2624–2635 *Nucleic Acids Research, 2018, Vol. 46, No. 5* *Published online 4 January 2018*
doi: 10.1093/nar/gkx1292

Affinity maturation of a portable Fab–RNA module for chaperone-assisted RNA crystallography

Deepak Koirala[1,†], Sandip A. Shelke[1,†], Marcel Dupont[1], Stormy Ruiz[2], Saurja DasGupta[2], Lucas J. Bailey[1], Steven A. Benner[3] and Joseph A. Piccirilli[1,2,*]

[1]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA, [2]Department of Chemistry, The University of Chicago, Chicago, IL 60637, USA and [3]Foundation for Applied Molecular Evolution, Firebird Biomolecular Sciences LLC, 13709 Progress Boulevard, Alachua, FL 32615, USA

Received September 26, 2017; Revised December 13, 2017; Editorial Decision December 14, 2017; Accepted December 28, 2017

ABSTRACT

Antibody fragments such as Fabs possess properties that can enhance protein and RNA crystallization and therefore can facilitate macromolecular structure determination. In particular, Fab BL3–6 binds to an AAACA RNA pentaloop closed by a GC pair with ∼100 nM affinity. The Fab and hairpin have served as a portable module for RNA crystallization. The potential for general application make it desirable to adjust the properties of this crystallization module in a manner that facilitates its use for RNA structure determination, such as ease of purification, surface entropy or binding affinity. In this work, we used both *in vitro* RNA selection and phage display selection to alter the epitope and paratope sides of the binding interface, respectively, for improved binding affinity. We identified a 5′-GNGACCC-3′ consensus motif in the RNA and S97N mutation in complimentarity determining region L3 of the Fab that independently impart about an order of magnitude improvement in affinity, resulting from new hydrogen bonding interactions. Using a model RNA, these modifications facilitated crystallization under a wider range of conditions and improved diffraction. The improved features of the Fab–RNA module may facilitate its use as an affinity tag for RNA purification and imaging and as a chaperone for RNA crystallography.

INTRODUCTION

Recent progress toward the understanding of cellular transcriptomes has revealed that beyond their typical roles in the central dogma of Biology, RNAs serve a multitude of other roles at almost every level of cellular function (1–4). As RNAs frequently contain domains that fold into complex three-dimensional structures to enable cellular function, obtaining structural information is an important step toward understanding their biological roles. However, among the more than 100 000 X-ray structures deposited in the PDB (www.rcsb.org) (5), only about 1% contain standalone RNA structures, underscoring the difficulties and challenges associated with RNA crystallography. Recent progress in cryo-electron microscopy (6–8) complemented by X-ray crystallography and nuclear magnetic resonance has provided promising outcomes in determining structures of large ribonucleoprotein complexes such as ribosomes (9), spliceosomes (10,11) and group II intron RNPs (12), but the ability of cryo electron-microscopy to provide detailed structural insights into standalone RNA molecules remains to be established.

One of the strategies for RNA crystallization relies on the use of RNA binding proteins as crystallization chaperones, (13–15) as surface diversity provided by many amino acid types enables lattice contacts that facilitate crystallization. RNA molecules of interest can be constructed to contain a motif that binds to its cognate RNA binding protein and the complex is advanced to crystallization trials. These so-called crystallization modules include the U1A RNA binding protein (13) and Fab BL3–6 (16), which bind to their cognate hairpins, and the L7ae family of proteins that bind to the kink turn motifs (15,17,18). The modules have yielded promising results for RNA crystallization and structure determination (19).

The higher molecular weight and β-sheet rich architecture of Fabs offer potential advantages as crystallization chaperones relative to other proteins because they possess increased surface area for lattice contacts and tend to self-associate via exposed β-sheet edges. For example, U1A has a MW of ∼11 kDa, 21% β-sheet composition and 5223 Å^2 total solvent accessible surface area (SASA); L7ae has MW ∼13 kDa, 17% β-sheet and 6314 Å^2 SASA; and Fab BL3–6 has MW ∼50 kDa, 48% β-sheet and 38712 Å^2 SASA (www.ebi.ac.uk/pdbe/pisa/) (20). Additionally, deploying crystal-

*To whom correspondence should be addressed. Tel: +1 773 702 9312; Fax: +1 773 702 0271; Email: jpicciri@uchicago.edu
† These authors contributed equally to the paper as first authors.

lization chaperones with known structures allows for facile phasing by molecular replacement.

Our laboratory has used Fabs to crystallize and solve structures of a variety of RNA targets including the P4-P6 domain from group I intron (14), the class I ligase ribozyme (16), the Spinach green fluorescent RNA aptamer (21) and a 12-mer single-stranded RNA (22). In the case of Fab BL3–6, the protein binds to a stem-loop structure containing an AAACA pentaloop closed by a G-C base pair and retains affinity to the RNA motif when grafted into other RNAs. For Spinach RNA, we replaced the P2-stem loop with the pentaloop and crystallized the aptamer as a complex with Fab BL3–6 (21).

Given the generality and potentially widespread utility of crystallization modules, enhancing their effectiveness can have beneficial impact on RNA structure determination. We envision a multiphase process to improve these crystallization chaperones that includes affinity maturation, surface entropy reduction and elbow angle restriction. Here, we focus on the first phase and report two independent affinity maturation strategies for Fab BL3–6 and its cognate 7-nt RNA motif. First, we improved the epitope RNA motif by challenging Fab BL3–6 to recognize with high-affinity RNA sequences from a library of randomized RNA hairpin loops (∼262 144 sequences) in an *in vitro* selection process. We isolated several variants of RNA hairpins that bind to Fab BL3–6 with higher affinity than the wild-type (WT), GAAACAC hairpin. One of the selected hairpins, GAGACCC, was used to tag and crystallize the Spinach RNA aptamer in complex with Fab BL3–6, and we solved the structure of this Fab–RNA complex at 2.09 Å resolution. Furthermore, an independent selection using a 40-nt randomized library unconstrained by secondary structure revealed a GNGACCC consensus sequence for Fab BL3–6 binding, underscoring the specificity of the RNA binding antibody. In our second approach, we generated a library of Fabs (∼10^{10} variants) by rationally varying the composition of residues in complimentarity determining regions (CDRs) of Fab BL3–6 and subsequently challenged the Fab library in a phage display selection to recognize and bind the parent cognate GAAACAC motif. After multiple rounds of phage display selection, we isolated Fab clones that bind to WT, GAAACAC hairpin with about 10-fold higher affinity than the parent Fab BL3–6 binds. As affinity matured Fabs also bind to affinity matured hairpins with low-nanomolar affinities, we grafted both our parent and affinity matured RNA hairpins into three structured RNAs – Spinach RNA aptamer, class I ligase ribozyme and P4-P6 domain of group I intron ribozyme—and performed the crystallization trials in complex with both our parent and affinity matured Fabs. The high affinity modules crystallize as effectively as the parent module and in some cases also facilitate crystallization under other conditions, setting the stage for future surface and conformational engineering.

MATERIALS AND METHODS

RNA synthesis and purification

The dsDNA templates for transcription reactions were prepared by polymerase chain reaction (PCR) amplification of the ssDNA oligomer purchased from integrated DNA technologies (www.idtdna.com). The first two nucleotides of the reverse primer contained 2′-OMe modifications to reduce transcriptional heterogeneity at the 3′ end (23). RNA was prepared by *in vitro* transcription for 3 h at 37°C in buffer containing 40 mM Tris–HCl pH 7.9, 2 mM spermidine, 10 mM NaCl, 25 mM $MgCl_2$, 10 mM 1,4-dithiothreitol (DTT), 30 U/ml RNase inhibitor, 2.5 U/ml TIPPase, 4 mM of each nucleotide triphosphate (NTP), 30 pmol/ml of DNA template, 40 μg/ml homemade T7 RNA polymerase. Transcription reactions were quenched by adding 10 U/ml RNase free DNase I (Promega, www.promega.com) and incubating at 37°C for 30 min. After the Phenol/Chloroform/Isopropanol, pH 4.3 extraction, the RNA was purified by denaturing polyacrylamide gel electrophoresis. Corresponding RNA band was visualized by UV shadowing and excised from the gel. RNA was eluted overnight at 4°C in 10 mM Tris, pH 8.0, 2 mM ethylenediaminetetraacetic acid (EDTA), 300 mM NaCl buffer. The buffer in eluted RNA was exchanged three times by pure water using 10 kDa cutoff size exclusion column (Amicon). RNA was collected, aliquoted into small fractions and stored at −80°C until further use.

RNA oligonucleotides for biochemical assays were prepared in house by solid-phase synthesis on a 1-μmol scale using an Expedite Nucleic Acid Synthesis System (8900) by following standard RNA synthesis protocols. The oligonucleotides were released from solid support with 3:1 NH_4OH/EtOH at 55°C for 8 h, desilylated with 300 μl 6:3:4 N-methylpyrrolidinone/ triethylamine/triethylamine-3HF at 65°C for 2 h and precipitated by n-BuOH. The oligomers were further purified by dPAGE, collected in pure water and stored at −80°C until further use.

In vitro selection with RNA libraries against Fab BL3-6

The *in vitro* selection of Fab binding RNA was performed by nitrocellulose membrane based immobilization strategy following similar procedures described elsewhere (24). The nitrocellulose membrane filter (Millipore Corporation, www.emdmillipore.com) was presoaked before use in the selection buffer, 10 mM Tris, 200 mM NaCl, 2.5 mM $MgCl_2$, pH 7.5 for 30 min at room temperature. To reduce the non-specific retention of RNA molecules in the filter during the selection, the negative selection was performed every other round of selection by passing the RNA libraries through the membrane. About 50 μg (∼2 nmols) of the RNA for the N9-library or ∼60 μg (∼2 nmols) for the N40-library was refolded in the selection buffer. For this, the RNA in water was heated at 90°C for 2 min and then incubated at 50°C for 15 min in the selection buffer followed by the incubation at room temperature for 5 min and kept on ice for another 5 min. This refolded RNA was then passed through the membrane for negative selection and collected flow-through was used for subsequent steps. In the first round, ∼1 nmol (∼10^{15} molecules) of the refolded RNA library was incubated with 2 μM Fab BL3–6 at room temperature for 30 min and the Fab–RNA mixture was passed through the membrane to capture the Fab-bound RNA sequences. The membrane was washed twice with 1 ml of selection buffer each time to remove the unbound RNA sequences. The bound RNA sequences were eluted from the membrane

in reverse transcription buffer (40 mM Tris–HCl, pH 7.5, 3 mM $MgCl_2$ and 75 mM KCl) supplemented with the 1 μM of the reverse primer to facilitate the elution. It was incubated at room temperature for 30 min with vigorous shaking, at 50°C for 30 min and finally at room temperature with vigorous shaking for 30 min. The RNA solution was then filtered using 0.2-μ filter unit (Millipore Corporation) to remove debris from the membrane and reverse transcribed to cDNA using SuperScript III (ThermoFisher Scientific) following the manufacturer's protocol. The cDNA was directly used for subsequent PCR amplification to generate dsDNA template library for producing the RNA pool for next round of selection.

The subsequent rounds of the selection were performed similarly except that the Fab concentration was reduced by 10 times and the number of washing steps by 1 ml of the selection buffer in each round of selection. The progress of the selection was monitored by running native polyacrylamide gel electrophoresis in each round. The output RNA library after five rounds of selection was subcloned into pUC19 vector between EcoRI and XbaI sites and sequenced. For cloning and sequencing purposes, the reverse primer used for reverse transcription was 5′- GCA CTCTAGAct ctt cgG CTG TTG TGA GCC TGT CGAA, where underlined sequence and lowercase sequence are XbaI and EarI restriction sites, respectively. The cDNA product in this case was PCR amplified by using the same reverse primer and a forward primer: 5′-CAG TGAATTCGG ACG TAA TAC GAC TCA CTA TAGG, where underlined sequence is EcoRI restriction site. This PCR product was double digested by EcoRI and XbaI enzymes to subclone into the pUC19 vector. The sequencing was done at core sequencing facility of the University of Chicago. RNA sequences were analyzed using WebLogo (http://weblogo.berkeley.edu/logo.cgi) (25), MEME-suite (http://meme-suite.org/tools/meme) (26) and Clustal omega sequence alignment (http://www.ebi.ac.uk/Tools/msa/clustalo/) (27). The individual hairpin clones were either prepared by *in vitro* transcription or by using in-house RNA synthesizer and binding affinity of each clone was characterized by filter-binding and hydroxyl-radical foot-printing assays.

Fab–RNA binding affinity measurements

The binding constants of selected RNA clones and related mutants were determined by nitrocellulose filter binding assay as reported previously (21). Briefly, ∼20 pmol of RNA was 5′-^{32}P radiolabeled and purified by denaturing polyacrylamide gel electrophoresis. A constant amount of radiolabeled RNA was incubated at 50°C for 10 min in an appropriate buffer (for hairpins: 10 mM Tris–HCl pH 7.4 containing 200 mM NaCl and 2.5 mM $MgCl_2$; Spinach RNA: 10 mM Tris–HCl, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$; P4-P6 RNA: 10 mM Tris–HCl, pH 7.4, 50 mM NaCl, 10 mM $MgCl_2$; class I Ligase ribozyme: 50 mM Tris, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$). The sample was cooled to room temperature for 10 min and incubated for 30 min with Fab BL3–6 ranging from 2 nM to 2μM in a final volume of 40 μl. The Bio-Dot apparatus from Bio-Rad was assembled by placing a BA85 nitrocellulose filter (Whatman) at the top and Hybond filter at the bottom (Amersham Pharmacia) and wells were pre-equilibrated with 100 μl of selection buffer. The Fab–RNA complex was applied and washed 2 times with 100 μl of the selection buffer at a time. Both filters were air dried, exposed to Phosphor-Imager screens, scanned with a Typhoon Trio imager (GE Healthcare) and the amount of RNA retained in each of the filters was quantified by using Image Quant software (Molecular Dynamics). The dissociation constants were calculated by fitting the data of fraction of RNA retained in the nitrocellulose membrane versus the concentration of the Fab to the following equation:

$$F = F_0 + F_{\max}\left(\frac{[\text{Fab}]}{K_d + [\text{Fab}]}\right)$$

where F represent the fraction of bound RNA at a given concentration of the Fab, K_d is the dissociation constant and F_0 and $F_{\max}$ are the minimum and maximum fractions of the bound RNA, respectively.

Hydroxyl-radical foot-printing

Hydroxyl radical foot-printing assay was performed as described previously (21). Briefly, ∼20 pmols of RNA was labeled at 5′ end by ^{32}P and purified by denaturing polyacrylamide gel electrophoresis. Binding reaction was performed for 30 min by a constant amount of radiolabeled RNA and 1μM BL3–6 Fab in a final volume of 5 μl. To initiate hydroxyl radical cleavage, 1 μl each of 0.6% H_2O_2, 10 mM sodium ascorbate and 1:2 Fe-EDTA (1 mM Fe (II) and 2 mM EDTA) solutions were spotted as separate droplets on the wall of the Eppendorf tube containing RNA–Fab complex or control samples and centrifuged briefly to mix the reagents together. The reaction was carried out for 5 min and quenched by adding 2 μl of a solution containing 100 mM thiourea, bromophenol blue and xylene cyanol dyes. The hydroxyl radical foot-printing patterns were then analyzed by performing denaturing polyacrylamide gel electrophoresis. Gels were exposed to Phosphorimager screens and scanned on a Typhoon Trio imager (GE Healthcare). The foot-printing bands were quantified by using Image Quant software (Molecular Dynamics).

Crystallization of Fab–RNA complex

An aliquot of RNA sample was refolded in appropriate buffer (Spinach RNA: 10 mM Tris–HCl, pH 7.4, 100 mM KCl, 5 mM $MgCl_2$; P4-P6 RNA: 10 mM Tris–HCl, pH 7.4, 50 mM NaCl, 10 mM $MgCl_2$; class I Ligase ribozyme: 50 mM Tris, pH 7.4, 150 mM NaCl, 10 mM $MgCl_2$). For refolding, RNA was heated at 90°C for 1 min in water and then incubated at 50°C for 15 min in folding buffer followed by incubation at room temperature for 5 min and in ice for 5 min. The refolded RNA was then incubated with 1.1 equivalents of the Fab at room temperature for 30 min and concentrated to 6 mg/ml using 10 kDa cut-off, Amicon Ultra-15 column. The formation of Fab–RNA complex was confirmed by native polyacrylamide gel electrophoresis (nPAGE). To decrease the number of nucleation events, RNA was then passed through the 0.2 μm

cutoff, Millipore centrifugal filter units. A Mosquito liquid handling robot (TTP Labtech) was used to set up high-throughput hanging-drop vapor-diffusion crystallization screens at room temperature using commercially available screening kits from Hampton Research, Sigma and Jena Bioscience. Several trials were reproduced in larger 1 μl + 1 μl hanging drops on siliconized glass slides. Crystals appeared and grew to full size within 2–3 days. For cryoprotection, drops containing suitable crystals were brought to 30% glycerol, keeping all other compositions isotonic. Crystals were immediately flash-frozen in liquid nitrogen after being fished out from the drops and taken to Argonne National Laboratory for collecting the X-ray diffraction data.

Structural data collection, processing and analysis

The X-ray diffraction data sets were collected at the Advanced Photon Source NE-CAT section beamline 24-ID-C. All the datasets were then integrated and scaled using its onsite RAPD automated programs (https://rapd.nec.aps.anl.gov/rapd/). Initial phases were obtained by molecular replacement with previously reported structure Spinach RNA aptamer in complex with Fab BL3–6 (PDB code: 4ZKE) as the search model using Phaser on Phenix. Model building was performed by using COOT, and then refinement was carried out with the Phenix package. The crystallographic statistics after the final refinement are in Supplementary Figures S8 and S11. Solvent-accessible surface area and area of interaction were calculated using PDBePISA. All figures were made in PyMOL (Schrodinger) and figure labels were edited in CorelDraw (Corel Corporation).

Phage display strategy for Fab selection

The construction of the phage display library and subsequent selection for the affinity maturation of Fab BL3–6 was performed by following similar strategies as described elsewhere (14,16). In brief, the mutagenic oligonucleotides were phosphorylated and annealed to the uracilated single stranded DNA template derived from parent Fab BL3–6 for Kunkel mutagenesis (28). The mutant closed circularized plasmid was electroporated into freshly prepared TG1 electrocompetent cells and after half an hour of recovery in SOC media infected with M13KO7 helper phage and amplified for 18 h. The phages were purified as per reported protocol (14,16). The class I ligase ribozyme (Bartel Ligase Product, BLP RNA) was used as a RNA target for the selection and the selection was performed at room temperature using streptavidin coated magnetic beads (Promega). The BLP RNA construct containing an additional 3′ overhang sequence, 5′-AGG UCG ACU CUA GAG GAU CCC CGG (x-module) was hybridized with the biotinylated DNA oligonucleotide, 5′-Biotin-ACC GGG GAT CCT CTA GAG TC and this RNA–DNA hybrid was immobilized on the magnetic beads via biotin-streptavidin affinity interaction. In the first round, 500 nM of RNA was immobilized using a predetermined amount of beads required for complete immobilization and then incubated with 10^{13} cfu of phages for 15 min in 1 ml of phosphate-buffered saline (PBS) buffer (8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 137 mM NaCl, and 3 mM KCl), 0.05% Tween 20, 2.5 mM EDTA, 12.5 mM $MgCl_2$, pH 7.4 supplemented with 0.1 mg/ml bovine serum albumin, 0.1 mg/ml streptavidin and 1 U/μl RNase inhibitor (NEB).The solution was then removed, and the beads were washed twice with the same buffer as above. In the subsequent rounds, purified phage pools were first incubated with streptavidin beads in the buffer for 30 min and the supernatant was used in the subsequent selection on a King Fisher magnetic particle processor (Thermo Electron Corporation). The 10^{11} cfu of phages were incubated for 15 min with (50 nM in second and third round and, 5 and 0.5 nM in fourth round) of the RNA in 100 μl of the buffer, supplemented with 0.1 mg/ml bovine serum albumin, 1 U/μl RNase inhibitor, and 1.5 μM X-module DNA–RNA hybrid as a competitor. Streptavidin magnetic beads were then added to the solution for 15 min to allow the capture of the biotinylated RNA construct together with the bound phages. The beads were then blocked with 50 μM biotin, washed five times with the buffer and eluted in 50 μl of elution buffer, PBS and 1 μg/ml biotinylated RNaseA). The biotinylated RNase A was removed from the resulting phage library by incubation with streptavidin beads. After each round of selection, recovered phages were amplified as described previously. After third and fourth round of selection, phages were sequenced. Eighteen unique clones were selected for expression as soluble proteins and purified as described elsewhere (14,16). The binding affinity of each clone with the BLP RNA without the x-module was determined by filter binding assay.

Fab expression and purification

Output clones were sequenced and reformatted with the introduction of stop codon on phagemids using Q5 site-directed mutagenesis kit (NEB). The desired Fab proteins were then expressed as soluble proteins according to published protocols (14,16). Collected cell pellets were lysed, and Fab proteins were purified using the AKTAxpress fast protein liquid chromatography (FPLC) purification system (Amersham) as described. Purified protein was dialyzed into 1 × PBS (pH 7.4), concentrated and analyzed by 12% sodium dodecyl sulphate-polyacrylamide gel electrophoresis using Coomassie Blue R-250 staining for visualization. Aliquots of Fab samples were tested for RNase activity using the RNaseAlert kit (Ambion). The aliquots of Fab samples were flash frozen in liquid nitrogen and stored at −80°C until further use. Additional methods and requests for material including Fab BL3–6 and Fab BL3–6S97N (if available), Fab plasmids and the detailed protocols for expression and purification of Fabs will be provided upon request.

Additional details on material and methods including the nucleic acid sequences are available in Supplementary Data.

RESULTS AND DISCUSSION

Selection and characterization of RNA epitopes against Fab BL3-6

Previously, the specificity of Fab BL3–6 for its cognate RNA hairpin was tested only against a limited number of mutants or unrelated sequences (16). Here, we performed a more exhaustive test of specificity by challenging Fab BL3–6 to recognize RNA sequences from a partially structured

library consisting of a constant 10-bp stem and randomized 9-nucleotide loop flanked by invariable primer binding sequences (Figure 1A). We employed a nitrocellulose membrane based *in vitro* selection approach (24) (Figure 1B) to screen this RNA library containing 262 144 different RNA sequences for binding to Fab BL3–6. Fab BL3–6 was incubated with the RNA library and the mixture was passed through the membrane to capture RNA sequences retained on the nitrocellulose membrane in the presence of Fab. Fab bound RNA was eluted from the membrane, reverse transcribed, PCR amplified and transcribed to generate RNA pool for the next round of selection. To facilitate enrichment of high affinity binders, we increased the selection stringency in successive rounds by decreasing the concentration of Fab BL3–6 and increasing the number of washing steps. The progress of the selection was monitored by native polyacrylamide gel electrophoresis (Supplementary Figure S1). We applied a negative selection in every other round by passing the RNA pool through a separate nitrocellulose membrane in the absence of Fab to reduce non-specific capture of RNA sequences by the membrane. After 5 rounds of selection, individuals from the RNA pool were cloned and sequenced.

Sequence analysis of the randomized 9 nt region within the RNA clones selected after five rounds revealed a highly enriched, stem-loop motif with consensus sequence GNGACCC (where N is any nucleotide), underscoring the sequence preference and motif specificity of Fab BL3–6 (Figure 1C). We synthesized the three most highly enriched RNA hairpins—named N9RNA1, N9RNA6 and N9RNA32—without the constant primer binding sites for biochemical characterization (see Supplementary Data for full RNA sequences). As predicted by mFold (http://unafold.rna.albany.edu/) (29), N9RNA1 and N9RNA32 form AGACC and UGACC pentaloops, respectively, whereas N9RNA6 is predicted to contain a 9 nucleotide AGAGACCCA loop (Supplementary Figure S2). First, we 5′-^{32}P-phosphate radiolabeled the hairpins along with the WT hairpin having AAACA loop and other controls and performed filter binding assays with Fab BL3–6.

As shown by the binding constant measurements (Figure 1D and Table 1), all three hairpins bind to Fab BL3–6 with greater affinity than does the WT hairpin, with binding constants ranging from 12 to 42 nM. In particular, the most enriched hairpin containing AGACC loop showed ∼8× higher affinity compared to that of WT hairpin (binding constants:12 versus 90 nM). Next, we performed mapping of the binding site of the Fab BL3–6 for these hairpins using hydroxyl radical foot-printing (Supplementary Figure S3). As expected, the Fab conferred protection in the loop regions of the hairpins similar to that in WT-RNA hairpin containing AAACA loop (Supplementary Figure S4).

Interestingly, the hairpin N9RNA6, which was predicted by mFold (http://unafold.rna.albany.edu/) (29) to form a 9-nt loop, AGAGACCCA also showed protection similar to that of N9RNA1 hairpin loop, suggesting the formation of an AGACC loop closed by a G-C base pair (underlined bases in above sequence), consistent with the preference of Fab BL3–6 for a pentaloop closed by a GC pair (see Supplementary Figure S3 for hydroxyl-radical foot-printing gel). Fab BL3–6 concentration dependent foot-printing also confirmed the values of binding constants obtained from filter binding assays (Supplementary Figure S5). Grafting of the stem-loop motifs of N9RNA1 and N9RNA6 into helix P2 of Spinach RNA and P5 stem-loop of the class I ligase ribozyme product (Bartel ligase product, BLP) had no effect on Fab BL3–6 binding affinities compared to respective, standalone hairpins (Table 1, see Supplementary Data for sequences of the RNA constructs), demonstrating the portability of these newly selected RNA motifs for chaperone assisted RNA crystallization.

Specificity of Fab chaperone modules

The N9 RNA library used above is structurally constrained and biased toward formation of hairpins. However, to utilize the Fab–RNA system as a portable chaperone module for RNA crystallization or as an epitope tag for affinity purification, immunoprecipitation or imaging, the Fab must recognize and bind its cognate RNA with high affinity and specificity. To test the sequence and motif specificity of Fab BL3–6 more broadly, we designed a larger and unconstrained N40 library consisting of randomized 40-nt region flanked by constant primer binding sequences (Supplementary Figure S6). We used the same *in vitro* selection procedure as described above to screen the binders from ∼10^{15} different RNA sequences. After performing four rounds of selection, RNA sequences were cloned and sequenced. Clustal Omega Multiple Sequence Alignment (27), Multiple Em for Motif Elicitation (MEME)—a motif search software (http://meme-suite.org/tools/meme) (26) and WebLogo 3 (http://weblogo.berkeley.edu/) (25) were used to analyze the RNA sequences. The majority of the sequences (∼70%) contained the GNGACCC consensus sequence (Supplementary Figure S6), and as predicted by mFold (http://unafold.rna.albany.edu/) (29) the consensus sequence formed a pentaloop hairpin motif closed by a GC pair, which matches the outcome of the N9-library selection (compare Figure 1D and Supplementary Figure S6). Representatives from the remaining ∼30% of the RNA sequences did not bind to Fab BL3–6 and showed a high background retention in the nitrocellulose membrane in the filter binding assay. Possibly the retention of these sequences during the selection process reflects the limited number of rounds (four) in combination with an ability to bind to nitrocellulose membrane non-specifically.

Next, we characterized selected clones by filter binding and hydroxyl radical foot-printing assays. As expected, the clones containing GNGACCC motif(s) bound Fab BL3–6 with nanomolar affinities (Table 1, see Supplementary Data for full RNA sequences). Moreover, Fab BL3–6-induced protection from hydroxyl-radical cleavage in the regions of RNA sequence where NGACC was predicted to be present as a loop within a hairpin motif. In the clones containing multiple GNGACCC sequences, the Fab-induced protection was observed only in the region predicted to be a stem-loop motif. These results establish that Fab BL3–6 is highly specific for binding to a stem-loop RNA structure containing consensus GNGACCC sequence as the loop and closing base pair. Our observations suggest that in addition to serving as a chaperone for RNA crystallography, Fab BL3–6 could also be developed as a reagent for affinity tag-

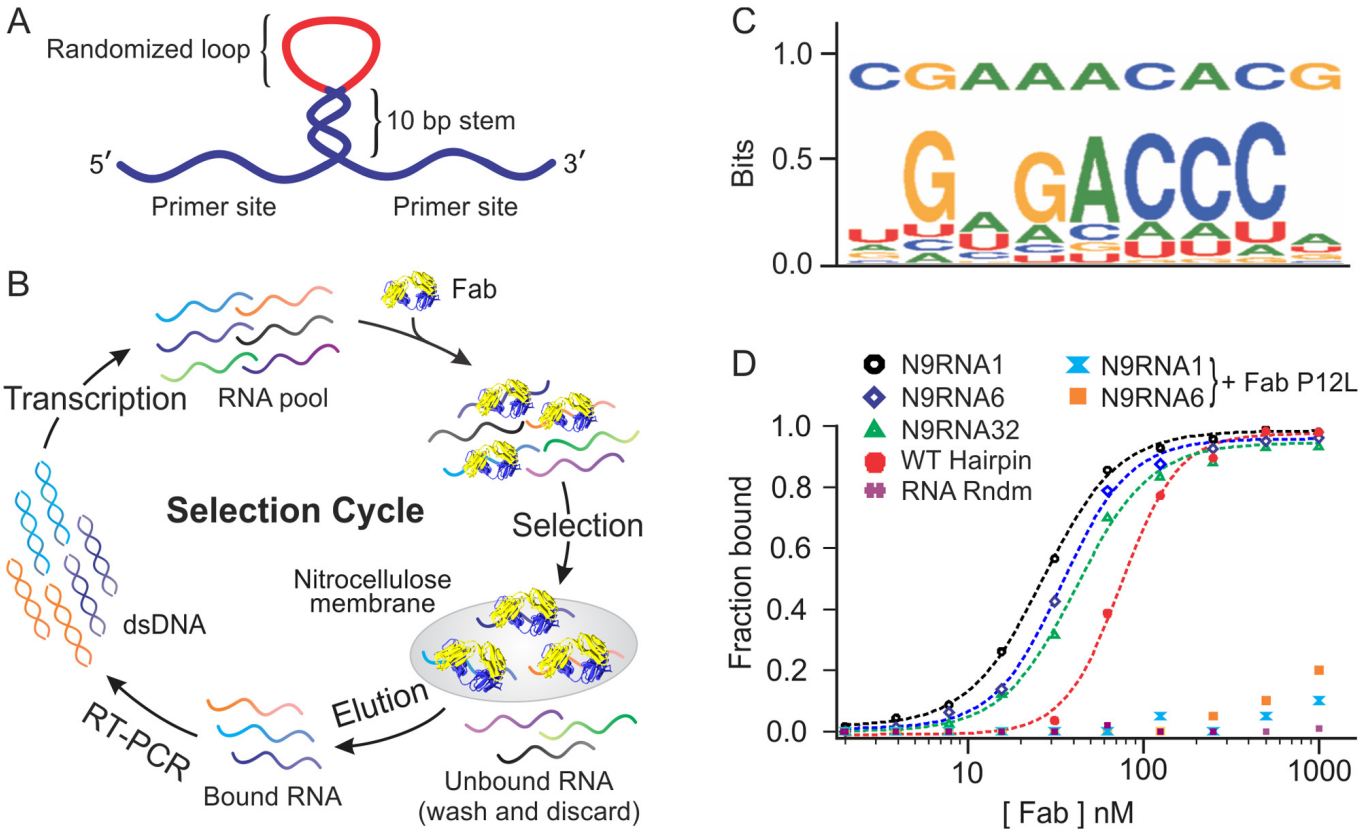

Figure 1. *In vitro* selection strategy for optimization of the RNA epitope that binds Fab BL3–6. (**A**) Design of the N9-library. (**B**) *In vitro* selection process. (**C**) Sequence logo of randomized region for the RNA sequences evolved after five rounds of selection. Sequence at the top represents the WT-RNA pentaloop hairpin that binds to Fab BL3–6. (**D**) Plots of fraction bound as a function of varying concentration of Fab BL3–6 for the select RNA clones as accessed from filter binding assays. Reported K_ds are the values (average ± standard deviation) obtained from ≥3 independent experiments.

Table 1. Dissociation constants (K_ds) for various RNA constructs that bind to Fab BL3–6

RNA constructs	BL3–6 binding sequence	K_d/nM (Avg ± SD)	RNA constructs	BL3–6 binding sequence	K_d/nM (Avg ± SD)
WT-RNA HP	GAAACAC	90 ± 12	Spincah6	GAGAGACCCAC	35 ± 10
N9RNA1 HP	GAGACCC	12 ± 6 (30 ± 1)	WT-BLP	GAAACAC	120 ± 10
N9RNA6 HP	GAGAGACCCAC	25 ± 6	BLP1	GAGACCC	6 ± 2
N9RNA32 HP	GUGACCC	42 ± 12	BLP6	GAGAGACCCAC	32 ± 5
A12U HP	GUAACAC	100 ± 14	N40RNA1	AGUG ACCU	118 ± 13
A13G HP	GAGACAC	16 ± 8	N40RNA26	GCGACCC	74 ± 9
A16C HP	GAAACC	108 ± 6	N40RNA52	GCAACAC	250 ± 8
A13G-A16C HP	GAGACCC	12 ± 4	N40RNA55	GUGACCC	75 ± 4
WT-Spinach	GAAACAC	100 ± 5	N40RNA59	AAGACCU	100 ± 6
Spinach1	GAGACCC	10 ± 2	N40RNA62	GAGACCC	45 ± 11

K_ds were obtained by filter-binding assays in 10 mM Tris–HCl, 200 mM NaCl, 2.5 mM MgCl2, pH 7.5 buffer at 23°C. Value in parenthesis for N9RNA1 represents the K_d that was obtained from hydroxyl-radical foot-printing. Standard Deviations are from ≥3 independent experiments. For comparison, only the Fab binding stem-loop motifs are shown. Full sequences of all RNA constructs are provided in the Supplementary Data.

based RNA purification, immunoprecipitation or visualization. Fab BL3–6 could also serve as a reagent to probe transcriptomes for the presence of the RNA motif. Supporting these hypothetical applications, preliminary experiments show that Fab BL3–6 can specifically and effectively bind its cognate RNA targets spiked into the cell lysates.

Crystallization of Fab BL3-6 bound to the newly selected GNGACCC motif

Since grafting of GAGACCC and GAGAGACCCAC stem-loop motifs into the helix P2 of Spinach RNA and P5 stem-loop of BLP RNA showed notably higher binding affinities to Fab BL3–6 compared to the corresponding GAAACAC grafted RNAs (Table 1), we advanced RNA constructs for crystallization trials corresponding to Spinach1 (grafted with GAGACCC motif), Spinach6

(grafted with GAGAGACCCAC motif), BLP1 (grafted with GAGACCC motif) and BLP6 (grafted with GAGAGACCCAC motif) along with the respective RNAs grafted with GAAACAC motif. These RNAs alone or in complex with the Fab BL3–6 were screened for crystallization in a total of 480 conditions from five different commercially available kits. None of these RNAs alone gave crystals, and constructs grafted with the 9-nt AGAGACCCA loop gave no hits even in the presence of Fabs, perhaps due to greater flexibility of the 9-nt loop compared to the 5-nt AAACA or AGACC loops. We observed several crystallization hits for Fab BL3–6: BLP1 RNA complex but observed no obvious improvement in crystallizability or quality of the crystals compared to that WT BLP RNA (with GAAACAC motif) in complex with Fab BL3–6 reported previously (16). For WT Spinach and Spinach1 constructs in complex with Fab BL3–6 (Supplementary Tables S2 and S3), we observed a similar number of hits; however, the crystals of Fab BL3–6:Spinach1 RNA complex were visually larger, single and isolated compared to those of Fab BL3–6:Spinach RNA complex under the same condition(s). Ultimately, to determine the broader utility of these motifs in RNA crystallography, it will be necessary to test wider variety of RNA targets in combination with Fab BL3–6 and its engineered variants (see below).

Previously, we used the 7-nt GAAACAC stem-loop (WT) motif as a portable binding site to crystallize the Spinach RNA aptamer using Fab BL3–6 as a crystallization chaperone (21). To understand the structural basis for the improved affinity of the N9RNA1 hairpin and to demonstrate its application as a crystallization module, we replaced the parent GAAACAC sequence in the helix P2 of previously crystallized Spinach RNA construct (WT Spinach) with the affinity matured GAGACCC motif and performed crystallization trials with the new Spinach RNA construct (Spinach1 RNA) in complex with Fab BL3–6. The crystals were grown using the hanging drop vapor diffusion method. We observed fully-grown hexagonal rod-shaped crystals within two weeks at room temperature, which diffracted to 2.09Å resolution. In side-by-side crystallization trails for Fab BL3–6 in complex with Spinach RNA the crystal hits for GAAACAC and GAGACCC motifs occurred under overlapping conditions. Thus, in contexts where one of these sequence motifs might alter the conformation of the RNA through formation of an alternative secondary or tertiary interaction(s), the other may be useful. Importantly, we also observed two conditions in which only the crystals of Spinach1 RNA–Fab complex formed, indicating that GAAACAC and GAGACCC portable motifs in complex with Fab BL3–6 collectively can enable crystallization under a wider range of conditions.

To solve the crystal structure of Fab–Spinach1 RNA complex, we obtained initial phases by molecular replacement using the previous crystal structure of WT Spinach in complex with Fab BL3–6 (PDB code: 4KZE) (21) as a search model. After model building and refinement, the final values of R_{free} and R_{work} were 22.6 and 19.3%, respectively. The overall crystal structure and the interactions between the Fab and RNA were very similar to those observed previously in the WT Spinach RNA–Fab complex (Figure 2A) except that we observed additional non-canonical pairing between G39 and C42 (Figure 2C). These nucleotides form two hydrogen bonds, whereas the corresponding nucleotides in the WT (A39 and A42) form only one (Figure 2B). Presumably, the non-canonical pair stabilizes the binding conformation of the RNA loop leading to increased affinity with Fab BL3–6. Mutational analysis confirmed the importance of the G39–C42 interaction. We generated a series of AAACA loop mutations (Supplementary Figure S7) and tested their binding affinities for Fab BL3–6 by filter binding assay (Table 1, see Supplementary Data for full RNA sequences). Consistent with the crystal structure, substitution of A39 with G and A42 with C enhanced the affinity of the hairpin mutant toward the Fab BL3–6 by eightfold. The enhanced affinity could reflect pre-organization of the loop for binding or, possibly, a smaller desolvation penalty upon Fab binding.

Affinity maturation of Fab BL3-6

Affinity maturation of a Fab antibody can be performed by introducing diversity in a directed (30) or non-directed (31,32) manner followed by selection for high affinity variants under increased selection stringency, either by varying the antigen concentration or salt concentration. Fab BL3–6 was obtained previously by affinity maturation of Fab BL3 using the non-directed mutagenesis approach of error-prone PCR to generate the library (16). Fab BL3 was derived from phage display selections against the BLP ribozyme using the YSGR Fab library derived from the humanized Fab4D5 framework (33,34). Compared to the parent Fab BL3, Fab BL3–6 bound to the BLP with 10-fold greater affinity (K_d ∼500 versus ∼50 nM). Here, we sought to improve further the binding affinity of Fab BL3–6 using a directed mutagenesis approach (35). In this mutagenesis approach, amino acid diversity is introduced at specific sites, such as CDRs that are involved in antigen recognition, and is tailored precisely toward favoring specific types of amino acids. Based upon the crystal structure of the BLP RNA and Spinach RNA in complex with Fab BL3–6 (PDB code: 3IVK, 4KZE) (16,21), we designed and prepared a Fab library that introduced diversity at specific sites in light chain CDR L3 and heavy chain CDRs H1 and H2 while maintaining the parent CDR H3.

Library design

To prescribe amino acid diversity for affinity maturation, we identified paratope positions in the crystal structures of both BLP and Spinach RNA Fab BL3–6 complexes where amino acids interact directly with the RNA hairpin GAAACAC via hydrogen bonding or stacking interactions (Figure 3A) (16,21). We also identified CDR residues that support the CDR conformations through contact with other amino acids (Supplementary Figure S9). Within the Fab RNA interface, CDR H3 makes the most dominant contribution (ca. 50%) to the total buried surface area (BSA), forming a cleft together with CDR H2 through which the RNA traverses. CDR H3 is enriched with arginines that interact with both the RNA nucleosides and the phosphodiester backbone. Retaining the CDR H3 from Fab BL3–6, we randomized positions in the other CDRs, challenging them to synergize with CDR H3 to enhance binding

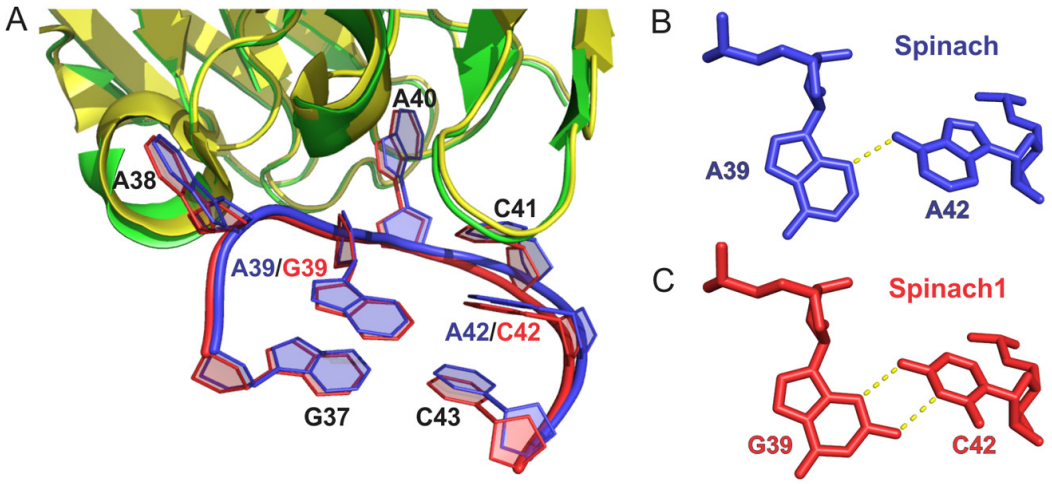

Figure 2. Crystal structures of Fab–RNA complex for Spinach RNA grafted with WT or newly selected, N9RNA1 portable sequence. (**A**) Superposition of the Fab–RNA interface for two structures. The complete superposition of these crystal structures and details of the crystallographic statistics are in Supplementary Figure S8. For the WT complex the RNA is colored blue and Fab is colored yellow. For the Spinach1 complex the RNA is red and the Fab is green. (**B** and **C**) The most significant difference between the two structures is the additional hydrogen bonding interaction between residues 39 and 42 enabled by the A39G and A42C mutations. Presumably this additional interaction accounts for the higher affinity of Fab for N9RNA1 motif compared to that of WT-RNA.

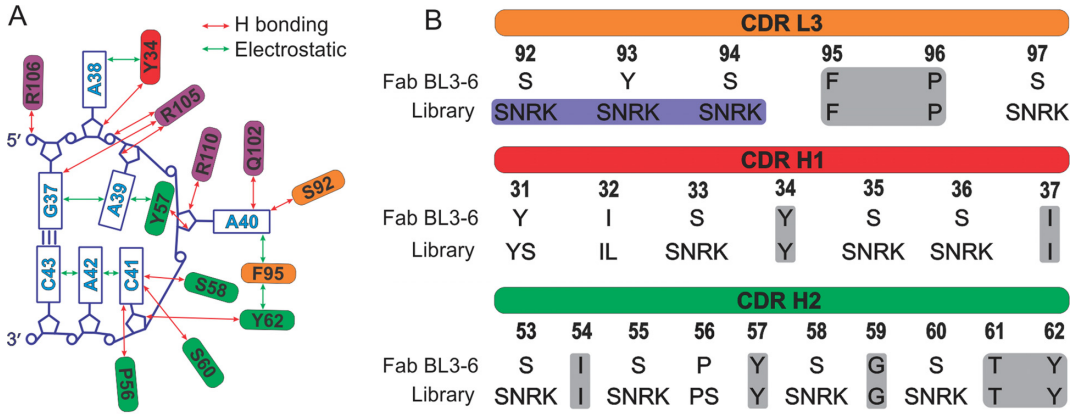

Figure 3. Design of Fab library for the affinity maturation of Fab BL3–6. (**A**) A schematic of epitope–paratope contacts between Fab BL3–6 and GAAACA loop (PDB ID: 4KZE (21) and (**B**), library design for the affinity maturation of Fab BL3–6. Gray shade indicates amino acids retained from the parent Fab and blue shade indicates length variability between 6 and 9 amino acids. The complete version of the library design strategy with amino acid sequences from all CDRs are given in Supplementary Figure S9. The nucleotide and residue numbers are according to Fab BL3–6: Spinach RNA co-crystal structure PDB ID: 4KZE (21).

affinity. We also retained the parent CDRs for L1 and L2, as neither interacts with the RNA. The peptide backbone of CDR L3 forms a curved structure that winds around the flipped-out nucleobase, A40 (Figure 3A). Note that this A40 corresponds to A62 in Fab BL3–6: ligase I ribozyme co-crystal structure, PDB code: 3IVK. A40 forms a hydrogen bond via its exocyclic amine to S92 of CDR L3 and stacks on the aromatic ring of F95, which in turn, stacks with Y62 of CDR H2 (Figure 3A and Supplementary Figure S9). The previous 10-fold increase in affinity provided by Fab BL3–6 over the parent Fab BL3 was attributed to the S95F mutation. At position 96 in CDR L3 proline occurred exclusively (51%, in all organisms according to Kabat database (www.abysis.org) (36). These observations implicate F95 and P96 as important for structural folding, and therefore we retained these residues.

We chose SNRK amino acids for randomization as they represent commonly enriched amino acids in RNA-protein or RNA-antibody binding interfaces and have enabled successful affinity maturation of Fab 5 for binding to the P4-P6 RNA domain. (J.D.Y. and J.A.P., unpublished results). Accordingly, we randomized positions 92–94 and 97 within L3 along with length variability ranging from 6 to 9 residues using SNRK (Figure 3B). CDRs H1 and H2 contribute ca. 40% of the total BSA within the Fab–RNA interface and play important roles in recognition of the RNA loop through hydrogen bonding and stacking interactions involving serines and tyrosines, respectively (16,21). Y34

(CDR H1), Y57 and Y62 (CDR H2) stack upon A38, A39 and F95, respectively (Figure 3 and Supplementary Figure S9). Thus, we retained these residues along with other CDR H1 and H2 residues, including I37, I54, G59 and T61, which are known to be important for structural/conformational folding and flexibility and occurred exclusively in all organisms (according to Kabat database, www.abysis.org) (36). Positions Y31, I32 (CDR H1) and P56 (CDR H2) were randomized with YS, IL and PS, respectively, whereas the remainder of CDR H1 and H2 positions (S33, S35, S36 from H1 and S53, S55, S58, S60 from H2) were randomized with SNRK (Figure 3 and Supplementary Figure S9). Our previous work has shown that the inclusion of positive-charged, basic amino acids within in CDR H3 enhances library performance (16). Here we wanted the test whether the additional inclusion of R and K residues in other CDRs could impart further benefit. Results from this design could inform design of naïve repertoires for RNA tailored Fab libraries.

Library preparation and Fab selection

Oligonucleotide directed mutagenesis was performed on a single-stranded, uracilated DNA template derived from parent Fab BL3–6. A single degenerate codon ARW was used for SNRK amino acids, which produces equal proportions of all four amino acids in each of the prescribed variable positions in CDRs L3, H1 and H2 (Figure 3B). As described above, a total of 14 positions were chosen for randomization, and length variability from 6 to 9 residues was included in CDR L3, yielding a maximum theoretical diversity of 2.3×10^9, which is close to the maximum achievable practical diversity of $\sim 10^{10}$. The resulting phage library was panned against BLP RNA immobilized on SAV magnetic beads and cycled through four rounds of binding selections. As the selection progressed, we decreased the antigen concentration from 500 nM in the first round to 0.5 nM in the final round. The binding reactions in rounds two to four included an excess of a DNA/RNA hybrid duplex as a competitor (1.5 μM). Phages were sequenced after the third and fourth rounds of selection. A total of 18 unique clones were found, all contained the parent Fab sequences in CDR L1, L2 and H3 as prescribed the library. Among the selected clones, CDR H2 remained unchanged, CDR H1 carried mutations in seven clones and L3 carried mutations in all sequenced clones. Individual clones were expressed in phosphate-depleted media and purified as an RNase free soluble protein. Three clones exhibited no detectable expression; five clones expressed in low yield (0.1–2 mg/l), and the remaining ten expressed in moderate to good yields (3–9 mg/l). Binding of these clones was assayed using nitrocellulose membrane filter binding with 5′-^{32}P radiolabeled BLP RNA and isolated hairpin (both containing the GAAACAC motif). All clones showed improved binding affinity compared to the parent Fab BL3–6 (Table 2 and Supplementary Table S1) with three clones showing ∼8- to10-fold higher affinity and the others ∼5-fold higher affinity.

Amino acid composition within CDRs and binding affinities of the affinity-matured Fabs

Given that the unchanged CDR-H3 likely fixes the binding orientation and conformation of the RNA with respect to the other CDRs, the selection challenges CDR L3, H1 and H2 variants in the library to identify better ways to bind the same RNA surface than do the corresponding CDRs in the parent. CDR H2 emerged from the selection as the WT sequence (Table 2 and Supplementary Table S1). The parent CDR H2 already makes significant contact with the RNA, and the limited diversification in the library likely contained no CDR H2 sequences that enhanced binding. A few clones showed mutations in CDR H1 (Fab clone 17 and 18, Table 2) along with recurring mutations in CDR L3. Within CDR L3, mutation S97N occurred consistently in all sequences and S92 exhibited a preference for either S or N over R and K (Supplementary Figure S10). The Spinach RNA: Fab BL3–6 co-crystal structure suggests that the S92 hydroxyl group might form a weak hydrogen bond with the A40 exocyclic amine (O – N distance ∼4 Å). S92N could maintain or strengthen the A40 interaction. The S97 hydroxyl group resides outside H-bond distance of the A40 amine (∼6 Å). The S97N preference might reflect formation of a new hydrogen bond(s) with the A40 amine, enabled by capacity of the longer asparagine side chain to position its keto oxygen closer to the amine. Below, we confirm this hypothesis through structural analysis. We performed site-directed mutagenesis on the Fab BL3–6 and mutated the S97 to N. Indeed, relative to the parent Fab BL3–6, clone BL3–6S97N showed 20-fold better binding (Table 2), indicating that this mutation alone can confer improved affinity.

We also tested binding of our affinity matured Fabs to the selected RNA hairpin motif (GAGACCC) to assess whether the Fab and RNA mutations act synergistically. Affinity matured Fabs also showed enhanced binding to the newly selected GAGACCC motif compared to the parent GAAACAC motif (by ∼2- to 3-fold). However, the difference in binding between the parent and newly selected RNA motifs is smaller in the background of the affinity matured Fab than with the parent Fab BL3–6 (∼8- to 10-fold, Table 2 and Supplementary Table S1).

Crystallization trials with affinity matured BL3-6 mutants

Two affinity matured Fab BL3–6 variants Fab 10 and Fab 3, chosen as representative clones, along with Fab BL3–6 and Fab BL3–6S97N mutant were advanced to side-by-side crystallization trials in complex with three different RNAs: BLP, Spinach RNA and P4-P6 domain of group I intron. The BLP RNA already contained the cognate stem-loop, and we engrafted the Fab BL3–6 binding loop (GAAACAC) into Spinach and the P4-P6 RNA. We used filter binding to confirm that Fab BL3–6 and the two affinity matured variants bind to the RNAs with the expected affinity. Using five different commercially available crystallization kits (Sigma, Sigma-Aldrich), Natrix, Index, PEG/Ion (Hampton) and JBS Screen Nuc-Pro 1–4 (Jena Bioscience) yielding a total of 480 conditions, crystallization trials were performed for these RNAs alone and in complex with either Fab BL3–6, Fab 10, Fab 3 and Fab BL3–6S97N. In

Table 2. Representative Fab sequences from the fourth round of selection of Fab BL3–6 L3 library, yield and dissociation constants (K_ds) determined by filter binding assay

Fab clone	CDR-L3						CDR-H1							Yield mg/L	K_d/nM WT BLP (Avg ± SD)	K_d/nM WT HP (Avg ± SD)	K_d/nM AM HP (Avg ± SD)
	92	93	94	95	96	97	31	32	33	34	35	36	37				
BL3-6	S	Y	S	F	P	S	Y	I	S	Y	S	S	I	4.0	120 ± 10	90 ± 12	12 ± 6
3	S	S	S	F	P	N	Y	I	S	Y	S	S	I	9.2	28 ± 2	25 ± 5	09 ± 1
5	S	N	K	F	P	N	Y	I	S	Y	S	S	I	5.0	20 ± 1	28 ± 4	08 ± 2
10	S	S	N	F	P	N	Y	I	S	Y	S	S	I	9.4	30 ± 1	28 ± 3	09 ± 3
12	S	S	R	F	P	N	Y	I	S	Y	S	S	I	5.4	37 ± 1	17 ± 1	07 ± 3
17	S	N	S	F	P	N	Y	I	N	Y	S	S	I	8.6	42 ± 1	21 ± 2	08 ± 3
18	N	N	S	F	P	N	S	I	K	Y	K	S	I	5.0	51 ± 3	22 ± 2	09 ± 5
BL3-6S97N	S	Y	S	F	P	N	Y	I	S	Y	S	S	I	5.0	5 ± 5	13 ± 1	09 ± 2

WT BLP = class I Ligase ribozyme, WT HP = RNA hairpin containing AAACA loop, AM HP = affinity matured hairpin containing AGACC loop. Standard deviations were determined from three independent measurements.

general, RNA-only trials gave far fewer hits, defined as visual observation of crystals under optical microscope if any, compared to RNA-Fab complex trials (Supplementary Table S2). The affinity matured variant, Fab 3 and Fab 10, gave a similar number of hits in the case of P4-P6 RNA. For BLP RNA, Fab 3 gave no hits; however, in the case of Spinach, Fab 3 and 10 gave far fewer hits than the parent Fab BL3–6 gave. The mutant Fab BL3–6S97N gave a greater number of hits compared to parent Fab BL3–6 with Spinach RNA (Supplementary Tables S2–S4). Mostly, these occurred under the same conditions as Fab BL3–6, but, for the small subset of conditions, we observed that the affinity matured Fabs facilitated crystal growth under conditions where Fab BL3–6 could not. Additionally, we are investigating the crystallization of RNA targets such as self-alkylating ribozyme (37), DIR RNA aptamer (38), yjdF riboswitch (39), IRES of Hepatitis C Virus (40) and XRN1-resistant RNA motif from a Flavivirus (41) grafted with the GAAACAC or GAGACCC in complex with Fab BL3–6 or Fab BL3–6S97N. In all cases, the grafted RNAs formed high affinity complexes with the Fabs. In several cases, we obtained crystals, and in at least one case, the crystals showed good diffraction.

Crystal Structure of Fab BL3-6S97N bound to the GAAACAC motif

To determine the structural basis for the higher affinity of mutant Fab BL3–6S97N compared to Fab BL3–6, we crystallized the Spinach RNA grafted with GAAACAC loop complexed with Fab BL3–6S97N. The crystals formed in several different conditions from the PEG/Ion, Index (Hampton) and JBS Screen Nuc-Pro 1–4 (Jena Bioscience) crystallization kits using the hanging drop vapor diffusion method as hexagonal rods with morphology similar to that for the corresponding complex with Fab BL3–6. Several crystals from different conditions diffracted in the range of 1.64 Å to 2.0 Å. The crystal structure was solved at 1.64 Å resolution, a significant improvement over our previous 2.4 Å structure (21), by molecular replacement using the crystal structure of Spinach RNA complex with Fab BL3–6 (PDB code: 4KZE) (21) as a search model. After model building and refinement, the values of R_{free} and R_{work} were 22.6 and 21.3%, respectively, after refinement. The overall paratope structure of the Fab BL3–6S97N mutant is very similar to that of Fab BL3–6 with Spinach RNA, except in CDR L3, where the side chain of S97 is replaced with the longer side chain of N97 (Figure 4A). The RNA epitope remained at the same position and aligns well with the Fab BL3–6: Spinach RNA complex crystal structure (r.m.s.d. for the overall complex = 0.568 Å, Supplementary Figure S11). In the structure, the longer side chain of the N97 positions the carboxamide group within the H-bond distance of A40, forming two H-bonds with its Watson–Crick face (Figure 4C). Presumably these additional hydrogen bonds account for the increased RNA binding affinity of Fab BL3–6S97N compared to Fab BL3–6 (Figure 4B). The only other structure difference between the two Fabs occurred at S92, where distinct side-chain rotamers were observed.

CONCLUSION

We have used two independent strategies for the affinity maturation of our existing Fab–RNA crystallization module. One strategy utilized an *in vitro* RNA selection process to improve the affinity of RNA epitope, in which Fab BL3–6 was challenged to recognize high affinity binders from a partially structured library of randomized RNA hairpin loops. After five rounds of selection, we isolated an RNA hairpin—GAGACCC, which binds to Fab BL3–6 with almost an order of magnitude higher affinity compared to WT GAAACAC hairpin. A co-crystal structure of Fab BL3–6 in complex with this motif suggests that the enhanced affinity derives from a non-canonical base pair between G_3 and C_6 of the hairpin loop. Importantly, the affinity of the hairpin for Fab BL3–6 remains high in the context of larger, structured RNAs, allowing this hairpin to serve as a modular affinity tag for chaperone assisted RNA crystallography, RNA purification or imaging. The GNGACCC consensus sequence also emerged from selections using a 40-nt randomized library unconstrained by

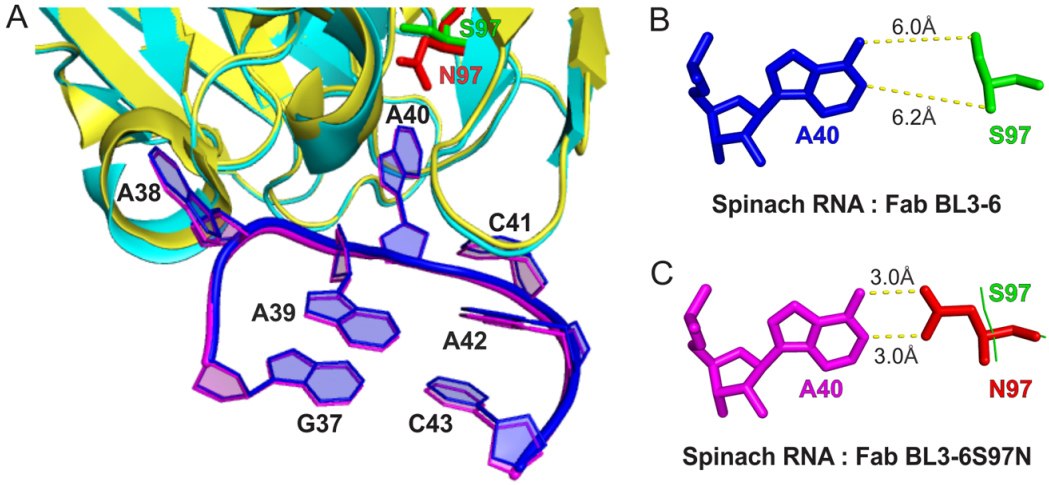

Figure 4. Crystal structures of GAAACA motif-grafted Spinach RNA in complex with Fab BL3–6 or BL3–6S97N. (**A**) Superposition of the Fab–RNA interface for two structures. The complete superposition of these crystal structures and details of the crystallographic statistics are in Supplementary Figure S11. For the Fab BL3–6: RNA complex, the RNA is colored blue and Fab is colored yellow whereas, for the Fab BL3–6S97N: RNA complex, RNA is colored magenta and Fab is colored cyan. (**B** and **C**) The most significant difference between the two structures is the two additional hydrogen bonding interactions between residue N97 from Fab BL3–6S97N and nucleobase A40, suggesting that these additional interactions account for the higher affinity of Fab BL3–6S97N for GAAACAC motif compared to that of Fab BL3–6.

secondary structure, underscoring the high specificity of Fab BL3–6 for its cognate RNA motif. During crystallization trials, we observed several conditions in which only parent chaperone module (GAAACAC hairpin with Fab BL3–6) or affinity matured module (GAGACCC hairpin with Fab BL3–6) crystallized, suggesting that these modules could be tested in parallel to increase the probability of RNA crystallization.

Our second strategy involved the affinity maturation Fab paratope, in which the composition of residues within specific CDR regions of Fab BL3–6 were randomized with SNRK to generate a library of Fabs. After four rounds of phage display selection, we isolated many Fabs that bound the GAAACAC hairpin with an order of magnitude greater affinity compared to the parent BL3–6 Fab. Interestingly, these newly developed Fabs also bound to affinity matured GAGACCC hairpin with slightly increased affinity (2- to 3-fold) compared to Fab BL3–6, for an overall enhancement of ∼10-fold. Affinity improvement was more striking for BLP RNA with overall affinity enhancement of ∼20-fold (Table 2, K_ds for Fabs BL3–6 and BL3–6S97N with BLP RNA are 120 ± 10 nM and 5 ± 5 nM, respectively). We attribute the increased binding affinity of these Fabs to the recurring S97N mutation in CDR L3. When two of these new Fabs together with the Fab BL3–6 were used in side-by-side crystallization trials with test RNAs containing a GAAACAC or GAGACCC motif, crystals formed under a greater variety of conditions compared to the parent Fab alone. As our affinity matured and portable Fab–RNA chaperone modules are highly specific, they have potential to serve as a reagent for affinity tag based RNA purification, immunoprecipitation and visualization or conceivably as a reagent for secondary structure identification within the transcriptomes. For crystallization trials, in particular, the enhanced affinity can allow more facile purification of the Fab and target RNA as a preformed complex. We envision a subset of these new Fabs will provide the starting point for the next stages of engineering the Fab BL3–6 crystallization module using surface entropy reduction and elbow angle restriction.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession numbers 6B3K and 6B14 for Spinach1 RNA aptamer—Fab BL3–6 complex and Spinach RNA aptamer—Fab BL3–6S97N complex, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

For crystallographic data collection, this research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under the contract No. DE-AC02–06CH11357. We would like to thank staffs of the Advanced Photon Source at Argonne National Laboratory for providing technical advice during crystallographic X-ray diffraction data collection. We also would like to thank Piccirilli laboratory members, especially to Benjamin Weissman, for critical review of the manuscript.

FUNDING

National Institutes of Health [R01AI081987, R01GM102489]; Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust (to J.A.P.); Defense Threat Reduction Agency

[HDTRA1–13-1–0004 to J.A.P., S.A.B.]; National Institute of General Medical Sciences [P41 GM103403]; National Institutes of Health; Funding for open access charge: National Institutes of Health [R01AI081987, R01GM102489].
Conflict of interest statement. None declared.

REFERENCES

1. Chen,K. and Rajewsky,N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.*, **8**, 93–103.
2. Ponting,C.P., Oliver,P.L. and Reik,W. (2009) Evolution and functions of long noncoding RNAs. *Cell*, **136**, 629–641.
3. Breaker,R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol. Cell*, **43**, 867–879.
4. Novikova,I.V., Hennelly,S.P. and Sanbonmatsu,K.Y. (2012) Sizing up long non-coding RNAs: Do lncRNAs have secondary and tertiary structure? *Bioarchitecture*, **2**, 189–199.
5. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
6. Elmlund,D., Le,S.N. and Elmlund,H. (2017) High-resolution cryo-EM: the nuts and bolts. *Curr. Opin. Struct. Biol.*, **46**, 1–6.
7. Frank,J. (2017) Advances in the field of single-particle cryo-electron microscopy over the last decade. *Nat. Protoc.*, **12**, 209–212.
8. Orlov,I., Myasnikov,A.G., Andronov,L., Natchiar,S.K., Khatter,H., Beinsteiner,B., Ménétret,J.-F., Hazemann,I., Mohideen,K., Tazibt,K. *et al.* (2017) The integrative role of cryo electron microscopy in molecular and cellular structural biology. *Biol. Cell*, **109**, 81–93.
9. von Loeffelholz,O., Natchiar,S.K., Djabeur,N., Myasnikov,A.G., Kratzat,H., Ménétret,J.-F., Hazemann,I. and Klaholz,B.P. (2017) Focused classification and refinement in high-resolution cryo-EM structural analysis of ribosome complexes. *Curr. Struct. Biol.*, **46**, 140–148.
10. Newman,A.J. and Nagai,K. (2010) Structural studies of the spliceosome: blind men and an elephant. *Curr. Opin. Struct. Biol.*, **20**, 82–89.
11. Nguyen,T.H.D., Galej,W.P., Fica,S.M., Lin,P.-C., Newman,A.J. and Nagai,K. (2016) CryoEM structures of two spliceosomal complexes: starter and dessert at the spliceosome feast. *Curr. Opin. Struct. Biol.*, **36**, 48–57.
12. Zhao,C. and Pyle,A.M. (2017) Structural Insights into the mechanism of Group II Intron Splicing. *Trends Biochem. Sci.*, **42**, 470–482.
13. Ferré-D'Amaré,A.R. and Doudna,J.A. (2000) Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.*, **295**, 541–556.
14. Ye,J.-D., Tereshko,V., Frederiksen,J.K., Koide,A., Fellouse,F.A., Sidhu,S.S., Koide,S., Kossiakoff,A.A. and Piccirilli,J.A. (2008) Synthetic antibodies for specific recognition and crystallization of structured RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 82–87.
15. Huang,L. and Lilley,D.M.J. (2013) The molecular recognition of kink-turn structure by the L7Ae class of proteins. *RNA*, **19**, 1703–1710.
16. Koldobskaya,Y., Duguid,E.M., Shechner,D.M., Suslov,N.B., Ye,J., Sidhu,S.S., Bartel,D.P., Koide,S., Kossiakoff,A.A. and Piccirilli,J.A. (2011) A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nat. Struct. Mol. Biol.*, **18**, 100–106.
17. Hamma,T. and Ferré-D'Amaré,A.R. (2004) Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure*, **12**, 893–903.
18. Zhang,J. and Ferre-D'Amare,A.R. (2013) Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature*, **500**, 363–366.
19. Zhang,J. and Ferré-D'Amaré,A.R. (2014) New molecular engineering approaches for crystallographic studies of large RNAs. *Curr. Opin. Struc. Biol.*, **26**, 9–15.
20. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
21. Huang,H., Suslov,N.B., Li,N.-S., Shelke,S.A., Evans,M.E., Koldobskaya,Y., Rice,P.A. and Piccirilli,J.A. (2014) A G-quadruplex–containing RNA activates fluorescence in a GFP-like fluorophore. *Nat. Chem. Biol.*, **10**, 686–691.
22. Shao,Y., Huang,H., Qin,D., Li,N.-S., Koide,A., Staley,J.P., Koide,S., Kossiakoff,A.A. and Piccirilli,J.A. (2016) Specific recognition of a single-stranded RNA sequence by a synthetic antibody fragment. *J. Mol. Biol.*, **428**, 4100–4114.
23. Kao,C., Rüdisser,S. and Zheng,M. (2001) A simple and efficient method to transcribe RNAs with reduced 3′ heterogeneity. *Methods*, **23**, 201–205.
24. Hall,B., Arshad,S., Seo,K., Bowman,C., Corley,M., Jhaveri,S.D. and Ellington,A.D. (2010) In vitro selection of RNA aptamers to a protein target by filter immobilization. *Curr. Protoc. Nucleic Acid Chem.*, **40**, 9.3.1–9.3.27.
25. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
26. Bailey,T.L., Boden,M., Buske,F.A., Frith,M., Grant,C.E., Clementi,L., Ren,J., Li,W.W. and Noble,W.S. (2009) MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
27. Goujon,M., McWilliam,H., Li,W., Valentin,F., Squizzato,S., Paern,J. and Lopez,R. (2010) A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.*, **38**, W695–W699.
28. Kunkel,T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 488–492.
29. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
30. Yelton,D.E., Rosok,M.J., Cruz,G., Cosand,W.L., Bajorath,J., Hellström,I., Hellström,K.E., Huse,W.D. and Glaser,S.M. (1995) Affinity maturation of the BR96 anti-carcinoma antibody by codon-based mutagenesis. *J. Immunol.*, **155**, 1994–2004.
31. Low,N.M., Holliger,P. and Winter,G. (1996) Mimicking somatic hypermutation: affinity maturation of antibodies displayed on bacteriophage using a bacterial mutator strain. *J. Mol. Biol.*, **260**, 359–368.
32. Chowdhury,P.S. (2002) Targeting random mutations to hotspots in antibody variable domains for affinity improvement. *Methods Mol. Biol.*, **178**, 269–285.
33. Clackson,T. and Wells,J.A. (1994) In vitro selection from protein and peptide libraries. *Trends Biotechnol.*, **12**, 173–184.
34. Eigenbrot,C., Randal,M., Presta,L., Carter,P. and Kossiakoff,A.A. (1993) X-ray structures of the antigen-binding domains from three variants of humanized anti-p185HER2 antibody 4D5 and comparison with molecular modeling. *J. Mol. Biol.*, **229**, 969–995.
35. Trower,M.K. (1994) Site-directed mutagenesis using a uracil-containing phagemid template. In: Harwood,AJ (ed). *Protocols for Gene Analysis*. Humana Press, Totowa, NJ, pp. 67–77.
36. Johnson,G. and Wu,T.T. (2000) Kabat database and its applications: 30 years after the first variability plot. *Nucleic Acids Res.*, **28**, 214–218.
37. McDonald,R.I., Guilinger,J.P., Mukherji,S., Curtis,E.A., Lee,W.I. and Liu,D.R. (2014) Electrophilic activity-based RNA probes reveal a self-alkylating RNA for RNA labeling. *Nat. Chem. Biol.*, **10**, 1049–1054.
38. Tan,X., Constantin,T.P., Sloane,K.L., Waggoner,A.S., Bruchez,M.P. and Armitage,B.A. (2017) Fluoromodules consisting of a promiscuous RNA aptamer and red or blue fluorogenic cyanine dyes: selection, characterization, and bioimaging. *J. Am. Chem. Soc.* **139**, 9001–9009.
39. Li,S., Hwang,X.Y., Stav,S. and Breaker,R.R. (2016) The yjdF riboswitch candidate regulates gene expression by binding diverse azaaromatic compounds. *RNA*, **22**, 530–541.
40. Spahn,C.M., Kieft,J.S., Grassucci,R.A., Penczek,P.A., Zhou,K., Doudna,J.A. and Frank,J. (2001) Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science*, **291**, 1959–1962.
41. Chapman,E.G., Costantino,D.A., Rabe,J.L., Moon,S.L., Wilusz,J., Nix,J.C. and Kieft,J.S. (2014) The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science*, **344**, 307–310.