging while such synthetic RNA–RBP pairs could significantly expand the RNA–RBP toolbox for various applications. Here, we report a novel library-vs-library in vitro selection strategy based on Phage Display coupled with Systematic Evolution of Ligands by EXponential enrichment (PD-SELEX). Starting with pools of 1.1×10^{12} unique RNA sequences and 4.0×10^8 unique phage-displayed L7Ae-scaffold (LS) proteins, we selected RNA–RBP complexes through a two-step affinity purification process. After six rounds of library-vs-library selection, the selected RNAs and LS proteins were analyzed by next-generation sequencing (NGS). Further deconvolution of the enriched RNA and LS protein sequences revealed two synthetic and orthogonal RNA–RBP pairs that exhibit picomolar affinity and >4000-fold selectivity.

GRAPHICAL ABSTRACT
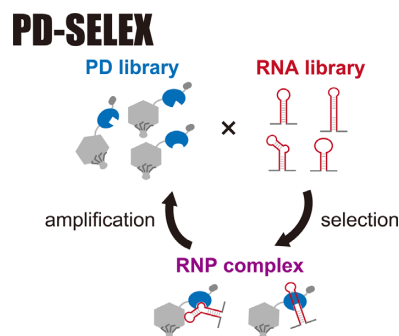
INTRODUCTION

RNA and RNA-binding protein (RBP) pairs, natural or engineered, are used as adaptors to link an RNA of interest with a protein of interest for various applications. For example, the specific interaction between the bacteriophage MS2 coat protein (MS2CP) and MS2 RNA stem–loop ([1]) has been exploited for fluorescence imaging of cellular mRNAs ([2,3]), proximity biotinylation of proteins using the labeling enzyme APEX2 ([4]), isolation of small non-coding RNA-binding proteins ([5]), transcriptional activation with engineered CRISPR–Cas9 complex ([6–8]), and post-transcriptional gene circuits ([9–11]). However, the number of RNA–RBP pairs used in these emerging applications, especially for mammalian synthetic biology, is limited ([12]).

A reasonable strategy to generate novel RNA–RBP pairs is to re-engineer the binding characteristic of an existing RNA–RBP pair. This has typically been done in two steps. (i) Mutate either the parental RNA or RBP to abolish binding ([13,14]). (ii) Construct a wholly or partially randomized library of the binding partner (RNA or RBP) and perform affinity selection ([13,15]). The first step is essentially designing a loss-of-function mutant which is relatively straightforward if the structural information (e.g. crystal structure)

---

is available. However, such a variant may not be optimal for binding a novel partner. The second step is a canonical affinity selection based on established methods such as Systematic Evolution of Ligands by EXponential enrichment (SELEX) [16,17] and Phage Display (PD) [18–20].

While these in vitro selection technologies usually involve the selection of an RNA or protein library for binding to a single target, an alternative approach is library-vs-library selection (or screening) which has been employed for identification of protein–protein [21–25], protein–peptide [26], peptide–peptide [27,28], small molecule–protein [29,30] and small molecule–RNA interactions [31]. To our knowledge, however, RNA library-vs-protein library selection has not yet been reported.

Here, we combined PD with SELEX (PD-SELEX) to develop a method to coevolve RNAs and RBPs through library-vs-library in vitro selection (Figure 1). First, we optimized the selection protocol using the box C/D kink-turn (Kt) RNA and L7Ae [32] as a model RNA–RBP pair. This was partly motivated by the frequent use of L7Ae and its RNA target motifs in synthetic biology applications, for example, mRNA delivery [33], mRNA loading into exosomes [34], and gene switches in vitro [32,35], in *Escherichia coli* [36], and in mammalian cells [9,11,12,32,37–45]. We then constructed a PD library displaying L7Ae protein mutants randomly mutated at 8 surface residues and a partially structured RNA library to execute library-vs-library selection (PD-SELEX). The enriched RNA–RBP pairs were deconvoluted by conventional single-RBP versus RNA library selections, and the binding affinity and selectivity of the discovered RNA–RBP pairs were rigorously characterized by surface plasmon resonance (SPR). Two orthogonal synthetic RNA–RBP pairs with picomolar affinities were discovered, and one of them was also found to be orthogonal to the natural L7Ae–RNA complex.

MATERIALS AND METHODS

Materials

Unmodified oligo DNAs were purchased from Eurofins Genomics, FASMAC, or Integrated DNA Technologies (IDT) Japan. Randomized oligo DNAs containing NNK (where N = A, C, G, or T and K = G or T) codons were purchased from Ajinomoto Bio-Pharma or IDT Japan. DNA oligo pool was purchased from Twist Bioscience. Synthetic genes were purchased from Eurofins Genomics, IDT Japan, or Twist Bioscience. Biotinylated DNA and 2′-OMe RNA were purchased from FASMAC. Q5 High-Fidelity 2× Master Mix and HiScribe T7 High Yield RNA Synthesis Kit were purchased from New England Biolabs (NEB) for polymerase chain reaction (PCR) and in vitro transcription, respectively. SuperScript IV reverse transcriptase was purchased from Thermo Fisher Scientific for reverse transcription (RT). All chemicals used for buffer preparation were purchased from Nacalai Tesque or FUJIFILM Wako Pure Chemical Corporation. Difco LB broth Miller was purchased from BD for bacterial culture. *E. coli* Rosetta-gami B 5615 strain is not commercially available, because the manufacturer has discontinued production. Accordingly, we generated the strain in the laboratory as follows: pAR5615 plasmid DNA encoding isopropyl β-D-1-thiogalactopyranoside

(IPTG)-inducible wild-type T7 phage capsid gene was purified from BLT5615 strain (Merck-Millipore) using Zyppy Plasmid Miniprep Kit (Zymo Research). The plasmid DNA was transformed into Rosetta-gami B competent cells (Merck-Millipore), and transformants were selected on an LB plate supplemented with antibiotics (50 μg/ml carbenicillin and 34 μg/ml chloramphenicol). MagStrep 'type3' XT beads (2-4090-002) was purchased from IBA Lifesciences. FG HM-NeutrAvidin beads (TAB8848N3171) and magnet stand (TAB4899N12) were purchased from Tamagawa SEIKI.

Preparation of proteins for surface plasmon resonance (SPR) and electrophoretic mobility shift assay (EMSA)

Synthetic genes encoding L7Ae, dL7Ae, LS4, or LS12 with dual tags (N-terminal Twin-Strep-tag and C-terminal 6× His-tag) were subcloned into a pET vector backbone. The bacterial expression vectors (Supplementary Figure S1) were transformed into *E. coli* BL21 (DE3) strain. Transformants were pre-cultured overnight at 37°C in 2 ml of LB medium supplemented with 100 μg/ml carbenicillin and transferred to 100 ml fresh LB medium supplemented with 50 μg/ml carbenicillin. After 1–2 h culture at 30°C, IPTG was added at a final concentration of 0.4 mM, and the cells were further cultured for 20 h at 30°C. The cells were harvested by centrifugation at 6000 × g for 10 min at 4°C and suspended in 20 ml ice-cold His-tag purification buffer A (50 mM sodium phosphate, pH 7.7, 300 mM NaCl, 1 mM 2-mercaptoethanol, 5 mM imidazole). After cell disruption by ultrasonication, lysozyme (4 mg, Nacalai), Triton X-100 (0.1% v/v), MgCl$_2$ (5 mM, NipponGene), DNase I (4 U, NEB), and RNase A (20 μg, NipponGene) were added to the crude cell extracts and incubated at 37°C for 30 min. The cell extracts were centrifuged at 8000 × g for 20 min at 4°C to remove cell debris, and the supernatants were filtered. Soluble proteins were loaded onto TALON metal affinity resin (TaKaRa) in a PD-10 (Cytiva) column and washed with 10 ml His-tag purification buffer A. His-tagged proteins were eluted by His-tag elution buffer (20 mM HEPES–KOH, pH 7.5, 200 mM NaCl, 1 mM 2-mercaptoethanol, 200 mM imidazole, 0.01% Tween 20 and 5% v/v glycerol), and the eluents were incubated with Strep-Tactin XT Superflow high capacity resin (IBA Lifesciences). After 5 min incubation, the resin was washed twice with 0.5 ml His-tag purification buffer A, and Twin-Strep-tagged proteins were eluted twice with 0.5 ml 1× Buffer BXT Strep-Tactin XT Elution Buffer (IBA Lifesciences). The eluents were desalted and concentrated using an ultrafiltration device (Amicon Ultra 0.5 ml filter, 3 kDa cut-off, Merck-Millipore) and buffer-exchanged into protein storage buffer (10 mM HEPES–KOH, pH 7.5, 140 mM KCl, 10 mM NaCl, 1 mM MgCl$_2$, 0.1 mM TCEP-Na). Quality of the purified proteins was confirmed by SDS-PAGE (Mini-PROTEAN TGX gel system, Bio-Rad) followed by Coomassie blue staining (CBB Protein Safe Stain, TaKaRa) (Supplementary Figure S2). The concentrations of the recombinant proteins were determined using the standard curve generated from the band intensities of serially diluted bovine serum albumin (BSA) (Pierce Bovine Serum Albumin Standard Ampules, Thermo Fisher Scientific).
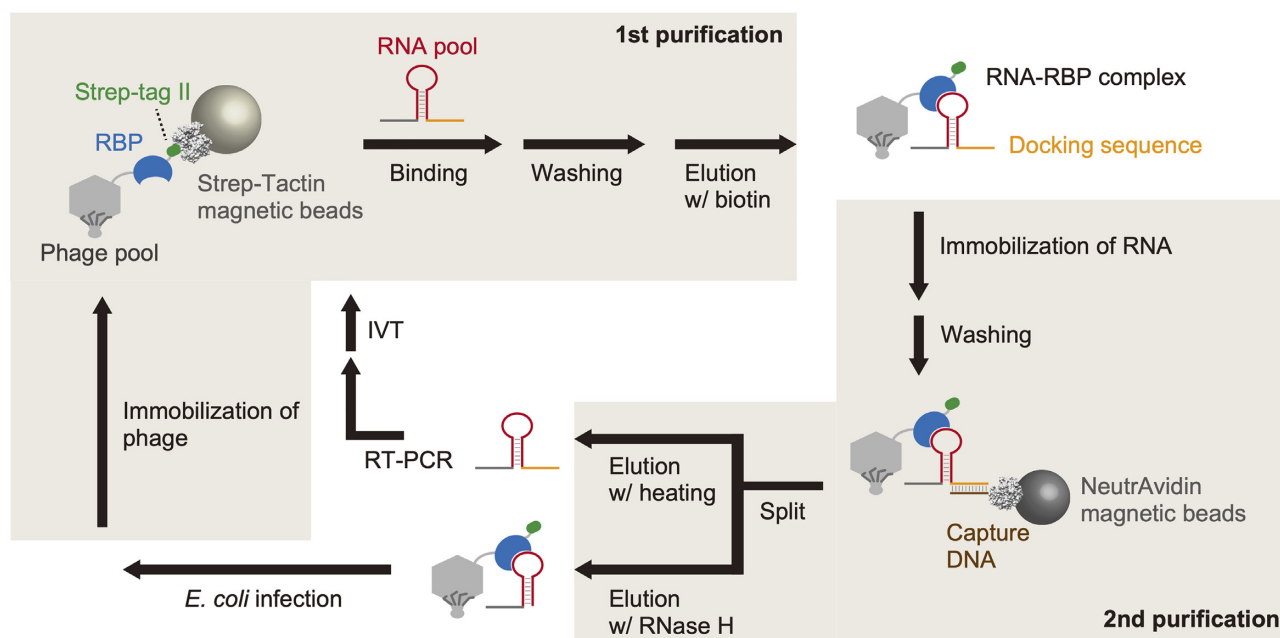
Figure 1. Schematic illustration of PD-SELEX. RBPs are displayed on T7 phage and immobilized on Strep-Tactin magnetic beads via Strep-tag II fused to the C-terminus of RBP. After capturing RBP-binding RNAs, the phage particles are eluted, and phage-displayed RBP–RNA complexes are captured by magnetic beads immobilized with an oligo DNA complementary to the fixed RNA sequence (docking sequence). RNA and phage particles are separately eluted and amplified for the subsequent round of selection.

Preparation of RNAs for SPR

Partially double-stranded DNAs or PCR products (46) were used as a template for *in vitro* transcription. RNA aptamers were synthesized by *in vitro* transcription at 42°C for 4 h using HiScribe T7 High Yield RNA Synthesis Kit (NEB) according to the manufacturer's manual for *in vitro* transcription of short templates (Supplementary Table S1). Transcription products were treated with 2 U TURBO DNase (Thermo Fisher Scientific) for 60 min at 37°C and purified with RNA Clean & Concentrator-25 Kit (Zymo Research). RNA concentrations were determined by absorbance at 260 nm according to OligoCalc (47).

SPR

Binding kinetics were determined using Biacore T200 (Cytiva) at 25°C. CM5 chip surface was washed by three injections of 10 mM NaOH at a flow rate of 20 μl/min for 30 s. Next, a mixture of 200 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and 50 mM *N*-hydroxysuccinimide (NHS) was injected into all flow cells at a flow rate of 25 μl/min for 7 min to activate carboxylic acid groups. After injection of ultrapure water for 1 min, 0.3 mg/ml NeutrAvidin in 10 mM sodium acetate (pH 5.2) was injected by setting the immobilization level to 5000 RU. Finally, 1 M ethanolamine–HCl (pH 8.5) was injected for 7 min to quench the activated carboxylic acid. To remove the non-covalently bound NeutrAvidin on the chip surface, 10 mM NaOH was injected three times at a flow rate of 20 μl/min for 20 s. Setting the immobilization level to 400 RU, biotinylated 2′-OMe RNAs (5′-/bio/cguucuguguucuuucgucgau-3′) (1 μM) in high salt

buffer (10 mM Tris–HCl, pH 7.5, 1 M NaCl, 1 mM EDTA-Na) was injected and the remaining biotin-binding sites on the chip were quenched by injection of 1 μM free biotin.

RNA aptamer solutions (500 nM) were prepared in water and denatured by heating at 80°C for 3 min. After cooling at room temperature for few minutes, the RNA solutions were diluted to 10–50 nM with the high salt buffer. Various concentrations of protein solutions were prepared in SPR running buffer (10 mM HEPES, pH 7.4, 150 mM NaCl, 0.05% v/v Surfactant P20, 1 mM $MgCl_2$, 0.25 mM TCEP-Na, 0.1 mg/ml BSA, 25 μg/ml yeast tRNA).

RNAs were injected at a flow rate of 10 μl/min for 30 s, and then the running buffer was injected for 1 min. Proteins were injected at a flow rate of 80 μl/min for 100 s (or 200 s) to monitor association, and then the dissociation kinetics was monitored for 420 s (or 1500 s) in SPR running buffer. The sensor surface was regenerated with 2 μl injection of 10 mM NaOH followed by 6 M urea at a flow rate of 20 μl/min and the running buffer was injected for 1 min. The raw data were analyzed by Biacore T200 Evaluation Software 1.0 (Cytiva) using 1:1 Langmuir interaction model. Background signal from the reference flow cell was subtracted from that of the sample flow cell, and no-protein sample (SPR running buffer only) was injected in each experiment (double referencing). Dissociation constant (K_D) was determined from the ratio of the association and the dissociation rate constants ($K_D = k_{off} / k_{on}$). The figures were generated using GraphPad Prism software 6.0h (GraphPad Software). The measurements were repeated at least twice using different flow cells (or different SPR chips) in different days to ensure reproducibility.

Preparation of affinity beads

MagStrep 'type3' XT beads were rinsed three times with ultrapure water and stored in blocking buffer (Pierce Protein-Free TBS Blocking Buffer, Thermo Fisher Scientific) at 4°C. FG HM-NeutrAvidin beads (100 μl) were rinsed once with immobilization buffer (10 mM Tris–HCl, pH 7.5, 1 M NaCl, 0.1 mM EDTA-Na), and mixed with 1.64 nmol of biotinylated capture oligo DNA (5′-/bio/CGTTCTGTGTCTTTCGTCGAT-3′) in the same buffer and incubated at room temperature for 5 min. After washing with the same buffer three times, FG beads were stored in blocking buffer at 4°C. The immobilized oligo DNA was estimated to be ∼6.3 pmol per μl suspended beads. The beads were pre-equilibrated with selection buffer (10 mM HEPES–KOH, pH 7.5, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$, 0.01% Tween 20, 0.1 mM TCEP-Na, 2% v/v glycerol) before use.

Construction of PD library

A. fulgidus rpl7ae gene was sub-cloned into T-vector pMD20 (TaKaRa), and the sequence was verified by Sanger sequencing. The partially randomized DNA library encoding L7Ae mutants was prepared by overlap extension PCR (48) as summarized in Supplementary Figure S3. Briefly, three fragments outside of the randomized regions were amplified by PCR using appropriate primer pairs (EcoRI-T7PD_Fw and L7Ae-Fr1_Rv, L7Ae-Fr2_Fw and L7Ae-Fr2_Rv, L7Ae-Fr3_Fw and HindIII-T7PD_Rv; Supplementary Table S2). Next, two of these fragments were used as templates to add degenerate codons (NNK) at the targeted positions in L7Ae using appropriate primers (L7Ae-Fr2_lib_Fw and L7Ae-Fr2_Rv, L7Ae-Fr3_lib_Fw and HindIII-T7PD_Rv; Supplementary Table S2). Finally, three DNA fragments containing overlap regions were assembled by overlap extension PCR using primers EcoRI-T7PD_Fw and HindIII-T7PD Rv (Supplementary Table S2). The final PCR product was digested with Eco RI-HF (NEB), Hind III-HF (NEB), Dpn I (NEB), and gel purified using Zymoclean Gel DNA Recovery Kit (Zymo Research). Sequence of the assembled DNA library was confirmed by Sanger sequencing (Supplementary Figure S3). L7Ae-scaffold T7 PD library was constructed using T7Select10-3b system (Merck-Millipore, 70550–3CN). The digested DNA library (∼0.7 pmol) was ligated with T7Select10-3b vector arms (5 μg) with 400 U T4 DNA ligase (NEB) at 16°C for 16 h in 20 μl scale. After addition of 125 μl of T7 packaging extract, in vitro DNA packaging reaction was performed at 22°C for 2 h and quenched by addition of LB medium. Diversity of the PD library was estimated by plaque assay (∼4.0×10^8). The PD library was amplified with logarithmic phase of Rosetta-gami B 5615 strain in 200 ml scale (multiplicity of infection < 0.01). Expression of Gp10A was induced in advance by addition of 1 mM IPTG. Upon lysis (usually 1.5–2 h after infection), 5 g of sodium chloride was added to the lysate and the *E. coli* debris was removed by centrifugation at 8000 × *g* for 20 min at 4°C. After collection of the supernatant, 21 g of polyethylene glycol (PEG) 8000 (Promega) was added and incubated at 4°C overnight. Lysate-PEG mixture was centrifuged at 8000 × *g* for 40 min at 4°C and the supernatant was discarded. The precipitated phage particles were suspended in T7 phage stock buffer (10 mM Tris–HCl, pH 8.0, 1 M NaCl, 1 mM EDTA-Na), and subjected to ultrasonication for few minutes. The phage solution was filtered through a 0.45 μm membrane filter (hydrophilic polytetrafluoroethylene, Merck-Millipore), and stored at 4°C until use. Titer of the amplified phage was quantified by plaque assay. Bacteriophage T7-displayed L7Ae or dL7Ae was confirmed by Western blotting (Supplementary Figure S4). The amino acid distributions of the randomized residues were analyzed by Illumina sequencing (Supplementary Figure S5). We found overrepresentation of amino acids arginine, leucine, serine and other amino acids as would be expected from the NNK degenerate codons.

Construction of RNA library

The template DNA was prepared by primer extension as described previously with modifications (49). Two oligo DNAs (200 pmol each of N20L_Fw and N20L_lib_Rv; Supplementary Table S2) were mixed with Q5 High-Fidelity Master Mix (NEB) in 200 μl scale, and the mixture was denatured at 98°C for 1 min. The oligos were annealed and extended at 72°C for 2 min. The product was analyzed by 2% agarose gel electrophoresis and purified with DNA Clean & Concentrator Kit (Zymo Research). Approximately 16.7 pmol (1.0×10^{13} molecules) of the purified DNA was used as a template in an in vitro transcription reaction at 37°C for 4 h in 200 μl volume (0.75× reaction buffer, 7.5 mM NTPs, 40 U Murine RNase inhibitor, 7.5 μl of T7 RNA polymerase mix). The transcription product was treated with 2 U TURBO DNase (Thermo Fisher Scientific) at 37°C for 60 min and purified by phenol-chloroform extraction followed by ethanol precipitation. Purified RNAs were dissolved in 1× RNA loading dye (NEB), heat-denatured at 80°C for 3 min, and separated by 8% PAGE. The separated sample was visualized by UV shadowing to excise the gel fragment and the RNA was eluted overnight in TE buffer. The eluents were filtered through a 0.22 μm membrane filter (hydrophilic polytetrafluoroethylene, Merck-Millipore), and concentrated using an ultrafiltration device (Amicon Ultra 0.5 ml filter, 3 kDa cutoff, Merck-Millipore). RNA concentration was determined from absorbance at 260 nm using NanoDrop One (Thermo Scientific).

PD-SELEX

A detailed summary of the selection parameters is shown in Supplementary Table S3. In the 1st purification, an appropriate titer of the PD pool was mixed with 1/5 volume of 20% PEG-8000, 2.5 M NaCl solution and precipitated by centrifugation at 18 000 × *g* for 10 min at 4°C. The phage particles were dissolved in 500 μl selection buffer and was incubated with 10 μl pre-equilibrated MagStrep 'type3' XT beads at room temperature for 30 min. The phage-immobilized beads were washed once with 500 μl selection buffer and re-suspended in 295 μl of the same buffer. RNA solution of an appropriate concentration was prepared in 200 μl volume, and the RNAs were folded by heating at

80°C for 3 min followed by incubation at room temperature. The folded RNAs were mixed with 4 μl yeast tRNA (40 μg, Ambion), 40 U Murine RNase Inhibitor (NEB), and 395 μl of the phage-immobilized magnetic beads. After washing with selection buffer, phage particles were eluted with 50 μl 1× Buffer BXT Strep-Tactin XT Elution Buffer (100 mM Tris–HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA-Na, 50 mM biotin) (IBA Lifesciences).

In the 2nd purification, the eluent from the 1st purification was mixed with 10 μl of capture oligo DNA-immobilized NeutrAvidin beads and 440 μl selection buffer and incubated. After washing with the same buffer, the beads were re-suspended in 20 μl of TE buffer. The bead suspension was split into two aliquots for separate RNA and phage recovery. To recover RNAs, 20 pmol of reverse primer (N20L_Rv, Supplementary Table S2) was added to the suspension and incubated at 80°C for 3 min. The beads were separated by a magnet and the supernatant was collected and kept on ice. To recover the phage particles, 60 U RNase H (TaKaRa) and 5× RNase H reaction buffer (250 mM Tris–HCl, pH 7.5, 375 mM NaCl, 15 mM MgCl$_2$, 5 mM dithiothreitol) were added to another aliquot of the beads and incubated at 30°C for 15 min. The beads were separated by a magnet and the supernatant containing phage particles was collected.

The recovered RNAs were reverse transcribed at 60°C for 15 min in 20 μl volume using SuperScript IV, and the reaction was quenched by incubation at 85°C for 5 min. The cDNAs were amplified by PCR (20 cycles of 98°C for 10 s, 71°C for 20 s, 72°C for 10 s) with appropriate primers (N20L_Fw and N20L_Rv, Supplementary Table S2). The PCR products were purified using Zymoclean Gel DNA Recovery Kit, and 75% of the purified DNA was used for in vitro transcription (20 μl scale) to produce the RNA pool for the next round. In vitro transcription products were treated with 2 U TURBO DNase (Thermo Fisher Scientific) at 37°C for 60 min and purified with RNA Clean & Concentrator-25 Kit. The recovered phage particles were amplified in 35 ml scale, and the amplified phage particles were purified by the PEG-NaCl method followed by filtration with a 0.45 μm membrane filter.

Illumina sequencing and data analysis

Illumina sequencing libraries were prepared by RT-PCR (for RNA library) or two-step PCR (for PD library). The RNAs were mixed with a barcoded reverse transcription primer (N20L_MiSeq_bc_RT, Supplementary Table S2) and denatured at 80°C for 3 min. After cooling on ice, reverse transcription was carried out at 60°C for 15 min. The cDNA was amplified by 10 cycles of PCR using appropriate primers (N20L_MiSeq_P7_F and MiSeq_R1seq_P5_R, Supplementary Table S2).

The phage genomic DNAs were amplified by 6 cycles of PCR using barcoded primers (R1_bc_LS_Fw and R2_LS_Rv, Supplementary Table S2), and the amplified DNAs were purified with DNA Clean & Concentrator Kit. The purified first PCR products were further amplified by 10 cycles of PCR using appropriate primers (P5_R1_Fw and P7_R2_Rv, Supplementary Table S2) to add P5 and P7 adapter sequences.

The barcoded Illumina libraries were separated by 3% agarose gel and purified with Zymoclean Gel DNA Recovery Kit. The purified libraries were quantified by real-time PCR using NEBNext Library Quant Kit for Illumina (NEB) according to manufacturer's instructions. PhiX control v3 DNA (10%, Illumina) was spiked into the libraries as an internal control, and sequencing was performed on Illumina MiSeq sequencer using MiSeq Reagent Kit v3 (150 cycle single read or 300 cycle paired-end read). Raw sequence data (fastq files) were analyzed and visualized by custom Python scripts.

Construction of second-generation RNA library

The oligo pool (426 fmol) containing 2000 distinct sequences was amplified by PCR (25 cycles of 98°C for 10 s, 70°C for 20 s, 72°C for 10 s) using N20L_Fw and N20L_Rev (Supplementary Table S2) as primers. The PCR product was purified using Zymoclean Gel DNA Recovery Kit, and the purified DNA (18.7 pmol) was used for in vitro transcription in 100 μl volume to obtain the second-generation RNA library. Transcription products were treated with 2 U TURBO DNase (Thermo Fisher Scientific) and 20 U Exonuclease I (NEB) at 37°C for 30 min and purified with RNA Clean & Concentrator-25 Kit.

Preparation of affinity beads for re-selection

Bacterial expression vectors (Supplementary Figure S1) were introduced into *E. coli* BL21 (DE3) strain. The transformants were cultured overnight at 30°C in 2 ml of MagicMedia Medium (Thermo Fisher Scientific) supplemented with 50 μg/ml carbenicillin. The cells were harvested and suspended in ice-cold His-tag purification buffer B (20 mM HEPES–KOH, pH 7.5, 200 mM NaCl, 1 mM 2-mercaptoethanol, 1 mM imidazole, and protease inhibitor cocktail without EDTA). After cell disruption by ultrasonication, Triton X-100 (0.05% v/v) was added to the crude cell extract. The extract was centrifuged at 18 000 × *g* for 10 min at 4°C to remove cell debris. The supernatant was incubated with TALON metal affinity resin (30 μl, Clonetech) for 10 min at 4°C. After washing several times with the same buffer, His-tagged protein was eluted with His-tag elution buffer. His-tag purified proteins were confirmed by Western blotting with Strep-Tactin-HRP conjugate (1:10 000 dilution, Bio-Rad) (Supplementary Figure S6).

MagStrep 'type3' XT beads (10 μl, Clonetech) were rinsed three times with ultrapure water and then incubated with blocking buffer (Pierce Protein-Free TBS Blocking Buffer, Thermo Fisher Scientific) at 4°C overnight. His-tag purified proteins were incubated with 10 μl of MagStrep 'type3' XT beads at 4°C for 30 min. After washing with 500 μl each of high salt buffer (10 mM Tris–HCl, pH 7.5, 500 mM NaCl, 1 mM EDTA-Na) and tTBS (25 mM Tris–HCl, pH 7.4, 137 mM NaCl, 27 mM KCl, 0.05% Tween 20) (Nacalai), the beads were stored in blocking buffer at 4°C.

Re-selection of second-generation RNA pool

RNA solution (0.2 μM) was prepared in 500 μl selection buffer supplemented with 0.5 M urea, and the RNAs were folded by heating to 80°C for 3 min followed by incubation at room temperature. The folded RNAs were mixed with 1 μl yeast tRNA (10 μg, Ambion) and 5 μl recombinant protein-immobilized beads, and incubated at 37°C for 30 min. After washing eight times with 500 μl sample buffer, RNAs were eluted with 50 μl 1× Buffer BXT Strep-Tactin XT Elution Buffer (IBA Lifesciences). The eluents were purified and concentrated by ethanol precipitation and then dissolved in 10 μl TE.

RESULTS

Design of PD-SELEX

We designed the following strategy to enrich complexes of RNA and phage-displayed protein from unbound species via two-step affinity purification (Figure 1). First (Figure 1, 1st purification), phage particles are immobilized on Strep-Tactin magnetic beads via Strep-tag II (50) which is fused to the C-terminus of the phage-displayed protein. The RNA pool is added to the phage-immobilized magnetic beads and allowed to form RNA–RBP complexes. Then the beads are washed to remove unbound RNAs. The phage particles (bound and unbound to RNA) are recovered from the beads by competitive elution with D-biotin. In the second step (Figure 1, 2nd purification), phage-bound RNAs are captured by DNA-immobilized NeutrAvidin magnetic beads. The DNA sequence is complementary to the fixed sequence at the 3′ end of the RNA library (docking sequence). It should be noted that the Strep-tag II fused to RBP does not bind to NeutrAvidin (51,52). The beads are then washed to remove unbound phage particles.

Through this two-step affinity purification protocol, complexes of RNA and phage-displayed protein are enriched. The beads are split into two aliquots to recover the enriched RNAs and phage particles separately. RNAs and phage particles are recovered by heating and ribonuclease H (RNase H) treatment, respectively. As RNase H cleaves the RNA strand in DNA–RNA hybrid, it helps to reduce phage particles that non-specifically bind to the capture DNA, NeutrAvidin, or magnetic beads. The recovered RNAs are amplified by reverse transcription (RT) followed by polymerase chain reaction (PCR) and in vitro transcription (IVT). The recovered phage particles are used to infect *E. coli* to be amplified. The amplified RNA and phage pools are then used for the subsequent round of selection. After several rounds of selection, the enriched RNA and phage-displayed RBP sequences are analyzed by next-generation sequencing (NGS).

PD-SELEX mock selection

To establish the PD-SELEX protocol described above, we designed and executed a mock selection using the natural box C/D Kt and L7Ae (32) as a model RNA–RBP pair, and a defective k-turn (dKt) RNA with no affinity to L7Ae (42) as a negative control RNA. Additionally, we designed dL7Ae, a non-functional variant of *A. fulgidus* L7Ae that

lacks affinity for box C/D Kt as a negative control protein (Figure 2A). Three charged amino acids (E34, K37, R41) in the α2 helix of L7Ae are evolutionarily conserved in archaea to eukaryotes (53). It is known that K37 and R41 interact with the phosphates of Watson-Click base pairs in Kt RNA, and E34 interacts with guanine in the G●A pair adjacent to the kinked buldge in Kt RNA (54). We substituted three of these charged amino acids with oppositely charged residues (E34K, K37E, R41E). The amino acid sequence of the loop connecting the α4 helix and β4 sheet is not conserved between archaea and eukaryotes (53), and three residues I88, E89, and V90 of *A. fulgidus* L7Ae recognize nucleobases in the bulge region of Kt RNA (54). Therefore, we also substituted five amino acids in the loop to glycine or serine (I88G, E89G, V90G, P91G, C92S). dLA7e, with 8 amino acid substitutions (E34, K37, R41, I88G, E89G, V90G, P91G, C92S) lost binding affinity to box C/D Kt and did not bind to dKt as measured by SPR at least within the range of measured protein concentrations (Figure 2B). We also confirmed that L7Ae does not bind to dKt which is consistent with the previous report (42). The observed dissociation constant (K_D) between L7Ae and box C/D Kt was 2.13×10^{-11} M in physiological salt and magnesium concentrations at pH 7.4 (Figure 2B and Table 1). This is comparable to the reported K_D value of 10 pM between L7Ae and another Kt RNA (ribosomal Kt-7) measured by stopped-flow kinetic analysis (55).

Next, we generated T7 phage particles displaying L7Ae or dL7Ae with a C-terminal Strep-tag II using the T7Select10-3b system (Figure 2C), and the protein expression was confirmed by Western blotting (Supplementary Figure S4). Also, box C/D Kt and dKt RNA sequences flanked by fixed regions (Figure 2C) were prepared by in vitro transcription. Mock PD libraries were prepared by mixing phage particles displaying L7Ae and dL7Ae at 1:99 or 5:95 ratio. Similarly, box C/D Kt and dKt RNAs were mixed in 1:99 or 5:95 ratio to prepare mock RNA libraries. As illustrated in Figure 1, 1×10^{11} plaque forming unit (PFU) of a mock PD library was immobilized on Strep-Tactin magnetic beads. Next, the beads were incubated with 0.2 μM mock RNA library in 500 μl buffer for 10 min, and the beads were washed three times with the same buffer. The phage particles and the RNAs were eluted with D-biotin and then incubated with DNA-immobilized NeutrAvidin magnetic beads in 500 μl of buffer for 10 min to capture RNA-phage complexes. After washing three times, RNAs were eluted, reverse transcribed, and amplified by PCR. Since dKt was designed to be 5-nt longer than box C/D Kt, relative abundance of the two RNA species could be evaluated by gel electrophoresis of the PCR products (Figure 2D). Box C/D Kt RNA was enriched 6.0-fold or 4.3-fold after a single round of selection as estimated from the intensities of the two bands (Figure 2D). Similarly, the mutated region of the L7Ae/dL7Ae proteins was PCR amplified from the phage genomic DNA before and after the mock selection. The PCR products were then digested with Hae III or Bam HI which specifically cleaves L7Ae or dL7Ae PCR products, respectively (Supplementary Figure S7). Again, 11.4-fold or 4.0-fold enrichment of L7Ae was observed after a single round of mock selection (Figure 2E). These results demonstrate that the two-step affinity purification protocol
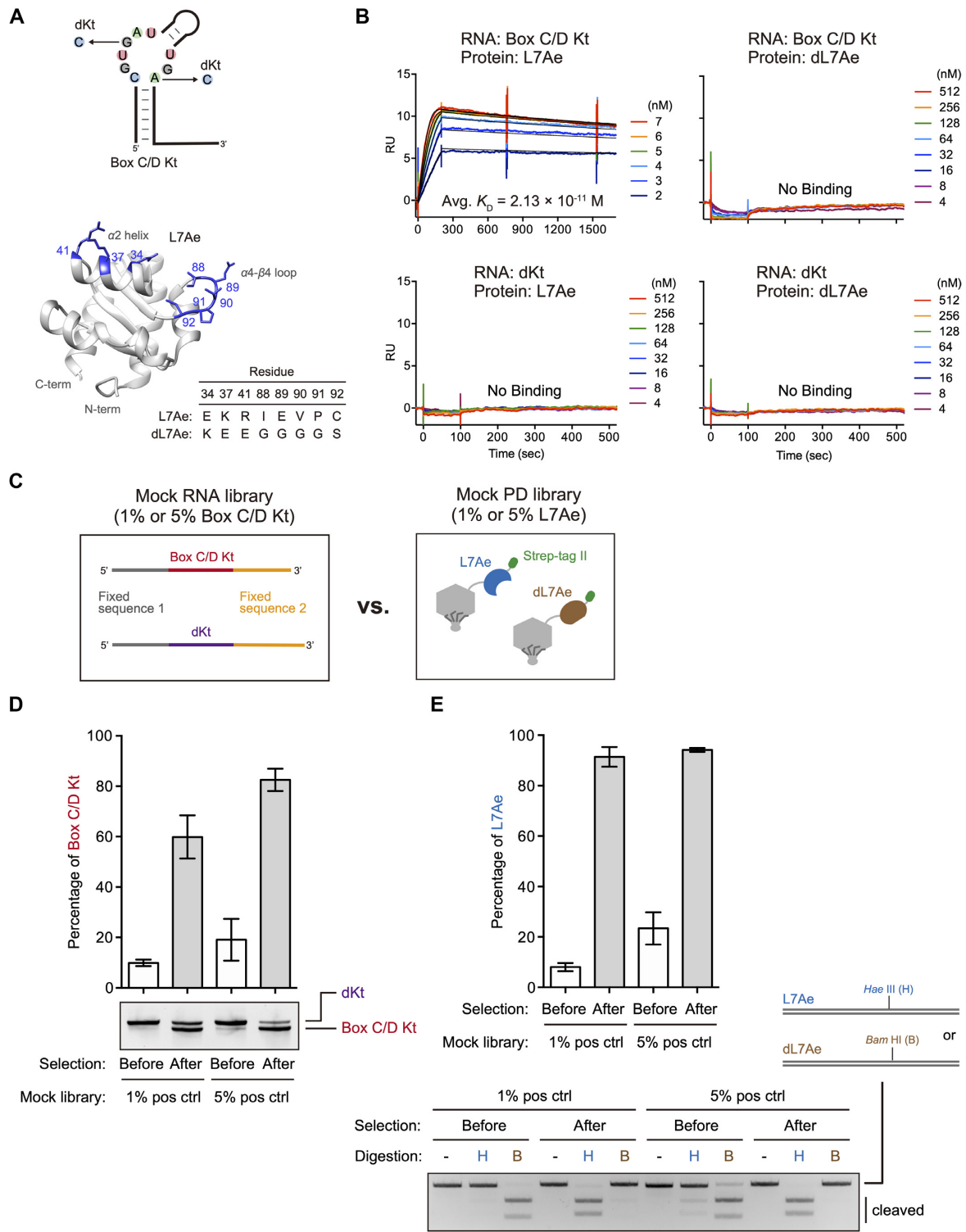
Figure 2. Mock selection. (**A**) Schematic representation of the box C/D Kt and dKt (defective Kt) RNAs, and *A. fulgidus* L7Ae structure (PDB ID, 4BW0) (54) depicting the mutated residues in dL7Ae. The graphic was generated using UCSF Chimera (71). (**B**) Confirmation of binding specificity of the RNAs (box C/D Kt or dKt) and the recombinant proteins (L7Ae or defective L7Ae: dL7Ae) by SPR. Sensorgrams are shown in colored lines, and black lines indicate curve fitting according to 1:1 binding model. RU stands for response unit. The observed kinetic values and dissociation constants are summarized in Table 1. For box C/D Kt–L7Ae, dissociation time was extended to 25 min from 7 min to monitor slow dissociation. (**C**) Schematic illustration of the mock libraries. (**D**) Mock RNA library selection. RT-PCR products (20 ng) were separated by native PAGE and visualized by SYBR Gold staining. The sizes of the box C/D Kt and dKt are 91 bp and 96 bp, respectively. The band intensities were quantified using ImageJ. The bar graph represents means and standard deviations of three independent experiments. (**E**) Mock PD library selection. PCR products (100 ng) were digested with a restriction enzyme Hae III (H) or Bam HI (B), and then separated by 2% agarose gel electrophoresis. DNAs were visualized by ethidium bromide staining. The band intensities were quantified using ImageJ, and the bar graph represents means and standard deviations of three independent experiments.

Table 1. Binding properties of model RNAs and proteins determined by SPR.

RNA	Protein	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_D (M)
Box C/D Kt	L7Ae	$(1.41 \pm 1.11) \times 10^7$	$(2.77 \pm 1.75) \times 10^{-4}$	$(2.13 \pm 0.43) \times 10^{-11}$
	dL7Ae	N.B.	–	–
dKt	L7Ae	N.B.	–	–
	dL7Ae	N.B.	–	–

k_{on}, k_{off}, K_D values are means and standard deviations of at least two independent experiments.
N.B.: no binding detected at the protein concentrations tested.

efficiently enriches RNA–RBP pairs from the background of non-binding RNAs and RBPs.

Library-vs-library selection of RNA–RBP pairs

After establishing the selection protocol, we set out to perform library-vs-library selection of novel RNA–RBP pairs. We designed a hairpin loop RNA library in which 20-nt in the loop region were randomized which contains up to 1.1×10^{12} unique sequences (Figure 3A, left) with an expectation that structured RNA pools contain higher frequency of RNA aptamers (56). Fixed sequences were added to the 5′ and 3′ ends of the hairpin loop library for amplification and affinity purification. *A. fulgidus* L7Ae was used as a scaffold for the RBP library. Guided by the reported crystal structure (54), three charged residues in the α2 helix (E34X, K37X, E40X, where X stands for 20 amino acids) and five residues (I88X, E89X, V90X, P91X, C92X) in the loop connecting α4 helix and β4 sheet were randomized. We did not randomize R41 because it is considered to form a universal hydrogen bond with a phosphate group in Kt RNA (57) which does not appear to be involved in RNA sequence selectivity. Although the theoretical diversity of the L7Ae scaffold (LS) PD library is 2.56×10^{10} ($= 20^8$), diversity of the actual library is limited by the efficiency of in vitro packaging reaction of the phage genomic DNA which was estimated to be approximately 4.0×10^8 based on plaque assay. The PD library size is comparable to those of the previously reported T7 PD libraries (20,58,59). Fewer amino acids should be randomized if complete or near-complete coverage of the amino acid combinations is desired. We opted to randomize more residues than can be fully represented in the PD library because our goal was to discover novel RBPs with new RNA sequence selectivity, and we did not know which residues are more important for recognizing new RNA motifs yet to be discovered.

Six rounds of library-vs-library selection were carried out. Stringency of the selection was gradually increased by lowering the amounts of RNA and PD pools used, shortening the binding time, and increasing the number of washing steps (Supplementary Table S3). In the first two rounds, negative selections were performed to avoid enrichment of non-specific variants that bind to magnetic beads and/or phage particles. In the T7Select10-3b system used for PD library construction, roughly 10 copies of proteins are displayed on each virion which could increase the apparent affinity to the immobilized targets due to multivalency. Therefore, in the last two rounds, we added 0.5 M urea as a denaturant to the binding/washing buffer to increase selection pressure.

After eluting the RNA and the PD pools, the initial and the selected sequences after rounds 4 (PD only), 5 and 6 were analyzed by Illumina sequencing. As expected, almost all sequences in the initial RNA and PD pools (99.8% of RNA and 99.2% of PD) were sequenced only once (Figure 3B and C). RNA and protein sequences were enriched through the selection (Figure 3B and C), and significant enrichment of both RNA and phage populations was observed prior to the 6th round. For each protein sequence, enrichment factor (EF) was calculated by dividing the abundance of the sequence after round N by that of round N-1. After filtering the protein sequence data to keep those that are enriched from rounds 4 through 6 (EFs from round 4 to 5 and round 5 to 6 ≥ 1.0), the 20 most abundant variants in round 6 of the PD pool (LS1-LS20) were selected for further analysis (Figure 3C). The enriched LS proteins showed no similarity to the parental L7Ae (Figure 3C) and the archaeal L7Ae homologs (Supplementary Figure S8A). However, some sequence trends were found among the laboratory-evolved proteins (Supplementary Figure S8B). For example, tryptophan or arginine was preferred in the 89th residue in the 20 most abundant sequences. Preference for arginine at residue 89 (but not tryptophan) was still observed when the top 1000 sequences were examined. Similarly, arginine was frequently observed in the 90th residue in both top 20 and top 1000 sequences. No strong preferences at other positions were observed (Supplementary Figure S8B).

Deconvolution of RNA–LS protein pairs

Sequencing of the enriched RNA and LS protein sequences does not indicate which RNA interacts with which LS protein. To discover specific RNA–LS protein pairs, we prepared recombinant LS1–LS20 and L7Ae proteins and immobilized them individually on magnetic beads. We also synthesized the 1991 most abundant RNA sequences after the 6th round and 9 control RNAs (L7Ae-binding RNAs: box C/D Kt, Kt-7, Kt-15, lysine riboswitch Kt, SAM riboswitch Kt, H/ACA sRNA loop, sR8 C'/D' box loop, H23 aptamer, and negative control: dKt) (13,42) by in vitro transcription from a dsDNA template prepared from a custom oligo pool. The 2000-sequence RNA library was selected for binding against each immobilized LS protein (and L7Ae) with stringent binding and washing steps (Figure 4A). The eluted RNAs were reverse transcribed with a barcoded primer and PCR amplified for Illumina sequencing. For each RNA, an EF was calculated as: EF = abundance after selection / abundance before selection.

A simple inspection of the RNA sequences after PD-SELEX revealed two highly enriched consensus sequences. Consensus sequence 1 (UUGUGASGC, S = C or G) was
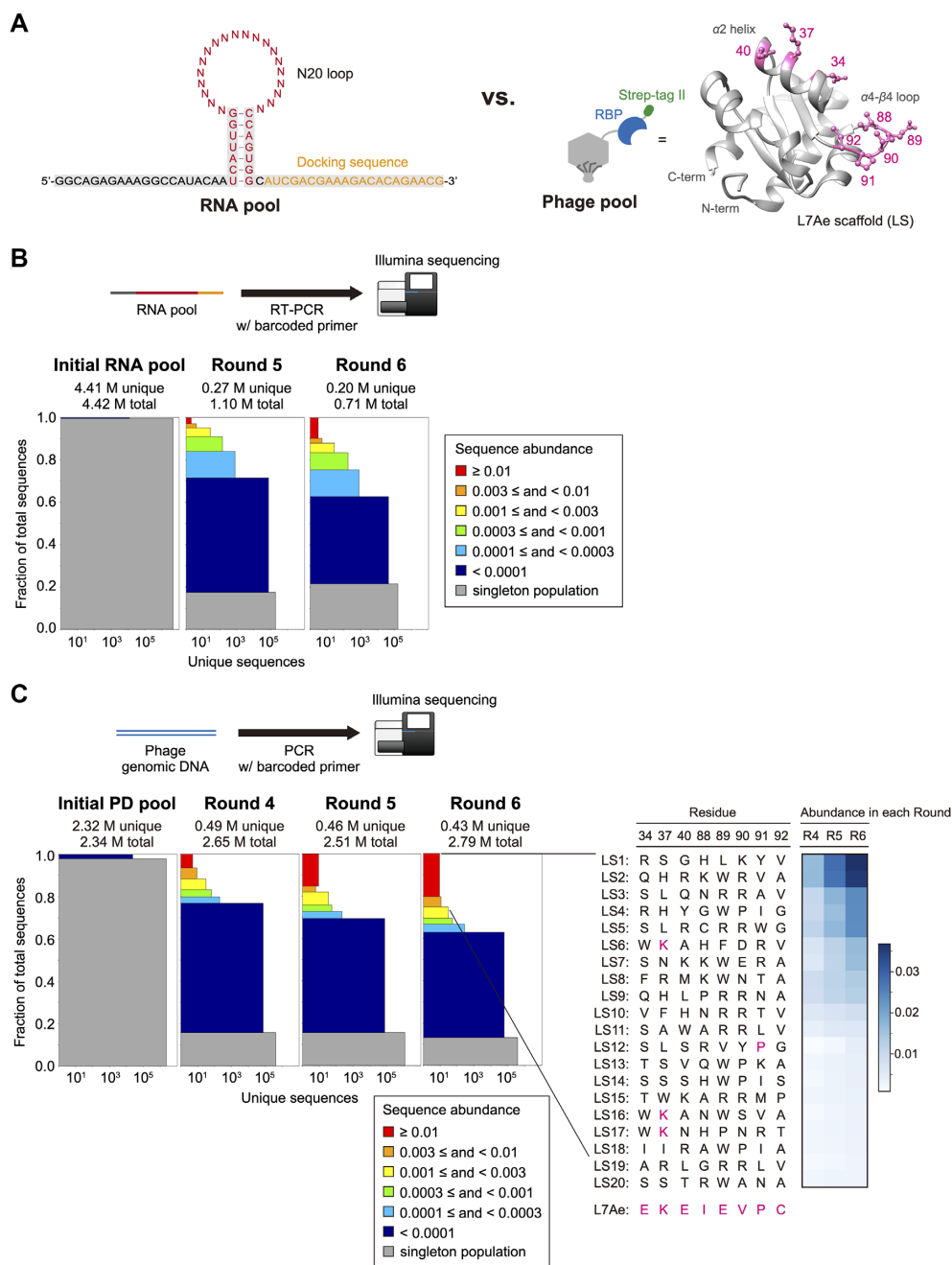
Figure 3. RNA library vs. PD library in vitro selection. (**A**) Design of the hairpin N20-loop RNA library (left, N = A, C, G or U) and L7Ae-scaffold protein library (right). Randomized amino acid residues are shown in pink. Population diversity (72,73) of the initial and the selected (**B**) RNA pool and (**C**: left) PD pool. Multi-copy sequences were sorted into six colored bins according to abundance. Single-count sequences are shown in the bottom grey bins (singleton population). The height of each bin represents its proportion in the population. The width represents the number of unique sequences. (**C**: right) Amino acid sequences of LS1-LS20. The amino acid residues found in L7Ae are shown in pink. The heatmap shows the abundance of each LS variant in rounds 4, 5 and 6.

found in 57% of the population, and consensus sequence 2 (UCCAUGACGC) was found in 25% of the population after round 6. The 2000-sequence RNA library also contained 59% and 33% of the consensus sequence 1 and 2, respectively. In contrast, the unselected pool contained 0.0109% (consensus sequence 1) and 0.0011% (consensus sequence 2) of the two sequences. Consequently, we focused on binding of these consensus sequences to the individual LS proteins.

Enrichment of the RNAs possessing consensus sequence 1 or 2 was statistically analyzed and visualized by box plots (Figure 4B and C). Judging from the median values, RNAs containing consensus sequence 1 were significantly enriched by LS4 and LS18 (Figure 4B). Interestingly, LS4 and LS18 share the same subsequence in the loop region (W89, P90, and I91), suggesting that these residues contribute to the affinity for consensus sequence 1 (Figure 3C). In the case
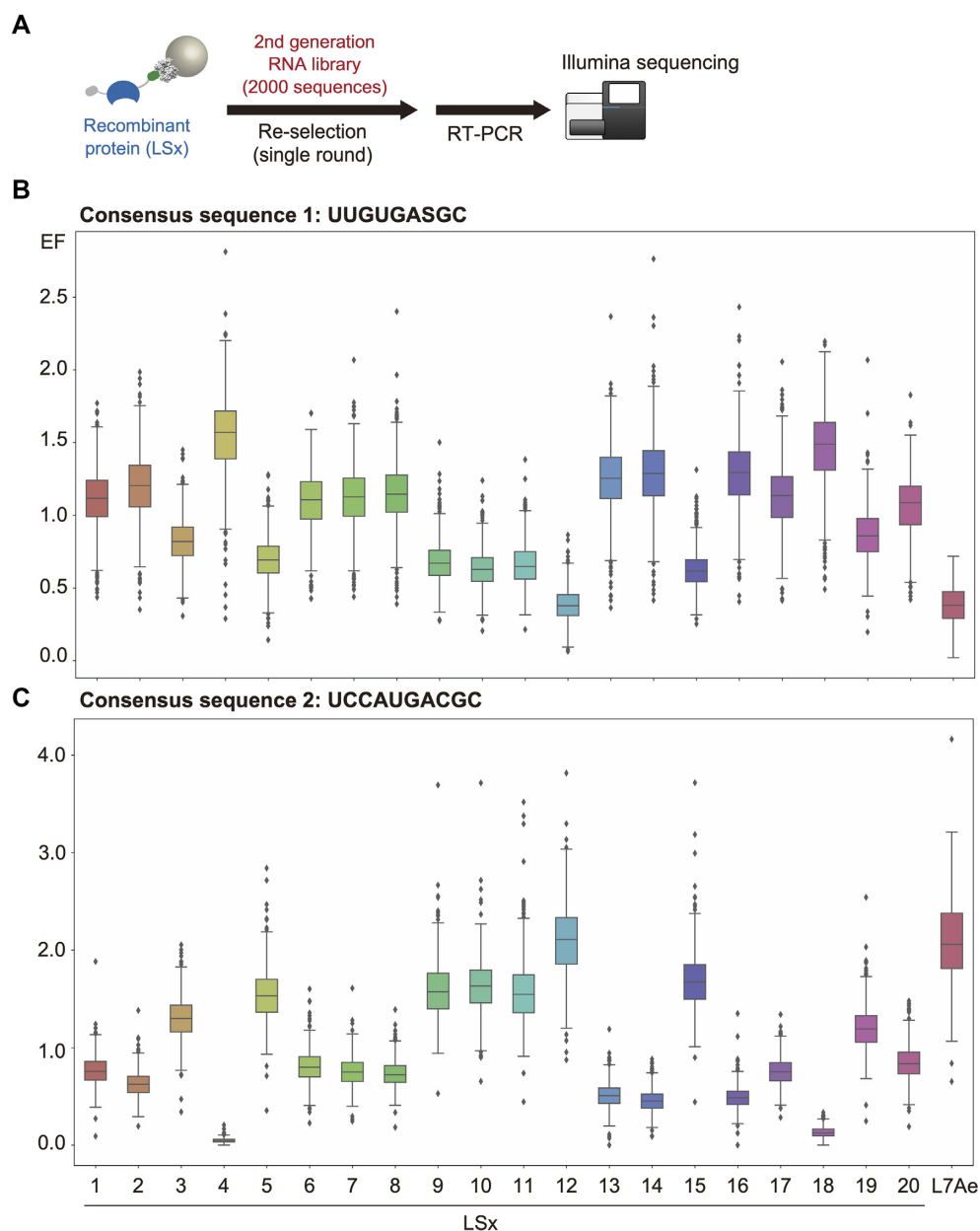
Figure 4. Re-selection of the second-generation RNA library against individual L7Ae mutants (LS1-LS20) and L7Ae. (**A**) Schematic illustration of the re-selection experiment. A focused RNA library containing the 1991 most abundant RNAs after PD-SELEX and 9 control RNAs was selected against individual L7Ae or its mutant immobilized on magnetic beads. (**B**, **C**) Box plots representing enrichment factors (EFs) of a subset of the RNAs that contains consensus sequence 1 (UUGUGASGC, S = C or G) (**B**) or consensus sequence 2 (UCCAUGACGC) (**C**). EF is defined as the abundance of an RNA after selection divided by the abundance before selection.

of consensus sequence 2, the RNAs were significantly enriched by LS5, LS9, LS10, LS11, LS12, LS15 and L7Ae (Figure 4C), and among them, LS5, LS9, LS10, LS11 and LS15 share the same amino acids at two positions (R89 and R90) (Figure 3C) while the 91st residue of LS12 is the same as that of L7Ae (P91).

Further inspection revealed that LS4 has low EF values (<0.5) for consensus sequence 2, and LS12 and L7Ae have low EF values (<0.5) for consensus sequence 1 (Figure 4B and C). This suggests that LS4 may selectively bind RNA sequences containing consensus sequence 1, and LS12 and L7Ae may selectively bind RNA sequences containing con-

sensus sequence 2, therefore could function as orthogonal RNA–RBP pairs.

Binding affinities of LS4 and LS12 with selected RNAs

We chose RNA sequences that exhibited the highest EF value for LS4 or LS12 (EF of LS4-1 = 2.8; EF of LS12-1 = 3.8) and measured their binding affinities using SPR (Figure 5). RNAs were captured on a 2′-OMe RNA-immobilized sensor chip, and varying concentrations of proteins were injected to determine the binding and dissociation kinetics. LS4-1 RNA bound LS4 with a K_D of 7.36×10^{-11} M,
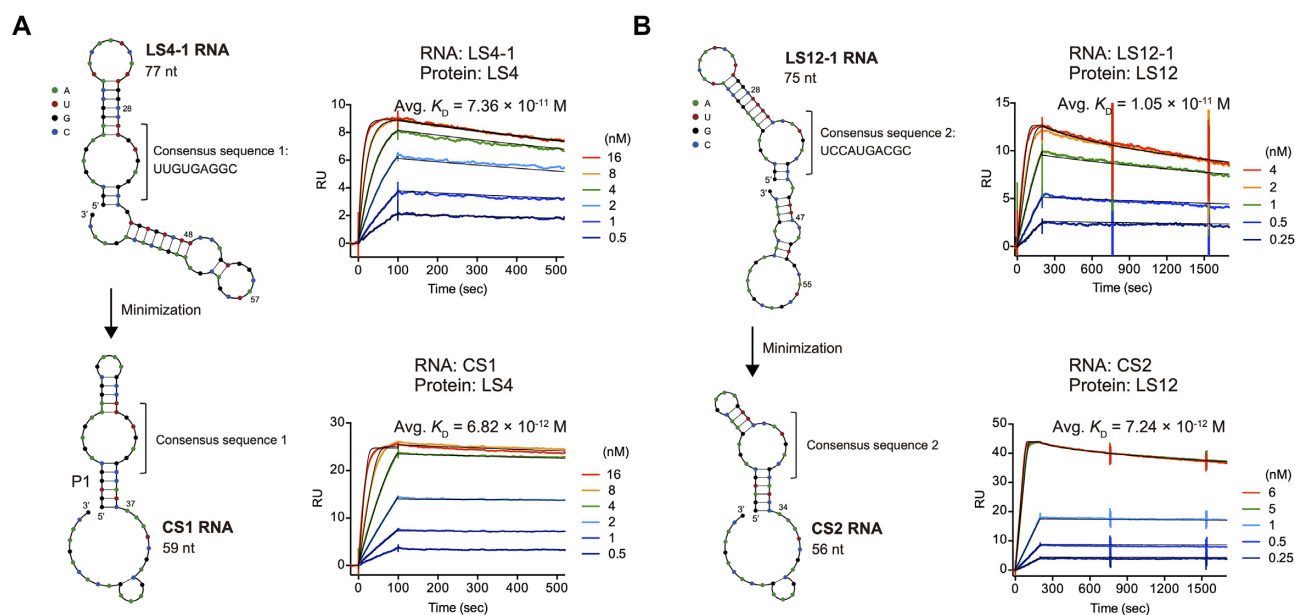
Figure 5. Characterization of binding affinity between the selected RNAs and RBPs. (**A**) SPR sensorgrams of LS4-1 RNA and LS4 protein (top), and the minimized CS1 RNA and LS4 protein (bottom). (**B**) SPR sensorgrams of LS12-1 RNA and LS12 protein (top), and the minimized CS2 RNA and LS12 protein (bottom). Observed kinetic values and dissociation constants are summarized in Table 2. It should be noted that 3′ end of the minimized RNAs (CS1: bases 37–59, CS2: bases 34–56) were used for immobilization on the SPR chip.

and LS12-1 RNA bound LS12 with a K_D of 1.05×10^{-11} M (Figure 5 and Table 2). Secondary structure prediction by NUPACK (60) suggested that the consensus sequence 1 in LS4-1 RNA likely formed an internal loop as depicted in Figure 5A. We hypothesized that this internal loop structure plays a critical role for binding to LS4 and designed CS1 RNA (Figure 5A), a minimized version of LS4-1 RNA. Strikingly, CS1 RNA bound to LS4 with a dissociation constant of 6.82×10^{-12} M, approximately 10-fold improvement relative to LS4-1 RNA presumably due to the stabilization of the P1 stem. Similarly, the consensus sequence 2 was predicted to form an asymmetric internal loop in LS12-1 RNA (Figure 5B) which comprises a putative kink-turn motif (61). Accordingly, LS12-1 RNA was minimized to CS2 RNA (Figure 5B) whose K_D for LS12 binding was measured to be 7.24×10^{-12} M which is comparable to the that of the parental RNA. Electrophoretic mobility shift assay (EMSA) of CS1 RNA–LS4 and CS2 RNA–LS12 using 0.5 nM Cy5-labeled RNA showed binding profiles characteristic of titration regime (62) which indicates $K_D \ll 5 \times 10^{-10}$ M (Supplementary Figure S9).

Selectivity and mutational analysis of CS1 and CS2 RNAs

Binding selectivity of CS1 RNA was evaluated by SPR. The dissociation constant of CS1 RNA and LS12 was 2.75×10^{-8} M, and that of CS1 RNA and L7Ae was 1.49×10^{-9} M (Supplementary Figure S10 and Supplementary Table S4). Therefore, CS1 RNA selectively binds LS4 over LS12 and L7Ae by 4030-fold and 218-fold, respectively (Figure 6A). The measured binding affinities of CS1 RNA with the RBPs are consistent with the observed enrichment factors in the re-selection experiment (Figure 4B).

It is known that G●A mismatches in the core of Kt RNAs form *trans* G (sugar)●A (Hoogsteen) type base pairs (55). We found four putative G●A base pairs (A7/G30, G8/A29, A9/G28, and A11/G26) that may exist in the symmetric internal loop of CS1 RNA (Figure 6A). To examine the contribution of the putative G●A base pairs to the affinity with LS4, we systematically mutated each G●A pair to a U.U mismatch (M1: A7U/G30U, M2: G8U/A29U, M3: A9U/G28U, M4: A11U/G26U). Interestingly, as the mutation gets closer to the P2 stem, the binding affinity clearly decreased (CS1-M1 and CS1-M2) and no binding was observed with CS1-M3 and CS1-M4. This suggests that G●A pairs close to the P2 stem (i.e., A9/G28 and A11/G26) play an important role in binding. Among these four mutants, CS1-M2 retained nanomolar affinity ($K_D = 2.21 \times 10^{-9}$ M) for LS4 but its affinity for LS12 and L7Ae was not measurable ($K_D > \sim 1$ μM). This prompted us to characterize additional mutants at these positions (M2a: G8U, M2b: G8A, M2c: A29U, M2d: G8A/A29G, M2e: G8A/A29U). CS1-M2b RNA exhibited respectable binding selectivity for LS4 (1273-fold over LS12) with a lower K_D (8.65×10^{-11} M). The possibility of tuning the binding affinity while preserving selectivity by mutating CS1 RNA may be advantageous for adapting the RNA–RBP pair for specific applications.

We next examined the binding affinities of CS2 RNA with LS4 and L7Ae. CS2 RNA bound LS4 at 1.14×10^{-7} M K_D which is 15800-fold greater than that for LS12 (7.24×10^{-12} M). On the other hand, the K_D value of CS2 RNA and L7Ae was 5.71×10^{-12} M (Supplementary Figure S10 and Supplementary Table S4) which was comparable to that of CS2 RNA and LS12 (Figure 5B). Therefore, although CS2 RNA cannot discriminate between LS12 and L7Ae, CS2 RNA selectively binds L7Ae 20 000-fold more strongly

Table 2. Binding characteristics of the RNA–RBP pairs from PD-SELEX

RNA	Protein	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})	K_D (M)
LS4-1	LS4	$(6.94 \pm 0.60) \times 10^6$	$(5.06 \pm 0.56) \times 10^{-4}$	$(7.36 \pm 1.43) \times 10^{-11}$
CS1		$(2.00 \pm 0.07) \times 10^7$	$(1.36 \pm 0.15) \times 10^{-4}$	$(6.82 \pm 0.98) \times 10^{-12}$
LS12-1	LS12	$(3.49 \pm 0.65) \times 10^7$	$(3.53 \pm 1.22) \times 10^{-4}$	$(1.05 \pm 0.46) \times 10^{-11}$
CS2		$(4.96 \pm 3.48) \times 10^7$	$(3.26 \pm 1.58) \times 10^{-4}$	$(7.24 \pm 2.29) \times 10^{-12}$

k_{on}, k_{off}, K_D values are means and standard deviations of at least two independent experiments.
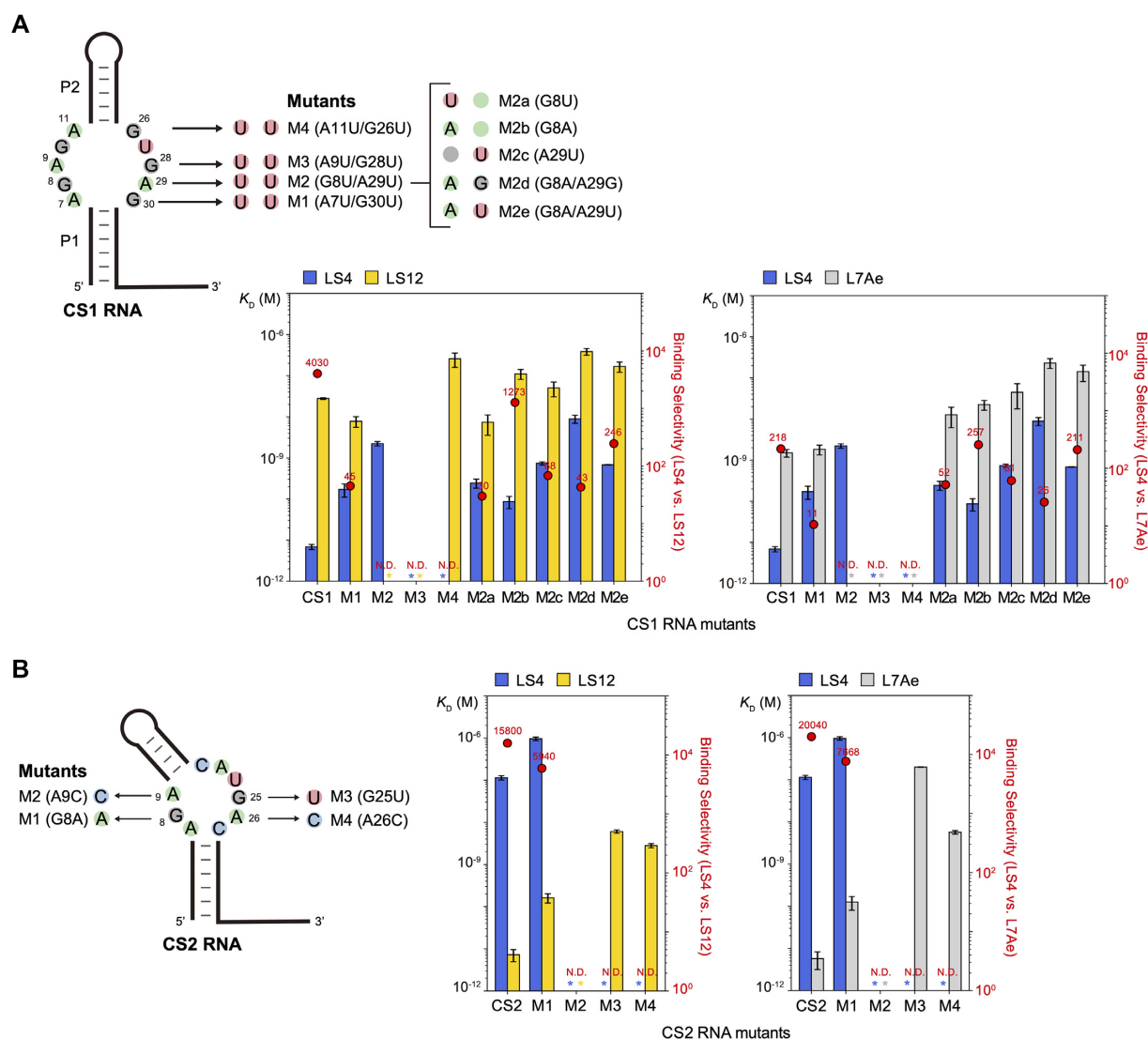
Figure 6. Binding selectivities of LS4, LS12, and L7Ae for (**A**) CS1 RNA and its mutants, and (**B**) CS2 RNA and its mutants. The bar graphs represent means and standard deviations of the K_D values derived from at least two independent experiments. Red dots represent binding selectivity (ratio of the K_D values) of the RNA for the two proteins. Asterisks indicate either no binding or weak binding with undetermined K_D value based on SPR. Binding selectivity could not be determined for these RNAs (N.D.). SPR sensorgrams and binding parameters are shown in Supplementary Figure S10 and Supplementary Table S4.

compared to LS4. These binding affinities were also consistent with the RNA binding trends in the re-selection experiment (Figure 4C).

We introduced several single-base mutations in the putative G●A pairs (G8/A26 and A9/G25) in the asymmetric internal loop of CS2 RNA (M1: G8A, M2: A9C, M3: G25U, M4: A26C) (Figure 6B). LS12 and L7Ae both showed sim-

ilar degree of increase in K_D except for M3. Impact of the M3 mutation on binding affinity was more pronounced for L7Ae ($K_D = 2.00 \times 10^{-7}$ M) than for LS12 ($K_D = 6.02 \times 10^{-9}$ M) (Supplementary Figure S10 and Supplementary Table S4). This observation may be explained by the E34S substitution in LS12 (Figure 3C). It was shown that E34 in L7Ae contacts guanine in the G●A pair adjacent to the

bulge in Kt-7 RNA ([54]) which corresponds to G25 in CS2 that was mutated to a uridine in CS2-M3 RNA.

The affinities of LS4 and LS12 for the box C/D Kt RNA were significantly impaired ($K_D = 5.99 \times 10^{-7}$ M and 1.45×10^{-8} M, respectively) compared to the box C/D Kt – L7Ae interaction ($K_D = 2.13 \times 10^{-11}$ M). (Supplementary Figure S10 and Supplementary Table S4). Therefore, these synthetic RBPs exhibit high selectivity for the cognate synthetic RNAs over the native L7Ae binding RNA motif. Overall, LS4 and CS1 RNA, as well as LS12 and CS2 RNA, were shown to be highly selective and orthogonal RNA–RBP pairs with picomolar affinities (Figure [6]A and B).

DISCUSSION

We used SPR to measure the L7Ae binding to box C/D Kt RNA resulting in a K_D value of 2.13×10^{-11} M (Figure [2]B and Table [1]) which is consistent with the result by the Lilley group who reported 10 pM K_D for another Kt motif (Kt-7) based on Förster resonance energy transfer (FRET) measurement using a stopped-flow mixer ([55]). However, Inoue and coworkers have reported significantly higher K_D values for box C/D RNA measured by SPR (between 1.6 and 2.5 nM) ([13,32,42]). We attribute the discrepancy to several possible factors. In two of the three studies ([13,42]), the researchers immobilized L7Ae protein on an SPR chip via EDC/NHS chemistry, and RNA was injected as the analyte. Immobilization of protein via primary amine groups results in heterogenous orientations of the immobilized protein some of which may negatively affect binding. In fact, some lysine residues in L7Ae are involved in RNA binding ([54,63]). Moreover, the negatively charged carboxymethylated dextran-functionalized sensor chip surface can reduce the local RNA analyte concentration due to electrostatic repulsion ([64]). The RNA was immobilized and L7Ae was injected as the analyte in another study that reported 1.6 nM K_D ([32]). While this experimental setup was similar to that of our study, we speculate that the recombinant L7Ae preparation may have caused the different observations. Turner and Lilley have pointed out that RNA contamination due to cellular RNAs copurified with L7Ae can affect affinity measurement, and they included a heparin chromatography step to minimize such interference ([55]). In our case, we treated the *E. coli* crude extracts containing recombinant proteins with ribonuclease A and deoxyribonuclease I to reduce nucleic acid contamination, and the proteins were affinity purified using two separate tags.

Hara *et al.* have reported their efforts to engineer L7Ae as a RBP scaffold to recognize non-natural RNA ([13]). They started with L7KK, a double mutant of L7Ae (K37A/K79A), which had a ~51-fold reduced affinity for box C/D RNA and performed SELEX to isolate RNA motifs that bind to L7KK. While they discovered a compact aptamer H23 that recognizes L7KK, H23 also bound L7Ae with a comparable affinity. Consequently, they were not able to discover an orthogonal RNA–RBP pair, highlighting the challenges in engineering orthogonal RNA–RBP pairs by a single-library strategy. In PD-SELEX, it was anticipated that the selected RNA–RBP pairs would be orthogonal to the parental pair (box C/D Kt and L7Ae). Indeed, both LS4 and LS12 showed reduced affinity to box C/D Kt. However,

while CS1 RNA showed 218-fold selectivity for LS4 over L7Ae, CS2 RNA bound LS12 and L7Ae with comparable affinity. In retrospect, inclusion of a negative selection step could have improved the chances of discovering other orthogonal RNA–RBP pairs. This could have been achieved by including excess L7Ae protein as a competitor during the RNA-phage binding step.

Since the first application of L7Ae and box C/D Kt RNA for translational regulation in mammalian cells by Saito *et al.* ([32]), the RNA–RBP pair has been a popular module for applications in synthetic biology. However, Randau *et al.* reported that overproduction of L7Ae causes cytotoxicity in *E. coli* suggesting that L7Ae binds to endogenous RNAs in *E. coli* ([36]). It has also been reported that overexpression of L7Ae in mammalian cells causes slower growth and/or cell death ([41,65]). This is probably because archaeal L7Ae can bind to broad Kt structures which are conserved in many species from archaea to humans ([53,57]). The evolved LS4 protein strongly binds CS1 RNA with a K_D value of 6.82×10^{-12} M (Figure [5] and Table [2]) while its affinity for the box C/D Kt and another putative Kt RNA (CS2 RNA) is significantly compromised ($K_D > 100$ nM) (Supplementary Figure S10 and Supplementary Table S4). Moreover, we confirmed that without nuclease treatment, *E. coli* expressed L7Ae and LS12 are copurified with abundant cellular DNA and RNA in cell lysate, consistent with the observations by Turner and Lilley ([55]), suggesting nonspecific interactions with nucleic acids. However, LS4 was copurified with a markedly lower amount of nucleic acids under the same conditions (Supplementary Figure S11). While it remains to be experimentally tested, LS4 may exhibit reduced cytotoxicity for synthetic biology applications compared to L7Ae.

To date, in vitro and *E. coli* or yeast cell-based in vivo library-vs-library selections have been reported to identify protein-protein ([21–25]), protein-peptide ([26]), peptide-peptide ([27,28]), and small molecule-protein ([29,30]), small molecule-RNA ([31]) pairs. This study is the first library-vs-library selection for RNA-protein interactions. In cell-based library-vs-library selections ([22,25–28]), two different genes are transformed into the cell, and binding between the two components results in expression of a drug resistance gene or a reporter gene which is used to enrich the interacting partners. However, combinations of the binding partners that can be explored using cell-based methods are limited by the transformation efficiency, or ~10^9 pairs ([66]). The sequence space that can be explored by in vitro library-vs-library selections significantly exceeds those that can be probed by cell-based experiments. For example, Lerner and coworkers carried out library-vs-library selection using yeast-displayed single-chain antibody fragment library with a diversity of 2×10^7 and phage-displayed antigen library with a diversity of ~10^7, and successfully enriched an antibody-antigen pair from >10^{14} possible combinations ([21]). Our PD-SELEX experiment started with up to 4.4×10^{20} possible RNA–RBP combinations, significantly higher than previous library-vs-library selection experiments. However, a caveat with highly complex library-vs-library selection is the low concentrations of the individual binding partners in solution. It can be imagined that if the binding affinity between an RNA motif and a pro-

tein mutant is weak, such a pair may not be able to survive the first round of selection due to the low concentration of the bound pair. Here, the relevant concentration is not necessarily the absolute concentration of a single RNA or protein sequence but could be the total effective concentration of multiple sequences that contain a consensus motif. Our PD-SELEX yielded RNA–RBP pairs with ∼7 pM K_D. This strong affinity combined with the relatively small RNA motifs (CS1 and CS2) that are represented more frequently in the randomized RNA library (relative to larger and more complex RNA motifs) may have contributed to the successful enrichment the RNA–RBP pairs. In principle, RNA–RBP pairs with lower affinity (e.g. nanomolar K_D) should also be selectable by PD-SELEX. In Bowley *et al.*, an antigen-antibody pair with a K_D of 50 nM was selected from phage-displayed antibody and yeast-displayed antigen libraries (21). However, it may be necessary to adjust the library complexity and/or the selection conditions to optimize the enrichment efficiency for the expected or desired affinity range. This can be achieved by performing a mock selection experiment as shown in Figure 2.

In conclusion, we designed and executed PD-SELEX to select orthogonal RNA–RBP pairs from RNA and phage-displayed RBP libraries. L7Ae mutants LS4 and LS12 and their cognate RNA binding motifs (CS1 and CS2 RNAs) show low picomolar affinities which are among the strongest RNA-protein interactions observed in natural or synthetic systems (67–70). We expect PD-SELEX to be a useful and efficient strategy for engineering and analysis of RNA-protein interactions.

DATA AVAILABILITY

All data and materials are available upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Johansson,H.E., Dertinger,D., LeCuyer,K.A., Behlen,L.S., Greef,C.H. and Uhlenbeck,O.C. (1998) A thermodynamic analysis of the sequence-specific binding of RNA by bacteriophage MS2 coat protein. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 9244–9249.
2. Grunwald,D. and Singer,R.H. (2010) In vivo imaging of labelled endogenous beta-actin mRNA during nucleocytoplasmic transport. *Nature*, **467**, 604–607.
3. Tutucci,E., Vera,M., Biswas,J., Garcia,J., Parker,R. and Singer,R.H. (2018) An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat. Methods*, **15**, 81–89.
4. Han,S., Zhao,B.S., Myers,S.A., Carr,S.A., He,C. and Ting,A.Y. (2020) RNA-protein interaction mapping via MS2- or Cas13-based APEX targeting. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 22068–22079.
5. Said,N., Rieder,R., Hurwitz,R., Deckert,J., Urlaub,H. and Vogel,J. (2009) In vivo expression and purification of aptamer-tagged small RNA regulators. *Nucleic Acids Res.*, **37**, e133.
6. Konermann,S., Brigham,M.D., Trevino,A.E., Joung,J., Abudayyeh,O.O., Barcena,C., Hsu,P.D., Habib,N., Gootenberg,J.S., Nishimasu,H. et al. (2015) Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature*, **517**, 583–588.
7. Shechner,D.M., Hacisuleyman,E., Younger,S.T. and Rinn,J.L. (2015) Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods*, **12**, 664–670.
8. Zalatan,J.G., Lee,M.E., Almeida,R., Gilbert,L.A., Whitehead,E.H., La Russa,M., Tsai,J.C., Weissman,J.S., Dueber,J.E., Qi,L.S. et al. (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, **160**, 339–350.
9. Auslander,S., Auslander,D., Muller,M., Wieland,M. and Fussenegger,M. (2012) Programmable single-cell mammalian biocomputers. *Nature*, **487**, 123–127.
10. Kennedy,A.B., Vowles,J.V., d'Espaux,L. and Smolke,C.D. (2014) Protein-responsive ribozyme switches in eukaryotic cells. *Nucleic Acids Res.*, **42**, 12306–12321.
11. Wroblewska,L., Kitada,T., Endo,K., Siciliano,V., Stillo,B., Saito,H. and Weiss,R. (2015) Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nat. Biotechnol.*, **33**, 839–841.
12. Kawasaki,S., Ono,H., Hirosawa,M. and Saito,H. (2020) RNA and protein-based nanodevices for mammalian post-transcriptional circuits. *Curr. Opin. Biotechnol.*, **63**, 99–110.
13. Hara,T., Saito,H. and Inoue,T. (2013) Directed evolution of a synthetic RNA-protein module to create a new translational switch. *Chem. Commun. (Camb.)*, **49**, 3833–3835.
14. Blakeley,B.D. and McNaughton,B.R. (2014) Synthetic RNA recognition motifs that selectively recognize HIV-1 trans-activation response element hairpin RNA. *ACS Chem. Biol.*, **9**, 1320–1329.
15. Crawford,D.W., Blakeley,B.D., Chen,P.H., Sherpa,C., Le Grice,S.F., Laird-Offringa,I.A. and McNaughton,B.R. (2016) An evolved RNA recognition motif that suppresses HIV-1 Tat/TAR-dependent transcription. *ACS Chem. Biol.*, **11**, 2206–2215.
16. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818–822.
17. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
18. Smith,G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, **228**, 1315–1317.
19. McCafferty,J., Griffiths,A.D., Winter,G. and Chiswell,D.J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature*, **348**, 552–554.
20. Danner,S. and Belasco,J.G. (2001) T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 12954–12959.
21. Bowley,D.R., Jones,T.M., Burton,D.R. and Lerner,R.A. (2009) Libraries against libraries for combinatorial selection of replicating antigen-antibody pairs. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 1380–1385.
22. Yu,H., Tardivo,L., Tam,S., Weiner,E., Gebreab,F., Fan,C., Svrzikapa,N., Hirozane-Kishikawa,T., Rietman,E., Yang,X. et al. (2011) Next-generation sequencing to generate interactome datasets. *Nat. Methods*, **8**, 478–480.
23. Gu,L., Li,C., Aach,J., Hill,D.E., Vidal,M. and Church,G.M. (2014) Multiplex single-molecule interaction profiling of DNA-barcoded proteins. *Nature*, **515**, 554–557.
24. Younger,D., Berger,S., Baker,D. and Klavins,E. (2017) High-throughput characterization of protein-protein interactions by

reprogramming yeast mating. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 12166–12171.

25. McClune,C.J., Alvarez-Buylla,A., Voigt,C.A. and Laub,M.T. (2019) Engineering orthogonal signalling pathways reveals the sparse occupancy of sequence space. *Nature*, **574**, 702–706.

26. Siau,J.W., Nonis,S., Chee,S., Koh,L.Q., Ferrer,F.J., Brown,C.J. and Ghadessy,F.J. (2020) Directed co-evolution of interacting protein-peptide pairs by compartmentalized two-hybrid replication (C2HR). *Nucleic Acids Res.*, **48**, e128.

27. Pelletier,J.N., Arndt,K.M., Pluckthun,A. and Michnick,S.W. (1999) An in vivo library-versus-library selection of optimized protein-protein interactions. *Nat. Biotechnol.*, **17**, 683–690.

28. Arndt,K.M., Pelletier,J.N., Muller,K.M., Alber,T., Michnick,S.W. and Pluckthun,A. (2000) A heterodimeric coiled-coil peptide pair selected in vivo from a designed library-versus-library ensemble. *J. Mol. Biol.*, **295**, 627–639.

29. McGregor,L.M., Jain,T. and Liu,D.R. (2014) Identification of ligand-target pairs from combined libraries of small molecules and unpurified protein targets in cell lysates. *J. Am. Chem. Soc.*, **136**, 3264–3270.

30. Chan,A.I., McGregor,L.M., Jain,T. and Liu,D.R. (2017) Discovery of a covalent kinase inhibitor from a DNA-encoded small-molecule library x protein library selection. *J. Am. Chem. Soc.*, **139**, 10192–10195.

31. Tran,T. and Disney,M.D. (2012) Identifying the preferred RNA motifs and chemotypes that interact by probing millions of combinations. *Nat. Commun.*, **3**, 1125.

32. Saito,H., Kobayashi,T., Hara,T., Fujita,Y., Hayashi,K., Furushima,R. and Inoue,T. (2010) Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nat. Chem. Biol.*, **6**, 71–78.

33. Zhitnyuk,Y., Gee,P., Lung,M.S.Y., Sasakawa,N., Xu,H., Saito,H. and Hotta,A. (2018) Efficient mRNA delivery system utilizing chimeric VSVG-L7Ae virus-like particles. *Biochem. Biophys. Res. Commun.*, **505**, 1097–1102.

34. Kojima,R., Bojar,D., Rizzi,G., Hamri,G.C., El-Baba,M.D., Saxena,P., Auslander,S., Tan,K.R. and Fussenegger,M. (2018) Designer exosomes produced by implanted cells intracerebrally deliver therapeutic cargo for Parkinson's disease treatment. *Nat. Commun.*, **9**, 1305.

35. Kopniczky,M.B., Canavan,C., McClymont,D.W., Crone,M.A., Suckling,L., Goetzmann,B., Siciliano,V., MacDonald,J.T., Jensen,K. and Freemont,P.S. (2020) Cell-free protein synthesis as a prototyping platform for mammalian synthetic biology. *ACS Synth. Biol.*, **9**, 144–156.

36. Daume,M., Uhl,M., Backofen,R. and Randau,L. (2017) RIP-Seq suggests translational regulation by L7Ae in archaea. *MBio*, **8**, e00730-17.

37. Saito,H., Fujita,Y., Kashida,S., Hayashi,K. and Inoue,T. (2011) Synthetic human cell fate regulation by protein-driven RNA switches. *Nat. Commun.*, **2**, 160.

38. Endo,K., Hayashi,K., Inoue,T. and Saito,H. (2013) A versatile cis-acting inverter module for synthetic translational switches. *Nat. Commun.*, **4**, 2393.

39. Auslander,S., Stucheli,P., Rehm,C., Auslander,D., Hartig,J.S. and Fussenegger,M. (2014) A general design strategy for protein-responsive riboswitches in mammalian cells. *Nat. Methods*, **11**, 1154–1160.

40. Cella,F., Wroblewska,L., Weiss,R. and Siciliano,V. (2018) Engineering protein-protein devices for multilayered regulation of mRNA translation using orthogonal proteases in mammalian cells. *Nat. Commun.*, **9**, 4392.

41. Wagner,T.E., Becraft,J.R., Bodner,K., Teague,B., Zhang,X., Woo,A., Porter,E., Alburquerque,B., Dobosh,B., Andries,O. *et al.* (2018) Small-molecule-based regulation of RNA-delivered circuits in mammalian cells. *Nat. Chem. Biol.*, **14**, 1043–1050.

42. Stapleton,J.A., Endo,K., Fujita,Y., Hayashi,K., Takinoue,M., Saito,H. and Inoue,T. (2012) Feedback control of protein expression in mammalian cells by tunable synthetic translational inhibition. *ACS Synth. Biol.*, **1**, 83–88.

43. Shen,C.C., Hsu,M.N., Chang,C.W., Lin,M.W., Hwu,J.R., Tu,Y. and Hu,Y.C. (2019) Synthetic switch to minimize CRISPR off-target effects by self-restricting Cas9 transcription and translation. *Nucleic Acids Res.*, **47**, e13.

44. Endo,K., Stapleton,J.A., Hayashi,K., Saito,H. and Inoue,T. (2013) Quantitative and simultaneous translational control of distinct mammalian mRNAs. *Nucleic Acids Res.*, **41**, e135.

45. Lin,M.W., Tseng,Y.W., Shen,C.C., Hsu,M.N., Hwu,J.R., Chang,C.W., Yeh,C.J., Chou,M.Y., Wu,J.C. and Hu,Y.C. (2018) Synthetic switch-based baculovirus for transgene expression control and selective killing of hepatocellular carcinoma cells. *Nucleic Acids Res.*, **46**, e93.

46. Milligan,J.F. and Uhlenbeck,O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol.*, **180**, 51–62.

47. Kibbe,W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.*, **35**, W43–W46.

48. Ho,S.N., Hunt,H.D., Horton,R.M., Pullen,J.K. and Pease,L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.

49. Dwidar,M., Seike,Y., Kobori,S., Whitaker,C., Matsuura,T. and Yokobayashi,Y. (2019) Programmable artificial cells using histamine-responsive synthetic riboswitch. *J. Am. Chem. Soc.*, **141**, 11103–11114.

50. Schmidt,T.G. and Skerra,A. (2007) The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nat. Protoc.*, **2**, 1528–1535.

51. Moayed,F., Mashaghi,A. and Tans,S.J. (2013) A polypeptide-DNA hybrid with selective linking capability applied to single molecule nano-mechanical measurements using optical tweezers. *PLoS One*, **8**, e54440.

52. Baumann,F., Bauer,M.S., Milles,L.F., Alexandrovich,A., Gaub,H.E. and Pippig,D.A. (2016) Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nat. Nanotechnol.*, **11**, 89–94.

53. Gagnon,K.T., Zhang,X., Qu,G., Biswas,S., Suryadi,J., Brown,B.A. and Maxwell,E.S. (2010) Signature amino acids enable the archaeal L7Ae box C/D RNP core protein to recognize and bind the K-loop RNA motif. *RNA*, **16**, 79–90.

54. Huang,L. and Lilley,D.M. (2013) The molecular recognition of kink-turn structure by the L7Ae class of proteins. *RNA*, **19**, 1703–1710.

55. Turner,B. and Lilley,D.M. (2008) The importance of G.A hydrogen bonding in the metal ion- and protein-induced folding of a kink turn RNA. *J. Mol. Biol.*, **381**, 431–442.

56. Chizzolini,F., Passalacqua,L.F.M., Oumais,M., Dingilian,A.I., Szostak,J.W. and Luptak,A. (2020) Large phenotypic enhancement of structured random RNA pools. *J. Am. Chem. Soc.*, **142**, 1941–1951.

57. Lilley,D.M. (2014) The K-turn motif in riboswitches and other RNA species. *Biochim. Biophys. Acta*, **1839**, 995–1004.

58. Krumpe,L.R., Schumacher,K.M., McMahon,J.B., Makowski,L. and Mori,T. (2007) Trinucleotide cassettes increase diversity of T7 phage-displayed peptide library. *BMC Biotechnol.*, **7**, 65.

59. Hatanaka,T., Matsugami,A., Nonaka,T., Takagi,H., Hayashi,F., Tani,T. and Ishida,N. (2017) Rationally designed mineralization for selective recovery of the rare earth elements. *Nat. Commun.*, **8**, 15670.

60. Zadeh,J.N., Steenberg,C.D., Bois,J.S., Wolfe,B.R., Pierce,M.B., Khan,A.R., Dirks,R.M. and Pierce,N.A. (2011) NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.*, **32**, 170–173.

61. Liu,J. and Lilley,D.M. (2007) The role of specific 2′-hydroxyl groups in the stabilization of the folded conformation of kink-turn RNA. *RNA*, **13**, 200–210.

62. Jarmoskaite,I., AlSadhan,I., Vaidyanathan,P.P. and Herschlag,D. (2020) How to measure and evaluate binding affinities. *Elife*, **9**, e57264.

63. Hamma,T. and Ferre-D'Amare,A.R. (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 A resolution. *Structure*, **12**, 893–903.

64. Chang,A.L., McKeague,M. and Smolke,C.D. (2014) Facile characterization of aptamer kinetic and equilibrium binding properties using surface plasmon resonance. *Methods Enzymol.*, **549**, 451–466.

65. Ono,H., Kawasaki,S. and Saito,H. (2020) Orthogonal protein-responsive mRNA switches for mammalian synthetic biology. *ACS Synth. Biol.*, **9**, 169–174.

66. Li,M. (2000) Applications of display technology in protein analysis. *Nat. Biotechnol.*, **18**, 1251–1256.

67. Heber,S., Gaspar,I., Tants,J.N., Gunther,J., Moya,S.M.F., Janowski,R., Ephrussi,A., Sattler,M. and Niessing,D. (2019)

Staufen2-mediated RNA recognition and localization requires combinatorial action of multiple domains. *Nat. Commun.*, **10**, 1659.

68. Amano,R., Takada,K., Tanaka,Y., Nakamura,Y., Kawai,G., Kozu,T. and Sakamoto,T. (2016) Kinetic and thermodynamic analyses of interaction between a High-Affinity RNA aptamer and its target protein. *Biochemistry*, **55**, 6221–6229.

69. Yang,X., Li,H., Huang,Y. and Liu,S. (2013) The dataset for protein-RNA binding affinity. *Protein Sci.*, **22**, 1808–1811.

70. Katsamba,P.S., Park,S. and Laird-Offringa,I.A. (2002) Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods*, **26**, 95–104.

71. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

72. Matochko,W.L., Li,Cory, Tang,S. and Derda,R. (2014) Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Res.*, **42**, 1784–1798.

73. Juds,C., Schmidt,J., Weller,M.G., Lange,T., Beck,U., Conrad,T. and Borner,H.G. (2020) Combining phage display and next-generation sequencing for materials sciences: a case study on probing polypropylene surfaces. *J. Am. Chem. Soc.*, **142**, 10624–10628.