TECHNICAL REPORTS

nature structural & molecular biology

A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination

Yelena Koldobskaya[1], Erica M Duguid[2], David M Shechner[3,4,7], Nikolai B Suslov[2], Jingdong Ye[5], Sachdev S Sidhu[6], David P Bartel[3,4], Shohei Koide[2], Anthony A Kossiakoff[2] & Joseph A Piccirilli[1,2]

RNA crystallization and phasing represent major bottlenecks in RNA structure determination. Seeking to exploit antibody fragments as RNA crystallization chaperones, we have used an arginine-enriched synthetic Fab library displayed on phage to obtain Fabs against the class I ligase ribozyme. We solved the structure of a Fab–ligase complex at 3.1-Å resolution using molecular replacement with Fab coordinates, confirming the ribozyme architecture and revealing the chaperone's role in RNA recognition and crystal contacts. The epitope resides in the GAAACAC sequence that caps the P5 helix, and this sequence retains high-affinity Fab binding within the context of other structured RNAs. This portable epitope provides a new RNA crystallization chaperone system that easily can be screened in parallel to the U1A RNA-binding protein, with the advantages of a smaller loop and Fabs' high molecular weight, large surface area and phasing power.

The advent of next-generation sequencing has triggered an explosion in discoveries of noncoding RNAs (ncRNAs), and the known repertoire of RNA functions is rapidly expanding[1,2]. Beyond the classic roles of ncRNA in protein synthesis, ncRNA engages in a wide range of other functions, including control of transcription, gene expression and embryonic development[3–5]. Recent analyses of cellular transcriptomes have revealed that, collectively across different cells, more than 90% of the eukaryotic genome (in humans, mice and other species) is transcribed, giving rise to vast numbers of RNA transcripts[6,7]. As a relatively small fraction of these transcripts code for proteins, there probably exists an expansive landscape of as-yet-undiscovered ncRNAs.

To perform their biological roles, many ncRNAs and noncoding regions of mRNAs adopt complex three-dimensional architectures. Defining these structures represents an important step toward understanding ncRNA function, and in recent years some general principles that govern RNA architecture have emerged from structures of ribozymes, riboswitches, ribosomes and other complexes of RNA and ribonucleoprotein[8–13]. Despite these advances, the pace of RNA structure determination has lagged behind that of protein structure determination: compared to nearly 57,000 unique X-ray protein structures in the Protein Data Bank, there are fewer than 2,100 experimentally determined RNA structures (http://www.rcsb.org/pdb/statistics/holdings.do). This difference reflects, in part, challenges associated with RNA crystallization. Whereas proteins have chemically diverse features that facilitate crystal-lattice formation, RNA surfaces have less chemical diversity and contain mutually repulsive phosphate groups that render lattice interactions less favorable and potentially irregular[14]. Additionally, RNAs frequently have flexible regions that enable sampling of alternative conformations or have a tendency to misfold, leading to conformational heterogeneity[15–18]. Confounding matters further, phasing of RNA crystals involves complex methods, in contrast to the well-established selenium-based phasing of protein crystals[15,19,20].

The crystallization bottleneck has led researchers to develop creative but laborious approaches to circumvent these problems and facilitate crystallization. Such methods include identifying well-folded RNA variants by screening phylogenetically related species, isolating robust crystallization targets through *in vitro* selection, rational engineering of RNA constructs, eliminating nonessential sequences to attenuate conformational dynamics, phasing by molecular replacement using idealized RNA domains, and facilitating intermolecular contacts by incorporating GNRA tetraloops (where N is any nucleotide and R is any purine)[15,16,18,21,22]. Another approach, one that supports the concept underlying the work described here, entails the use of the U1A RNA-binding protein as a portable crystallization chaperone. This strategy involves replacing a nonessential region of an RNA with the ten-nucleotide sequence recognized by the U1A protein and crystallizing the RNA in complex with the U1A protein[23]. Despite the well-documented success of the chaperone approach for protein crystallography over the past decade[24–26], U1A remains the only general chaperone available for RNA crystallization.

In protein crystallography, Fab and scFv antibody fragments that bind their antigens with high affinity and specificity have served as crystallization chaperones, enabling successful structure determination

[1]Department of Chemistry, University of Chicago, Chicago, Illinois, USA. [2]Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois, USA. [3]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. [4]Department of Biology and Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [5]Department of Chemistry, University of Central Florida, Orlando, Florida, USA. [6]Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada. [7]Current address: Broad Institute and Harvard University Department of Stem Cell and Regenerative Biology, Cambridge, Massachusetts, USA. Correspondence should be addressed to J.A.P. (jpicciri@uchicago.edu) or A.A.K. (koss@bsd.uchicago.edu).

Received 12 May; accepted 1 October; published online 12 December 2010; doi:10.1038/nsmb.1945

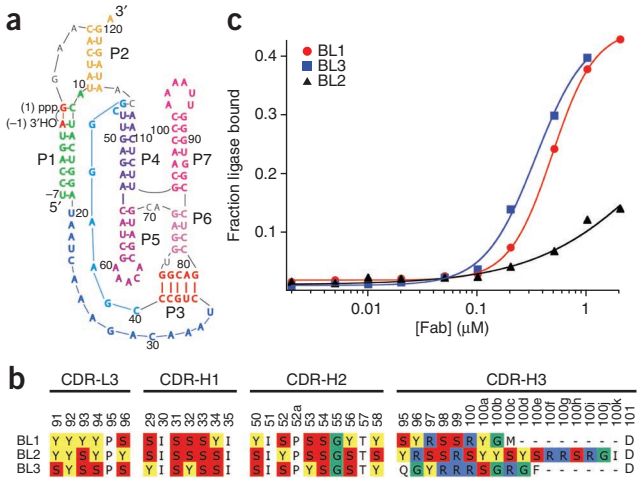

Figure 1 Selection of class I ligase–binding Fabs from the YSGR Fab superlibrary. (**a**) Secondary structure of the ligase ribozyme and its substrate[36]. The arrow indicates the self-ligation reaction, in which the substrate's 3′-terminal hydroxyl (nucleotide A(–1), red) attacks the α-phosphate of the ribozyme's 5′-terminal triphosphate (nucleotide G1, red). (**b**) CDR sequences of ligase-binding Fabs selected from the YSGR superlibrary. Red, serine; yellow, tyrosine; green, glycine; blue, arginine. Residues are numbered according to the Kabat system[53]. (**c**) Nitrocellulose binding assay reveals low-micromolar K_d values for Fab–class I ligase binding. BL1 K_d = 478 nM; BL2 K_d > 2,000 nM; BL3 K_d = 338 nM. Reactions were carried out as described in Online Methods.

of several 'difficult' protein targets[24,25,27–29]. Antibody chaperones seem particularly well suited for overcoming some of the challenges inherent to RNA crystallization[25]. With a higher molecular weight (50 kDa) than the U1A protein (11 kDa), Fab chaperones provide more surface area for crystal contacts, and their β-rich architecture is predisposed to making effective crystal contacts[29]. This, in turn, could enhance the probability of crystallization and therefore reduce the number of RNA constructs screened during crystallization trials. Moreover, the Fab scaffold can serve as the search model for molecular replacement and provide initial phase information, simplifying the process of solving the structure of the RNA target[25].

In recent years, the development of natural and synthetic immune repertoires and selection methodologies has enabled antibody production without host immunization[30–33]. Using a phage platform to display libraries of synthetic Fabs, we have recently established a general approach for obtaining Fab antibody fragments that bind to RNA. We first targeted an independently folding domain from the group I intron and obtained antibodies that recognize the RNA tertiary structure with high affinity and specificity[34]. We used these antibodies successfully to crystallize the target RNA and to solve the structure of the P4-P6 RNA domain at 1.95-Å resolution.

Here we target the class I ligase, an artificial ribozyme, originally isolated from a random pool of RNA sequences, that efficiently catalyzes a reaction analogous to that of RNA-dependent RNA polymerases[35,36]. Several rounds of selection using an arginine-enriched Fab library, followed by affinity maturation using error-prone PCR, generated Fabs that bind the ligase with 30- to 50-nM affinities. We solved the crystal structure of the Fab–ligase complex at 3.1-Å resolution using molecular replacement with Fab (chaperone) coordinates. The structure reveals the global architecture of the ligase, the molecular interactions underpinning Fab recognition, and a substantial Fab contribution to crystal lattice contacts. Moreover, we show that the crucial antigenic elements recognized by the ligase-binding Fab reside in a motif that retains high-affinity Fab binding in the context of other structured RNAs, thereby providing a portable epitope tag with potential for widespread use for RNA affinity capture or chaperone-assisted crystallography.

RESULTS

Phage-display selections against the ligase ribozyme

We have previously used a reduced-codon[37,38] synthetic phage antibody library (the YSG Fab library) to obtain Fabs that recognize the tertiary structure of the P4-P6 RNA domain in a Mg^{2+}-dependent manner[34]. Subsequent selections using the YSG library produced no RNA-binding clones against a variety of RNA targets. Considering the enrichment of cationic lysine and arginine residues in the P4-P6–binding Fabs[34], we deployed another library already enriched in lysine and arginine residues. This library (referred to hereafter as the YSGR superlibrary) consisted of a mixture of four reduced-codon libraries designated YSG, YSR, YSGR and YSGRZ to reflect the codon bias at selected positions. Complementarity-determining regions (CDRs) L3, H1 and H2 contained equimolar mixtures of tyrosine and serine. For CDR-H3, the YSG and YSR sublibraries had the same design as reported previously[31] with the following components: YSG, 50% tyrosine, 25% serine, 25% glycine; YSR, 25% tyrosine, 50% serine, 25% arginine. H3 CDRs in YSGR and YSGRZ sublibraries were designed as follows: YSGR, 38% tyrosine, 25% serine, 25% glycine, 12% arginine; YSGRZ, 19% tyrosine, 25% serine, 25% glycine, 12% arginine, 19% Z, where Z represents an equimolar mixture of all amino acids except tyrosine, serine, glycine, arginine and cysteine.

We used the product form of an improved variant of the class I ligase ribozyme[36] (**Fig. 1a**) as the target for phage-display selections. Two YSGR superlibrary selections yielded three unique Fab clones, BL1, BL2 and BL3 (**Fig. 1b**), which tested positive by phage ELISA for binding to the target RNA. A parallel selection with the YSG library yielded no positive clones. Fabs BL1, BL2 and BL3 were expressed as soluble proteins and analyzed for binding to the ligase product with a nitrocellulose filter binding assay, revealing K_d values in the range of 300–500 nM for BL1 and BL3, and >2,000 nM for BL2 (**Fig. 1c**). Ribozyme assays in the presence of Fab showed that BL3 had no effect on ligase activity (**Supplementary Fig. 1a**), whereas BL1 inhibited ligase activity in a concentration-dependent manner (**Supplementary Fig. 1b**; see **Supplementary Fig. 2a** for the BL1-ligase footprinting pattern). We therefore chose Fab BL3 for further analysis and crystallization screens. As most antibody chaperones successfully used for crystallization bind their targets with K_d < 250 nM[26,34,39,40], we pursued affinity maturation of BL3 (K_d = 338 nM) before moving on to crystallization trials.

Affinity maturation of class I ligase–binding Fabs

We chose error-prone PCR as a technique to introduce random mutations throughout the variable domains of BL3, mainly because this method has minimal bias toward any specific class of substitutions and allows mutations in Fab CDRs and in scaffold regions. We generated error-prone PCR Fab libraries with mutated heavy-chain and light-chain variable domains from the Fab-BL3 parent sequence. Subsequent phage-display selections using these error-prone PCR phage libraries consisted of three rounds of equilibrium phage selection in which we varied the concentration of target RNA (class I ligase product) from 2.5 to 0.025 nM. After each round of selection, phages that eluted from the selection with the lowest target concentration with enrichment ratio >10 (phages eluted from the target selection/phages eluted from a blank selection) relative to a blank selection were either amplified for the next round of selection

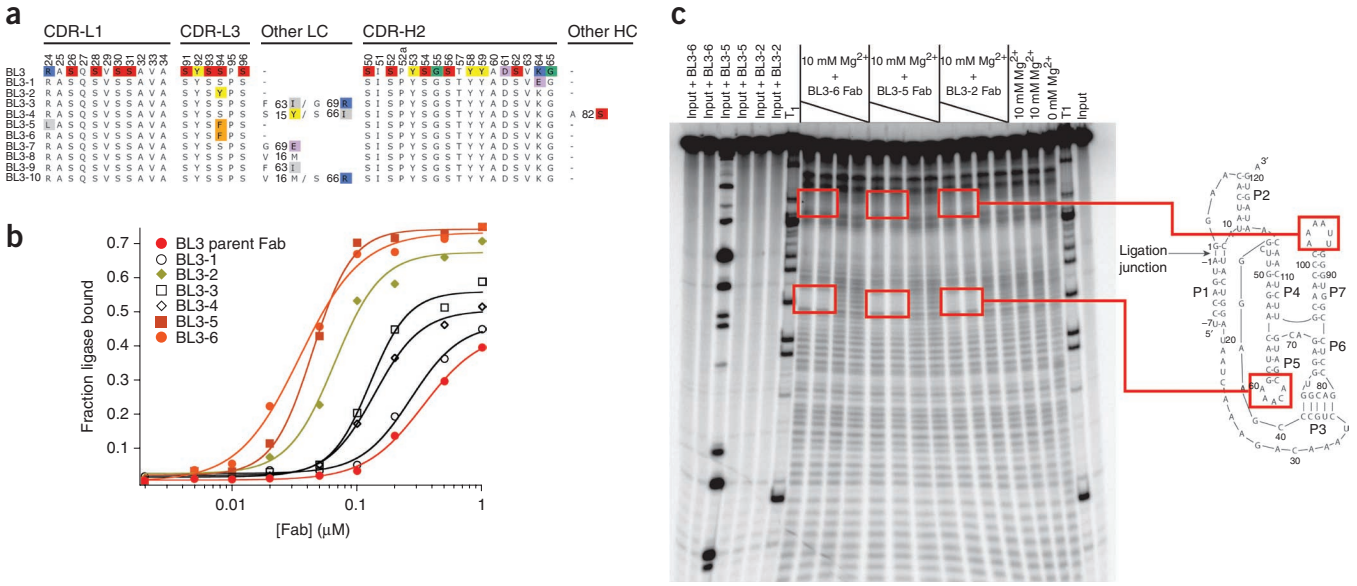

Figure 2 Affinity maturation of class I ligase–binding Fab-BL3 by error-prone PCR. (**a**) CDR sequences of BL3-derived affinity-matured Fabs selected from error-prone PCR Fab libraries. Red, serine; yellow, tyrosine; green, glycine; blue, arginine or lysine; lavender, aspartate or glutamate; gray, leucine or isoleucine. (**b**) BL3-derived Fabs bind to the class I ligase product with nanomolar K_d values. Parent BL3 Fab, K_d = 338 nM; BL3-1, K_d = 270 nM; BL3-2, K_d = 66 nM; BL3-3, K_d = 127 nM; BL3-4, K_d = 138 nM; BL3-5, K_d = 44 nM; BL3-6, K_d = 35 nM. Nitrocellulose filter binding assay conditions were identical to those for **Figure 1c**. The variation of endpoint (fraction RNA bound) values observed for BL3 Fab binding to the ligase product, ranging from 0.4 to 0.75, probably reflects the faster dissociation of weaker-binding Fabs in the context of the nitrocellulose filter binding assay. Accordingly, we used this assay only as a convenient, qualitative method to assay Fab-RNA affinity. As an independent method to assess BL3 Fab K_d values, we used hydroxyl-radical footprinting (**Supplementary Fig. 2b**, BL3-6 Fab). (**c**) Hydroxyl-radical footprinting of self-labeled class I ligase in the presence of 5 nM to 1 mM BL3 affinity-matured Fabs BL3-2, BL3-5 and BL3-6. Increasing Fab concentrations are denoted by black triangles; red boxes show P7 and P5 Fab binding sites and the corresponding areas of protection. T1, samples treated with ribonuclease T1. Ligase ribozyme secondary structure (see **Fig. 1a**) is shown at right.

(round 2) or sequenced (round 3). Sequencing analysis revealed ten unique Fab clones after the third round of selection, including five CDR mutations and seven scaffold (non-CDR) mutations (**Fig. 2a**). All BL3-derived Fabs tested (six of ten) showed some improvement in binding relative to the parent Fab-BL3 (**Fig. 2b**). Scaffold mutations, represented by Fabs BL3-3 and BL3-4, produced moderate (two- to three-fold) improvements in binding to the class I ligase. Mutations S94Y and S94F in CDR-L3 (BL3-2, BL3-5 and BL3-6 Fabs) improved binding five- to ten-fold, resulting in K_d values of 30–70 nM, which we viewed as sufficient for Fab–RNA crystallization screening.

Mapping of Fab–class I ligase binding sites

Hydroxyl-radical footprinting of the class I ligase product in complex with affinity-matured BL3 Fabs (**Fig. 2c**) revealed two regions of Fab-induced protection: one corresponding to the ligase P5 loop (nucleotides (nt) 59–63, GAAAC) and a portion of the P5 stem (nt 55–56), and the other corresponding to the P7 hairpin loop (nt 87–93; **Fig. 2c**). The Fab concentrations at which these protected regions appeared confirmed that the K_d was ~75 nM, as observed by filter binding (**Supplementary Fig. 2b**). Additionally, the Fab–ligase complex retained the Mg^{2+}-dependent protected regions observed in the absence of Fab, indicating that the ligase undergoes no substantial change in global conformation when bound to the Fab (**Fig. 2c**). The same range of Fab BL3-6 concentrations (**Supplementary Fig. 1b**) had no detectable effect on ligase activity, consistent with the absence of any large distortion of the active ribozyme structure.

Crystal structure of the Fab–ligase complex

To test the usefulness of Fab BL3-6 as an RNA crystallization chaperone, we set up crystallization trials for the BL3-6–class I ligase product complex. We found that BL3-6 Fab was a successful crystallization chaperone that facilitated both crystallization and phasing of the Fab–RNA structure. Crystals of the Fab BL3-6–ligase product complex were grown by the hanging-drop vapor diffusion method. Small cube-shaped crystals were observed within 1 week at 4 °C, and provided diffraction data to 3.1 Å. **Table 1**, the **Supplementary Methods** and **Supplementary Figure 3** provide information regarding data collection, phasing and refinement.

The Fab–ligase complex crystallized in the $C222_1$ space group, with two complexes per asymmetric unit. A parallel effort to solve the structure of the ligase, initiated several years earlier, replaced loop 5 (L5) with the U1A-binding loop and yielded crystals of U1A–ligase complex that diffracted to 3.0 Å but were much more challenging to grow and phase[41]. For the Fab–ligase cocrystals, we used molecular replacement with Fab coordinates to readily obtain initial phases that provided sufficient information to resolve the RNA backbone (128 nt). By comparison, phasing the U1A–ligase cocrystals involved substantially more effort, with several established methods attempted before one was found that yielded useful phasing information (D.M.S. and D.P.B., unpublished data). The ligase structures in the Fab and U1A complexes, solved independently in separate labs, show excellent agreement, with an all-atom r.m.s. deviation of 1.26 Å[41]. These observations, together with the observation that the ligase remains fully active with Fab bound (**Supplementary Fig. 1a**), demonstrate that Fab-ligase interactions do not substantially distort the ligase RNA backbone and that Fab BL3-6 serves as an effective crystal chaperone.

As was recently reported[41], the ligase product in both structures forms a tripod-like arrangement of three coaxially stacked domains—P1-P2, P4-P5 and P3-P6-P7, all roughly equal in length—that are joined by J3/4 and the P4-P5-P6-P7 four-helix junction (**Fig. 3**).

Table 1 Data collection and refinement statistics

	Fab BL3-6–ligase
Data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	207.5, 206.5, 135.9
Resolution (Å)	50.0–3.1 (3.15–3.10)
R_{sym}	0.104 (0.557)
I / σI	34.3/2.5 (5.2/3.2)
Completeness (%)	98.2 (79.6)
Redundancy	6.7 (3.4)
Refinement	
Resolution (Å)	50.0–3.1 (3.18–3.10)
No. reflections	49,591
R_{work} / R_{free}	20.8/22.5 (31.7/34.5)
No. atoms	12,299
Protein	6,604
RNA	5,554
Ions	141
B-factors	68.09
Protein	63.25
RNA	83.57
Ions	71.20
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.323

The number of crystals is one. Values in parentheses are for highest-resolution shell.

Two single-stranded regions—J1/3, which spans the length of the ribozyme and docks into the minor grooves of P1 and P6, and J3/4, which connects the P3 pseudoknot to P4—support the overall tripod architecture. Here we focus on crystallographic features relevant to chaperone activity: the Fab-RNA interface and Fab-RNA packing in the crystal lattice.

Structural basis for Fab–RNA antigen recognition

Residues from five CDRs (all except CDR-L1), as well as several scaffold residues, contribute to the Fab-RNA interface (**Fig. 4** and **Supplementary Fig. 4a**). Consistent with the results from hydroxyl-radical footprinting, the Fab recognizes two distinct epitope regions of the ligase tripod, the P7 backbone and the P5 loop, accounting for 29.7% and 70.3% of the buried surface area, respectively. In the smaller Fab-P7 interface, backbone phosphates of P7 interact with nonrandomized residues in CDR-L2 and the light-chain scaffold—specifically, Arg19, Gly65, Ser66, Arg67, Ser68, Gly69 and Ser77, which account for the majority (83.5%) of the 326-Å^2 surface burial in the P7 paratope. Several mutations in this region also emerged during affinity maturation (Fabs BL3-3, BL3-4, BL3-7 and BL3-10, **Fig. 2a**) that may contribute favorably toward antigen binding (see **Supplementary Note** for analysis of the amino acid composition of the Fab paratope).

The more extensive P5 interface involves heavy-chain CDRs H1, H2 and H3, which contribute tyrosine, serine and arginine residues, and light-chain CDR-L3, which contributes the Phe95 residue selected during affinity maturation. Within the P5 paratope, CDR-H3 arginines (105, 106 and 110) contact RNA nucleotide A62 and the G59-C65 terminal base pair (**Fig. 4b**). Phe95 stacks between Tyr62 (CDR-H2) and the A62 base (**Fig. 4c**), accounting for the improved binding affinity of the Fab. Ser58 and Ser60 (CDR-H2) both contact C63 (**Fig. 4d**), consistent with the observed importance of this nucleotide in Fab binding (see below). Finally, Tyr57 (CDR-H2) stacks with A61, which in turn stacks with G59 of the terminal base pair (**Fig. 4e**). Several 2′-hydroxyl groups within the AAACA loop reside within 3.5–4.5 Å of residues located in CDR-H2 and CDR-H3 (Tyr62 of CDR-H2 and A62; Tyr62 of CDR-H2 and C63; Tyr57 of CDR-H2 and A60; Arg105 of CDR-H3 and A61; **Supplementary Fig. 4b**) and may engage in direct or water-mediated hydrogen bonds, consistent with the observation below that DNA does not bind to the BL3 Fab. Together, the CDR residues form a pocket around the ligase L5 loop that imparts specific recognition of the AAACA sequence and the closing G-C base pair (**Fig. 4a**).

The overall structure of Fab BL3-6 is similar to that of the parent Fab 4D5 variant that served as the scaffold for the antibody library (**Supplementary Fig. 4c**; r.m.s. deviation of 0.51 Å)[42]. Although the elbow angles differ slightly owing to distinct crystal packing, the main differences reside in CDR loops L3, H1, H2 and H3. These loops harbor the diversified positions and form most of the interactions with the ligase antigen, underscoring the conformational plasticity of the hypervariable regions[43]. Fab BL3-6 binds to the ligase with a relatively large buried surface area (1,140 Å^2 on the Fab side alone), substantially greater than that of most Fab-protein antigen interfaces (777 ± 135 Å^2)[44] but similar to that observed previously for Fab2 binding to P4-P6 RNA domain[34]. Despite similar degrees of surface burial, Fab2 and Fab BL3-6 paratopes have unique topological features that mediate distinct modes of RNA recognition. Fab2 has a relatively flat topology that recognizes global features of P4-P6 through interactions with the duplex RNA regions brought together by the RNA fold. In contrast, Fab BL3-6 forms a binding cleft for recognition of a local feature of the ligase (P5) predominantly through interactions with single-stranded (loop) nucleotides. These distinct modes of recognition show that Fabs can provide a highly versatile scaffold for binding to a wide range of RNAs. Fab-ligase recognition also shares several features with

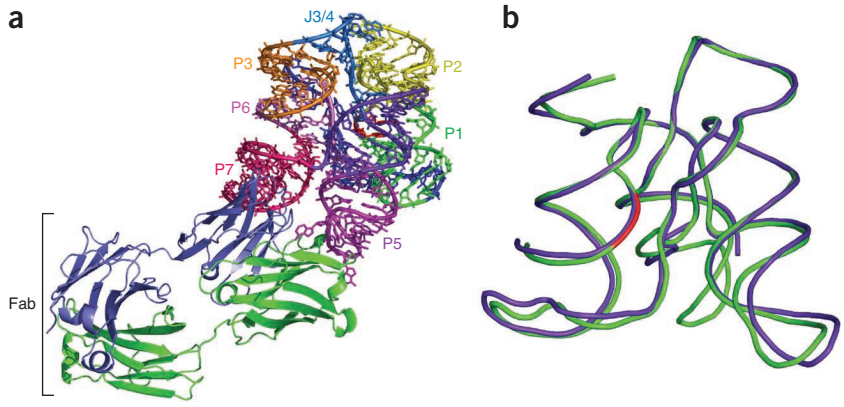

Figure 3 Crystal structure of the Fab BL3-6–ligase complex. (**a**) Crystal structure of the BL3-6 Fab–ligase complex at 3.1 Å resolution. (**b**) Overlay of the Fab–ligase structure (green) with the U1A–ligase structure[41] (blue); red indicates the ligation junction. Nucleotides in the U1A-binding loop have been omitted. All-atom r.m.s. deviation is 1.26 Å, omitting the residues in L5.

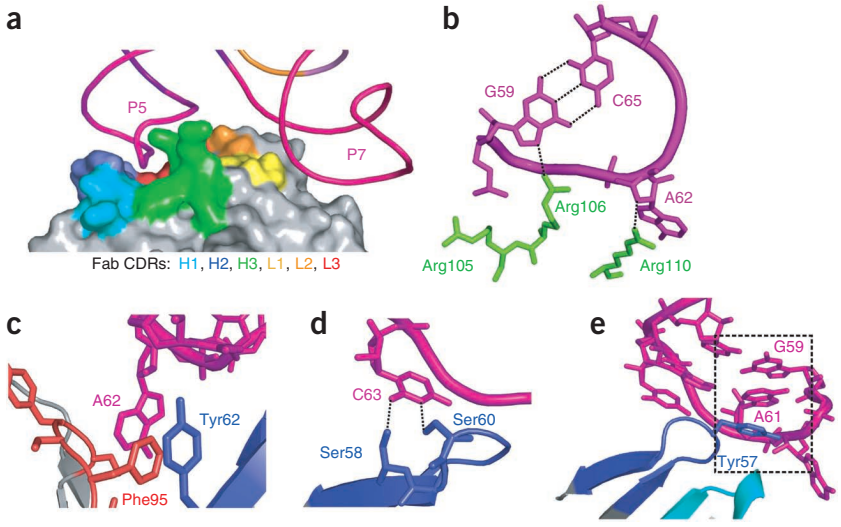

Figure 4 Details of Fab-P5 loop interactions. (**a**) Contacts to P5 loop are formed by Fab CDRs L3, H1, H2 and H3. (**b**) CDR-H3 arginine residues form contacts to A62 and to the G59-C65 terminal base pair. (**c**) Phe95 (CDR-L3) stacks with Tyr62 (CDR-H2) and the A62 base. (**d**) C63 forms contacts to Ser58 and Ser60 (CDR-H2). (**e**) A61 and Tyr57 (CDR-H2) stack with G59 from the terminal base pair.

single-stranded RNA–binding proteins[45,46] (see the **Supplementary Note** for comparison of similarities and differences).

Features of Fab–ligase crystal packing

A view of the Fab–ligase crystal lattice reveals a prominent role for the BL3 Fab in mediating Fab–RNA crystal contacts. A projection in the *a*-*b* plane (**Fig. 5a**) shows a checkerboard pattern formed by Fab dimers (green), in which alternate squares are filled by clusters of RNA (blue) with extensive RNA-RNA contacts. Large channels ~100 Å × 100 Å run perpendicular to the *a*-*b* plane. The *b*-*c* projection (**Fig. 5b**) again shows rows of Fabs alternating with rows of RNA. With approximately equal molecular weights (~48 kDa), Fab BL3-6 and the ligase ribozyme have approximately equal surface areas: 19,740 Å^2 (average of the two Fabs in the asymmetric unit) and 20,654 Å^2 (average of the two ligase molecules in the asymmetric unit). Analysis of buried surface area with the CCP4 PISA web server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) suggests that Fab-RNA contacts account for the majority of surface burial in the structure (65%, or 3,818.5 Å^2). RNA-RNA dimerization via P2-to-J3/4 base triples (**Supplementary Fig. 5a**–**c**) provides the next-largest contribution to buried surface area, 21.3% (1,251.4 Å^2), followed by limited Fab-Fab contacts (13.8% of buried surface area, or 813.6 Å^2). Including Fab-RNA contacts of the original complex, Fab-mediated crystal contacts account for 78.7% of buried surface area in the structure, suggesting that Fab BL3-6 serves as an effective crystallization chaperone (see **Supplementary Note** for further discussion of crystal packing).

Antigen portability of the P5 hairpin loop

Because the BL3 Fab–ligase epitope appeared to reside within two small hairpin loops of the ligase (the P5 hairpin loop and the P7 hairpin loop), we tested affinity-matured BL3 Fabs for binding directly to isolated 25-mer oligonucleotide hairpins corresponding to P5 and P7 (**Fig. 6a**). BL3-derived Fabs bound the P5 hairpin with affinities similar to those for the full-length ligase (**Fig. 6b**), indicating that ligase tertiary structure is not necessary for Fab recognition. BL3 Fabs showed no detectable binding to the isolated P7 hairpin RNA. Considering the reported proximity of P7 and P5 in the ligase tertiary structure[47], these results suggest that high-affinity binding to P5 might cause the Fab to also protect the neighboring P7 from hydroxyl radicals.

We next tested Fab BL3-6 binding to a series of mutant P5 oligonucleotides using filter-binding assays. Alteration of the loop sequence—in variants carrying a UUCG loop (not shown), a C63 deletion, a C63A substitution or a A61C C63A double substitution (**Fig. 6c**)—abolished detectable Fab binding, consistent with crystallographically inferred interactions between the C63 nucleobase and CDR-H2 (Ser58 and Ser60). In addition, no binding activity was detected for the linear AAACA RNA pentanucleotide alone, for an unstructured 15-mer RNA containing the AAACA sequence, for a double-stranded RNA containing the AAACA sequence in one of the strands, or for a DNA oligonucleotide corresponding to the P5 sequence (**Fig. 6d**). In contrast to the highly deleterious loop mutations, mutations in the stem had minimal effects on binding (**Fig. 6c**), with the exception of the closing G-C base pair, where substitution with A-U or C-G pairs resulted in Fab-oligonucleotide K_d values three-fold and five-fold higher, respectively, than that of the original oligonucleotide with the G-C closing base pair. The interaction between Arg106 and N7 of G59 (**Fig. 4b**) may engender the Fab's preference for the G-C base pair over C-G. We conclude that affinity-matured BL3 Fabs recognize the GAAACAC sequence in a loop conformation with high affinity and high sequence specificity.

To test the potential of the BL3 Fab antigen to serve as a tag for antibody recognition in the context of other RNAs, we engineered the GAAACAC sequence into the P4-P6 domain and the Varkud satellite (VS)[48] ribozyme. For P4-P6 we engineered the antigen into the P6b stem-loop (**Supplementary Fig. 6a**). As the closing base pair was already G-C, we only replaced the P6b loop (AUCUU) with the AAACA sequence, and did not introduce new base pairs. For the VS constructs, we engineered the antigen into stem-loops IV and VI, again replacing only the loop and the closing base pair and introducing no additional base pairs. The BL3-6 Fab bound these chimeric RNAs with affinities comparable to those for the isolated P5 hairpin and the full-length class I ligase (**Figs. 2** and **6**, and **Supplementary Fig. 6b**). These results demonstrate the portability of the GAAACAC antigen for general use as a crystallization chaperone or affinity tag.

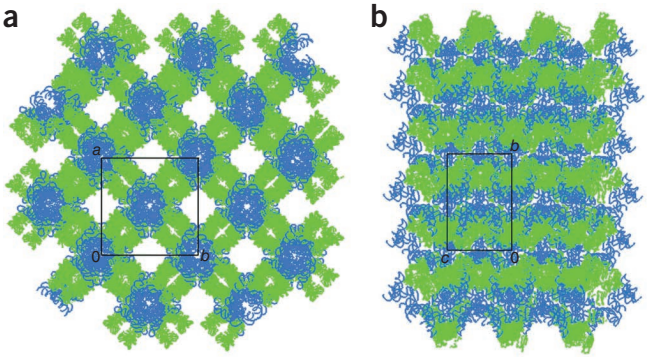

Figure 5 Fab–ligase crystal packing. (**a**) Crystal packing along the *a*-*b* plane. (**b**) Crystal packing along the *b*-*c* plane. Green, Fab; blue, ligase.

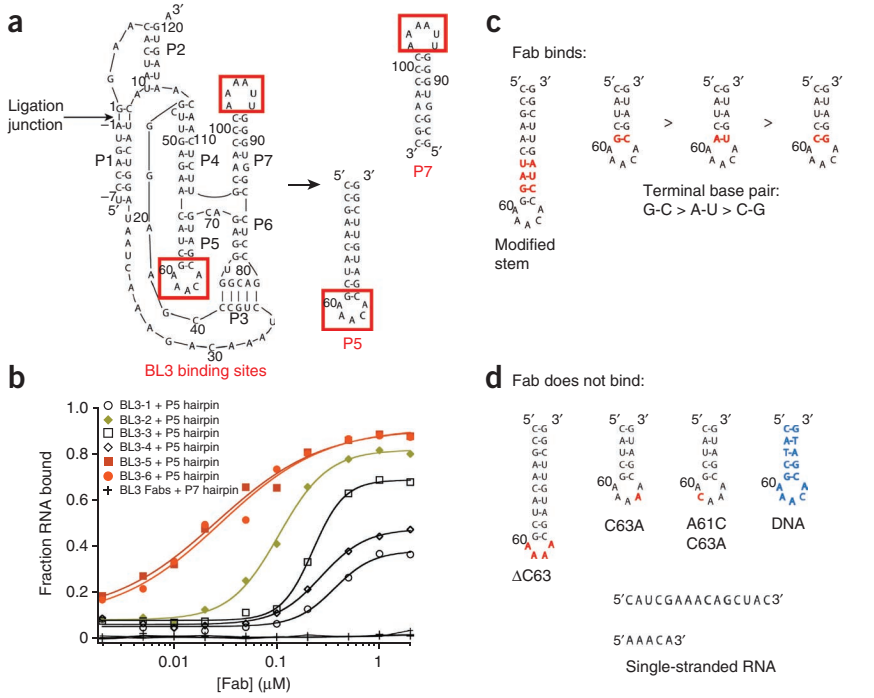

Figure 6 Analysis of the Fab–ligase P5 epitope. (**a**) Design of P5 and P7 hairpin oligos to test for Fab binding to isolated RNA hairpin loops. Red boxes indicate Fab–class I ligase binding sites. (**b**) BL3 Fabs bind to the P5 hairpin-loop RNA oligo but not to the P7 oligo. Binding curves show results for the P5 hairpin incubated with BL3-1 (K_d = 360 nM), BL3-2 (K_d = 105 nM), BL3-3 (K_d = 230 nM), BL3-4 (K_d = 270 nM), BL3-5 (K_d = 26 nM) and BL3-6 (K_d = 28 nM). Also plotted are results for the P7 hairpin incubated with BL3 Fabs. Nitrocellulose filter binding assay conditions were identical to those for **Figure 1c**. (**c**) P5 RNA oligonucleotides that bind the BL3-6 Fab. (**d**) P5 RNA and DNA oligonucleotides that do not bind the BL3-6 Fab.

DISCUSSION

Our previous work has proven the concept of using Fabs as crystallization chaperones for RNA. The work has shown that the phage-display technology can be used to obtain an anti-RNA Fab and that the Fab can be used as a chaperone for P4-P6 crystallization (see **Supplementary Note** for a timeline of the process). However, the library previously used in the P4-P6 selections had limited capacity to produce binding clones against other RNA targets. Here we have identified a library (the YSGR superlibrary) that performs much more robustly against a variety of RNA targets and, when used in combination with affinity maturation, markedly enhances the potential for obtaining high-affinity Fabs to RNA targets. Additionally, we show here that Fab BL3-6, identified in selections against the class I ligase ribozyme, serves as a successful RNA crystallization chaperone that facilitates both the crystallization and phasing. Notably, Fab BL3-6 recognizes its GAAACA RNA loop within a variety of RNA contexts, providing a transplantable module for RNA crystallization and allowing the use of the Fab as a general antibody crystallization chaperone without the need to generate Fabs using artificial libraries and phage display.

In recent years, it has become routine to incorporate U1A-binding loops into RNA constructs targeted for crystallization[23]. Fab BL3-6 and the corresponding RNA loop provide an additional RNA crystallization chaperone and RNA tag system that can easily be screened in parallel to U1A, with the added advantages of the smaller size of the GAAACAC loop and the high molecular weight, large surface area and increased phasing power provided by Fabs. Future structure-determination projects using the Fab-GAACAC system can begin the process by using the Fab-P5 stem-loop coordinates directly as a search model for molecular replacement, which should facilitate the placement and rebuilding of the RNA backbone coordinates. We also note that Fab BL3-6 is a strongly expressed protein (with >3 mg protein isolated from 1 l *Escherichia coli* culture) that can easily be purified RNase-free for use in RNA crystallization. We anticipate that the ability of the Fab to recognize the AAACA loop in a variety of contexts—class I ligase, isolated hairpin oligonucleotides, and within the P4-P6 domain and the VS ribozyme—will enable the use of the Fab-loop system as a simple RNA crystallization tag or as an epitope tag orthogonal to the U1A, MS2 and PP7 tags currently in use for RNA affinity purification and visualization[49–52].

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Protein Data Bank: Coordinates for the Fab BL3-6–ligase complex have been deposited with accession code 3IVK.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank members of the Piccirilli, Koide and Kossiakoff laboratories for assistance with phage display and for insightful comments; D. Lilley and T. Wilson for helpful discussions regarding the VS ribozyme; and V. Torbeev, I. Dementieva, P. Rice and V. Tereshko for assistance with crystallography. This work was funded by Howard Hughes Medical Institute (J.A.P.), US National Institute of General Medical Sciences Medical Scientist National Research Service Award no. 5 T32 GM07281 (Y.K.) and US National Institutes of Health grants R01-GM72688 and U54-GM74946 (to S.K. and A.A.K.) and R01-GM61835 (to D.P.B.). Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (grant 085P1000817).

AUTHOR CONTRIBUTIONS

All authors designed research; Y.K., E.M.D., D.M.S. and N.B.S. performed experiments; Y.K., E.M.D., D.M.S., S.K., A.A.K. and J.A.P. analyzed data; Y.K., E.M.D., D.M.S., N.B.S., D.P.B., S.K., A.A.K. and J.A.P wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Eddy, S.R. Noncoding RNA genes. *Curr. Opin. Genet. Dev.* **9**, 695–699 (1999).
2. Taft, R.J., Pheasant, M. & Mattick, J.S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288–299 (2007).
3. Mattick, J.S. The genetic signatures of noncoding RNAs. *PLoS Genet.* **5**, e1000459 (2009).
4. Cooper, T.A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777–793 (2009).
5. Hogg, J.R. & Collins, K. Structured non-coding RNAs and the RNP Renaissance. *Curr. Opin. Chem. Biol.* **12**, 684–689 (2008).
6. Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488 (2007).

7. Mattick, J.S. & Makunin, I.V. Non-coding RNA. *Hum. Mol. Genet.* **15** (Spec No 1), R17–R29 (2006).
8. Montange, R.K. & Batey, R.T. Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys* **37**, 117–133 (2008).
9. Serganov, A. & Patel, D.J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* **8**, 776–790 (2007).
10. Fedor, M.J. Comparative enzymology and structural biology of RNA self-cleavage. *Ann. Rev. Biophys.* **38**, 271–299 (2009).
11. Hoogstraten, C.G. & Sumita, M. Structure-function relationships in RNA and RNP enzymes: recent advances. *Biopolymers* **87**, 317–328 (2007).
12. Jinek, M. & Doudna, J.A. A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**, 405–412 (2009).
13. Reichow, S.L., Hamma, T., Ferré-D'Amaré, A.R. & Varani, G. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res.* **35**, 1452–1464 (2007).
14. Doudna, J.A. & Cate, J.H. RNA structure: crystal clear? *Curr. Opin. Struct. Biol.* **7**, 310–316 (1997).
15. Golden, B.L. & Kundrot, C.E. RNA crystallization. *J. Struct. Biol.* **142**, 98–107 (2003).
16. Edwards, A.L., Garst, A.D. & Batey, R.T. Determining structures of RNA aptamers and riboswitches by X-ray crystallography. *Methods Mol. Biol.* **535**, 135–163 (2009).
17. Ke, A. & Doudna, J.A. Crystallization of RNA and RNA-protein complexes. *Methods* **34**, 408–414 (2004).
18. Ferré-D'Amaré, A.R., Zhou, K. & Doudna, J.A. A general module for RNA crystallization. *J. Mol. Biol.* **279**, 621–631 (1998).
19. Carrasco, N., Buzin, Y., Tyson, E., Halpert, E. & Huang, Z. Selenium derivatization and crystallization of DNA and RNA oligonucleotides for X-ray crystallography using multiple anomalous dispersion. *Nucleic Acids Res.* **32**, 1638–1646 (2004).
20. Höbartner, C. *et al.* Syntheses of RNAs with up to 100 nucleotides containing site-specific 2′-methylseleno labels for use in X-ray crystallography. *J. Am. Chem. Soc.* **127**, 12035–12045 (2005).
21. Keel, A.Y., Rambo, R.P., Batey, R.T. & Kieft, J.S. A general strategy to solve the phase problem in RNA crystallography. *Structure* **15**, 761–772 (2007).
22. Robertson, M.P. & Scott, W.G. A general method for phasing novel complex RNA crystal structures without heavy-atom derivatives. *Acta Crystallogr. D Biol. Crystallogr.* **D64**, 738–744 (2008).
23. Ferré-D'Amaré, A.R. & Doudna, J.A. 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 (2000).
24. Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature* **376**, 660–669 (2002).
25. Koide, S. Engineering of recombinant crystallization chaperones. *Curr. Opin. Struct. Biol.* **19**, 449–457 (2009).
26. Uysal, S. *et al.* Crystal structure of full-length KcsA in its closed conformation. *Proc. Natl. Acad. Sci. USA* **106**, 6644–6649 (2009).
27. Dutzler, R., Campbell, E.B., Cadene, M., Chait, B.T. & MacKinnon, R. X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294 (2002).
28. Tereshko, V. *et al.* Toward chaperone-assisted crystallography: protein engineering enhancement of crystal packing and X-ray phasing capabilities of a camelid single-domain antibody (VHH) scaffold. *Protein Sci.* **17**, 1175–1187 (2008).
29. Koide, S. Engineering of recombinant crystallization chaperones. *Curr. Opin. Struct. Biol.* **19**, 449–457 (2009).
30. Carter, P.J. Potent antibody therapeutics by design. *Nat. Rev. Immunol.* **6**, 343–357 (2006).
31. Birtalan, S. *et al.* The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. *J. Mol. Biol.* **377**, 1518–1528 (2008).
32. Sidhu, S.S., Lowman, H.B., Cunningham, B.C. & Wells, J.A. Phage display for selection of novel binding peptides. *Methods Enzymol.* **328**, 333–363 (2000).
33. Laird-Offringa, I.A. & Belasco, J.G. In vitro genetic analysis of RNA-binding proteins using phage display libraries. *Methods Enzymol.* **267**, 149–168 (1996).
34. Ye, J.-D. Synthetic antibodies for specific recognition and crystallization of structured RNA. *Proc. Natl. Acad. Sci. USA* **105**, 82–87 (2008).
35. Ekland, E.H., Szostak, J.W. & Bartel, D.P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364–370 (1995).
36. Bagby, S.C., Bergman, N.H., Shechner, D.M., Yen, C. & Bartel, D.P. A class I ligase ribozyme with reduced Mg^{2+} dependence: selection, sequence analysis, and identification of functional tertiary interactions. *RNA* **15**, 2129–2146 (2009).
37. Fellouse, F.A., Wiesmann, C. & Sidhu, S.S. Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc. Natl. Acad. Sci. USA* **101**, 12467–12472 (2004).
38. Fellouse, F.A. *et al.* Molecular recognition by a binary code. *J. Mol. Biol.* **348**, 1153 (2005).
39. Decanniere, K. *et al.* A single-domain antibody fragment in complex with RNase A: non-canonical loop structures and nanomolar affinity using two CDR loops. *Structure* **7**, 361–370 (1999).
40. Desmyter, A. *et al.* Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology. *J. Biol. Chem.* **277**, 23645–23650 (2002).
41. Shechner, D.M. *et al.* Crystal structure of the catalytic core of an RNA polymerase ribozyme. *Science* **326**, 1271–1275 (2009).
42. Eigenbrot, C., Randal, M., Presta, L., Carter, P. & Kossiakoff, A.A. 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 (1993).
43. Kossiakoff, A.A. & Koide, S. Understanding mechanisms governing protein-protein interactions from synthetic binding interfaces. *Curr. Opin. Struct. Biol.* **18**, 499–506 (2008).
44. Jones, S. & Thornton, J.M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**, 13–20 (1996).
45. Lunde, B.M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **8**, 479–490 (2007).
46. Messias, A.C. & Sattler, M. Structural basis of single-stranded RNA recognition. *Acc. Chem. Res.* **37**, 279–287 (2004).
47. Bergman, N.H., Lau, N.C., Lehnert, V., Westhof, E. & Bartel, D.P. The three-dimensional architecture of the class I ligase ribozyme. *RNA* **10**, 176–184 (2004).
48. Saville, B.S. & Collins, R.A. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell* **61**, 685–696 (1990).
49. Chao, J.A., Patskovsky, Y., Almo, S.C. & Singer, R.H. Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.* **15**, 103–105 (2008).
50. Hogg, J.R. & Collins, K. RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA* **13**, 868–880 (2007).
51. Keryer-Bibens, C., Barreau, C. & Osborne, H.B. Tethering of proteins to RNAs by bacteriophage proteins. *Biol. Cell* **100**, 125–138 (2008).
52. Piekna-Przybylska, D., Liu, B. & Fournier, M.J. The U1 snRNA hairpin II as a RNA affinity tag for selecting snoRNP complexes. *Methods Enzymol.* **425**, 317–353 (2007).
53. Kabat, E.A. & Wu, T.T. Attempts to locate complementarity-determining residues in the variable positions of light and heavy chains. *Ann. NY Acad. Sci.* **190**, 382–393 (1971).

ONLINE METHODS

Fab phage-display library and general phage manipulation. YSG and YSGR Fab libraries were constructed at Genentech[31,32,54]. Class I ligase ribozyme product RNAs modified with a 3′ tag for phage display, as well as untagged *trans*- and *cis*-ligase ribozymes used in binding and activity assays, and class I ligase product for crystallization screens, were prepared according to published protocols[47]. Ligase substrate (5′-UCCAGUA-3′ RNA) and P5 and P7 RNA hairpin oligonucleotides were purchased from Integrated DNA Technologies.

Fab selection and expression. Phage-display selections were done at room temperature by the magnetic-beads method[33]. General phage library manipulation and soluble Fab expression methods were as described[34].

Affinity maturation by error-prone PCR. To generate mutated Fab libraries by error-prone PCR from the Fab BL3 parent sequence, we amplified the Fab BL3 variable heavy-chain and light-chain domains with an error-prone polymerase (Mutazyme II, Stratagene) at two mutation rates: ~18 per kilobase (kb) and ~27 per kb, estimated by the PCR amplification factor. Mutated light-chain and heavy-chain fragments were then annealed to the parent (BL3 Fab) single-stranded DNA template, and the Fab constant domains and the phagemid vector were extended with a high-fidelity DNA polymerase. Primer sequences were 5′-CAGATGACCC AGTCCCCGAGCTCCCTG-3′ and 5′-CGTTTGATCTCCACCTTGGTACCCT GTCCGAA-3′ for light-chain amplification, and 5′-GTTCAGCTGGTGGAGTC TGGCGGTGGC-3′ and 5′-CGAGGAGACGGTGACCAGGGTTCCTTGACC-3′ for heavy-chain amplification (all from Integrated DNA Technologies). We used 0.1 ng template (BL3 single-stranded DNA) per 50 µl PCR reaction to generate the maximum mutation rate. We then reamplified 0.1 ng of the mutated heavy-chain and light-chain mixtures to produce a mixture of fragments with a mutation rate of ~27 per kb. Mutated light-chain and heavy-chain fragments were then used as mutant mega-primers in a QuikChange Multi (Stratagene) amplification of the parent BL3 plasmid. The mutant phagemid library was electroporated into *E. coli* SS320 cells and amplified with M13-KO7 helper phage and ampicillin.

Phage-display selections using the error-prone PCR phage libraries were similar to those described[34], but with reduced ligase and biotin concentrations (2.5 nM, 0.25 nM and 0.025 nM). After each round, before phage amplification, phage titers were used to determine the lowest target concentration that produced an enrichment ratio >10. After round 3, phages eluted from the lowest target concentration with enrichment >10 were sequenced, and Fabs were reformatted and expressed as described[32].

Nitrocellulose filter binding assay. The nitrocellulose binding assay was carried out as described[34], using TEM buffer (50 mM Tris pH 7.6, 10.1 mM $MgCl_2$, 0.1 mM EDTA, 150 mM NaCl) as the binding buffer, with 0.5 mg ml^{-1} heparin, trace ^{32}P-labeled class I ligase product RNA and 2 nM to 2 µM Fabs.

Hydroxyl-radical footprinting. Hydroxyl-radical footprinting relied on the ability of the class I ligase ribozyme to prepare self-radiolabeled RNA upon reaction with ^{32}P 5′ end–labeled substrate[47]. We prepared self-labeled ligase by incubating 2 µM 5′-triphosphate ligase ribozyme at 80 °C for 5 min and then at 22 °C for 10 min, followed by 30 min at 22 °C in 1× PBS supplemented with 10 mM $MgCl_2$ to fold the ribozyme. Addition of ^{32}P 5′ end–labeled substrate (5′-^{32}P-UCCAGUA-3′) initiated the self-labeling ligation reaction. Hydroxyl-radical footprinting was then carried out as described in ref. 34 and **Supplementary Methods**. Electrophoresis gels contained 10% acrylamide, 8 M urea and 0.5× TBE buffer.

Crystallization of the Fab BL3-6–ligase complex. Class I ligase ribozyme product was synthesized and annealed as described[6]. Crystals were grown by the hanging-drop vapor diffusion method at 4 °C; 1 volume of Fab–ligase complex (final concentration 7 µg ml^{-1}) was combined with 1 volume of crystallization buffer (50 mM cacodylate pH 6.0, 20 mM NaCl, 1 mM $CdCl_2$, 32% (v/v) 2-methyl-2,4-pentanediol), and equilibrated against 1 ml of the crystallization buffer. Small cube-shaped crystals were observed within 1 week.

Data collection, phasing and refinement. Crystals were mounted on Hampton Cryoloops, coated with Paratone-N and frozen in liquid nitrogen for data collection. A data set extending to 3.1-Å resolution was collected at 1.127-Å wavelength at beamline 21-ID-D of the Advanced Photon Source. Complete information regarding data collection, phasing and refinement is provided in **Table 1** and **Supplementary Notes**.

Additional methods and requests for materials. BL3-6 plasmids, detailed protocols for expression and purification of Fab BL3-6, and Fab BL3-6 itself (if available) will be provided upon request.

54. Fellouse, F.A. *et al.* High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J. Mol. Biol.* **373**, 924 (2007).