Published in final edited form as:
Nat Struct Mol Biol. 2011 January ; 18(1): 100–106. doi:10.1038/nsmb.1945.

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,*], **Nikolai B. Suslov**[2], **Jingdong Ye**[5], **Sachdev S. Sidhu**[6], **David P. Bartel**[3,4], **Shohei Koide**[2], **Anthony A. Kossiakoff**[2,†], and **Joseph A. Piccirilli**[1,2,†]

[1] Department of Chemistry, University of Chicago, Chicago, IL 60637, USA

[2] Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA

[3] Whitehead Institute for Biomedical Research and Howard Hughes Medical Institute, 9 Cambridge Center, Cambridge, MA 02142, USA

[4] Department of Biology and Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[5] Department of Chemistry, University of Central Florida, Orlando, FL 32816, USA

[6] Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, MS5 3E1, Canada

Abstract

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 3.1Å 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 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 the smaller size of the loop and high molecular weight, large surface area, and phasing power provided by Fabs.

Introduction

The advent of next-generation sequencing has triggered an explosion in the pace of non-coding RNA (ncRNA) discovery, with an expanding repertoire of corresponding functions.

[†]To whom correspondence should be addressed: Joseph A. Piccirilli, jpicciri@uchicago.edu, phone: 773-702-9312, fax: 773-702-0439, Anthony A. Kossiakoff, koss@bsd.uchicago.edu, phone: 773-702-5073, fax: 773-702-0439.
[*]Current address: Broad Institute and Harvard University Department of Stem Cell and Regenerative Biology, Cambridge, MA 02138, USA.

Author Contributions

Y.K., E.M.D., D.M.S., N.B.S., J.Y., S.S.S., D.P.B., S.K., A.A.K. and J.A.P. 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.

[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 (human, mice and others) is transcribed, giving rise to vast numbers of RNA transcripts.[6,7] As a relatively small fraction of these transcripts code for proteins, there likely exists an expansive landscape of yet undiscovered ncRNAs.

To perform their biological roles, many ncRNAs and non-coding regions of mRNAs adopt complex three-dimensional architectures. Defining these structures represents an important step towards understanding ncRNA function, and in recent years, some general principles that govern RNA architecture have emerged from structures of ribozymes, riboswitches, ribosomes and other RNA and RNP complexes.[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.[14] 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.[15] Additionally, RNAs frequently have flexible regions that enable sampling of alternative conformations or have a tendency to misfold, leading to conformational heterogeneity.[16–19] Confounding matters further, phasing of RNA crystals involves complex methods in contrast to the well-established selenium-based phasing of protein crystals.[16,20,21]

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.[16,17,19,22,23] Another approach, and one that supports the proof-of-concept that drives this work, 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 10-nucleotide sequence recognized by the U1A protein, and crystallizing the RNA in complex with the U1A protein.[24] Despite the well-documented success of the chaperone approach for protein crystallography over the past decade,[25–27] 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 of several ‘difficult’ protein targets.[25,26,28–30] Antibody chaperones seem particularly well suited for overcoming some of the challenges inherent to RNA crystallization.[26] With a higher molecular weight (50 kDa) relative to the U1A protein (11 kDa), Fab chaperones provide more surface area for crystal contacts and their beta-rich architecture is predisposed to making effective crystal contacts.[30] 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.[26]

In recent years, the development of natural and synthetic immune repertoires and selection methodologies has enabled antibody production without host immunization.[31–34] Using a phage platform to display libraries of synthetic antigen-binding fragments (Fabs), we recently established a general approach to obtain 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.[35] These antibodies were used successfully to crystallize the target RNA and to solve the structure of the P4-P6 RNA domain at 1.95Å resolution.

In this work 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.[36,37] 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 30–50 nM affinities. We solved the crystal structure of the Fab:ligase complex at 3.1Å 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 demonstrate that the critical 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 previously used a reduced codon[38,39] 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.[35] Subsequent selections using the YSG library produced no RNA-binding clones against a variety of RNA targets. Considering the enrichment of cationic K and R residues in the P4–P6-binding Fabs,[35] we deployed another library already enriched in K and R residues. This library (referred to hereafter as the YSGR super library) consisted of a mixture of four reduced-codon libraries (YSG, YSR, YSGR, and YSGRZ), with CDRs L3, H1, and H2 again containing equimolar mixtures of Tyr and Ser. For CDR-H3, YSG and YSR sub-libraries adopted the same design as reported previously[32] with the following components: YSG (50% Tyr, 25% Ser, 25% Gly) and YSR (YSR (25% Tyr, 50% Ser, 25% Arg), while H3 CDRs in YSGR and YSGRZ sub-libraries were designed as YSGR (38% Tyr, 25% Ser, 25% Gly, 12% Arg) and YSGRZ (19% Tyr, 25% Ser, 25% Gly, 12% Arg, 19% Z, where Z represents an equimolar mixture of all amino acids except Tyr, Ser, Gly, Arg, and Cys), respectively.

We used the product form of an improved variant of the class I ligase ribozyme[40] (Fig. 1a) as the target for phage display selections. Two YSGR super library 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 using a nitrocellulose filter binding assay, revealing K_d's in the 300–500 nM range for BL1 and BL3, and >2000 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. Since most antibody chaperones successfully used for crystallization bind their targets with K_d's < 250

nM27,35,41,42, 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 selected error-prone PCR to introduce random mutations throughout the variable domains of BL3, mainly because the method has minimal bias toward a 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 (HC) and light chain (LC) 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 target RNA (class I ligase product) concentration 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 >10 relative to a blank selection were either amplified for the next round of selection (round 2) or sequenced (round 3). Sequencing analysis revealed 10 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 (6 of 10) 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 (2- to 3-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 5- to 10-fold, resulting in 30–70 nM K_d's, 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) showed two regions of Fab-induced protection: one corresponding to the ligase P5 loop (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 appearance of these protections as a function of Fab concentration confirmed a ~75 nM K_d observed by filter binding (Supplementary Fig. 2b). Additionally, the Fab-ligase complex retained the Mg^{2+}-dependent protections 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 3.1Å. Table 1, the Supplementary Methods, and Supplementary Fig. 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, initiated several years earlier, replaced loop 5 (L5) with the U1A-binding loop and yielded crystals of the ligase-U1A complex that diffracted to 3.0Å but were much more challenging to grow and phase.[43] For the Fab-ligase co-crystals, 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 co-crystals involved substantially more effort, with several

established methods attempted before finding one 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 RMSD of 1.26Å.[43] These observations, together with the observation that the ligase remains fully active with Fab bound (Supplementary Fig. 1a), illustrate that Fab-ligase interactions do not substantially distort the ligase RNA backbone, and that Fab BL3-6 serves as an effective crystal chaperone.

As recently reported,[43] 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 (Fig. 3a)—that are joined by J3/4 (light blue) and the P4-P5-P6-P7 4-helix junction. Two single-stranded regions, J1/3 (dark blue), which spans the length of the ribozyme and docks into the minor grooves of P1 and P6, and J3/4 (light blue, Fig. 3a), 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), possibly contributing 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 Tyr, Ser, and Arg 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 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 3.5-4.5Å of residues located in CDR-H2 and CDR-H3 (Tyr62 of CDR-H2 and A62; Tyr62 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 (see below and Fig. 6d). 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. 5a).

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, RMSD = 0.51Å).[44] Although the elbow angles differ slightly due 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.[45] Fab-BL3-6 binds to the ligase with a relatively large buried surface area (1140 $Å^2$ on the Fab side alone), substantially greater

than that of most Fab-protein antigen interfaces (777±135 Å^2)[46] but similar to that observed previously for Fab2 binding to ΔC209 P4-P6. 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 illustrate that Fabs can provide a highly versatile scaffold for binding to a wide range of RNAs. Fab-ligase recognition also shares several features in common with single-stranded RNA binding proteins [47,48] (see the Supplementary Note for comparison of similarities and differences).

Features of Fab-ligase crystal packing

A view of the Fab-ligase crystal lattice illustrates 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, with alternate squares filled by clusters of RNA with extensive RNA-RNA contacts. Large ~100x100Å channels 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 kD), Fab BL3-6 and the ligase ribozyme have approximately equal surface areas: 19740 Å^2 (average of the two Fabs in the asymmetric unit) and 20654 Å^2 (average for the two ligase molecules in the asymmetric unit). Buried surface area analysis using the CCP4 PISA webserver[49] suggests that Fab-RNA contacts account for the majority of surface burial in the structure (65%, or 3818.5Å^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% (1251.4 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 (the P5 hairpin loop and the P7 hairpin loop of the ligase), 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,[50] these results suggested 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 abolished detectable Fab binding, including variants carrying a UUCG loop (not shown), C63 deletion, C63A substitution, and a A61C/C63A double substitution (Fig. 6c), consistent with crystallographically inferred interactions between the C63 nucleobase and CDR-H2 (Ser58 and Ser60). In addition, no detectable binding activity was observed for the linear AAACA RNA pentanucleotide alone, an unstructured 15-mer RNA containing the AAACA sequence, a double-stranded RNA containing the AAACA sequence in one of the strands, or a DNA oligonucleotide corresponding to the P5 sequence (Fig. 6d). In contrast to the highly deleterious loop mutations, mutations in the stem affected binding minimally (Fig. 6c) with the exception of the closing G-C base pair, where substitution with A-U or C-G pairs increased Fab-

oligonucleotide K_d's 3x and 5x, respectively, relative to the original oligonucleotide with the G-C closing base pair. The interaction between Arg106 and N7 of G59 may engender the Fab's preference for the G-C base pair versus C-G (Figs. 4b and 6c). 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 VS 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 with no introduction of 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. 2b, 6b, and Supplementary Fig. 6b). These results demonstrate the portability of the GAAACAC antigen for general use as a crystallization chaperone or affinity tag.

Discussion

Our previous work established proof-of-concept for using Fabs as potential crystallization chaperones for RNA. The work demonstrated that the phage-display technology could be used to obtain an anti-RNA Fab and that the Fab could be used as a chaperone for P4-P6 crystallization (see Supplementary Note for a timeline of the process). However, the library employed 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, dramatically enhances the potential for obtaining high affinity Fabs to RNA targets. Additionally, we have shown 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. Importantly, 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 for generating 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.[24] Fab-BL3-6 and the corresponding RNA loop provide an additional RNA crystallization chaperone/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 determinations using the Fab/GAACAC system can begin the process 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 *E. coli* culture) that can be purified RNase-free in a facile manner 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[51–54].

Database accession numbers

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

Methods

BL3-6 plasmids, detailed protocols for expression and purification of Fab BL3-6, and Fab BL3-6 itself (pending availability) will be provided upon request.

Fab phage display library and general phage manipulation

YSG and YSGR Fab libraries were constructed at Genentech.[32,33,55] 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.[50] 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 using the magnetic beads method.[34] General phage library manipulation and soluble Fab expression methods were as described by Ye et al.[35]

Affinity maturation by error-prone PCR

To generate mutated error-prone PCR Fab libraries from the Fab-BL3 parent sequence, we amplified the Fab-BL3 variable heavy chain (HC) and light chain (LC) domains with an error-prone polymerase (Mutazyme II, Stratagene) at two mutation rates: ~18 per kb and ~27 per kb, estimated by the PCR amplification factor. Mutated LC and HC fragments were then annealed to the parent (BL3 Fab) ssDNA template, and the Fab constant domains and the phagemid vector were extended with a high-fidelity DNA polymerase. Primer sequences were 5′-CAGATGACCCAGTCCCCGAGCTCCCTG-3′ and 5′-CGTTTGATCTCCACCTTGGTACCCTGTCCGAA-3′ for light chain amplification, and 5′-GTTCAGCTGGTGGAGTCTGGCGGTGGC-3′ and 5′-CGAGGAGACGGTGACCAGGGTTCCTTGACC-3′ for heavy chain amplification (all from Integrated DNA Technologies). 0.1 ng template (BL3 ssDNA) was used per 50 μL PCR reaction to generate the maximum mutation rate. 0.1 ng of the mutated HC and LC mixtures were then re-amplified to produce a mixture of fragments with a mutation rate of ~27 per kb. Mutated LC and HC 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 previously,[35] but with reduced ligase/biotin concentrations (2.5 nM, 0.25 nM, and 0.025 nM). Prior to phage amplification after each round, phage titers were used to determine the lowest target concentration that produced an enrichment ratio greater than 10. After round 3, phages eluted from the lowest target concentration with enrichment >10 were sequenced, and Fabs were reformatted and expressed as described previously.[33]

Nitrocellulose filter binding assay

The nitrocellulose binding assay was carried out as described by Ye et al.[35], using "TEM" (50 mM Tris pH 7.6, 10.1 mM $MgCl_2$, 0.1 mM EDTA, 150 mM NaCl) as the binding buffer.

Hydroxyl radical footprinting

Hydroxyl radical footprinting relied on the ability of the class I ligase ribozyme to prepare self-radiolabeled RNA upon reaction with 5′-^{32}P-labeled substrate.[50] Self-labeled ligase was prepared by incubating 2 μM 5′-triphosphate ligase ribozyme at 80°C for 5 min and 22°C for 10 min, followed by 30 min at 22°C in 1x PBS supplemented with 10 mM $MgCl_2$ at for 30 min to fold the ribozyme. Addition of 5′-^{32}P-labeled substrate (5′-^{32}P- UCCAGUA-3′) initiated the self-labeling ligation reaction. Hydroxyl radical footprinting was then carried out as described by Ye et al.[35] and in the Supplementary Methods.

Crystallization of the BL3-6 Fab – ligase complex

Class I ligase ribozyme product was synthesized and annealed as described previously.[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 $CdCl_2$, and 32% MPD), 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 3.1Å resolution was collected at 1.127Å wavelength at Beamline 21-ID-D at the Advanced Photon Source. Complete information regarding data collection, phasing, and refinement is provided in Table 1 and in the Supplementary Notes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Piccirilli, Koide, and Kossiakoff labs for assistance with phage display and for insightful comments; David Lilley and Timothy Wilson for helpful discussions regarding the VS ribozyme; and Vladimir Torbeev, Irina Dementieva, Phoebe Rice, and Valentina Tereshko for assistance with crystallography.

Funding: HHMI (J.A.P.); National Institute of General Medical Sciences: Medical Scientist National Research Service Award # 5 T32 GM07281 (Y.K.); NIH 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 U. S. 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).

References

1. Eddy SR. Noncoding RNA genes. Curr Opin Genet Dev. 1999; 9:695–699. [PubMed: 10607607]
2. Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays. 2007; 29:288–299. [PubMed: 17295292]

3. Mattick JS. The genetic signatures of noncoding RNAs. PLoS Genet. 2009; 5:e1000459. [PubMed: 19390609]
4. Cooper TA, Wan L, Dreyfuss G. RNA and Disease. Cell. 2009; 136:777. [PubMed: 19239895]
5. Hogg JR, Collins K. Structured non-coding RNAs and the RNP Renaissance. Current Opinion in Chemical Biology. 2008; 12:684. [PubMed: 18950732]
6. Kapranov P, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007; 316:1484–1488. [PubMed: 17510325]
7. Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet. 2006; 15(Spec No 1):R17–29. [PubMed: 16651366]
8. Montange RK, Batey RT. Riboswitches: emerging themes in RNA structure and function. Annu Rev Biophys. 2008; 37:117–133. [PubMed: 18573075]
9. Serganov A, Patel DJ. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat Rev Genet. 2007; 8:776–790. [PubMed: 17846637]
10. Fedor MJ. Comparative enzymology and structural biology of RNA self-cleavage. Annu Rev Biophys. 2009; 38:271–299. [PubMed: 19416070]
11. Hoogstraten CG, Sumita M. Structure-function relationships in RNA and RNP enzymes: recent advances. Biopolymers. 2007; 87:317–328. [PubMed: 17806104]
12. Jinek M, Doudna JA. A three-dimensional view of the molecular machinery of RNA interference. Nature. 2009; 457:405–412. [PubMed: 19158786]
13. Reichow SL, Hamma T, Ferre-D'Amare AR, Varani G. The structure and function of small nucleolar ribonucleoproteins. Nucleic Acids Res. 2007; 35:1452–1464. [PubMed: 17284456]
14. Research Collaboratory for Structural Bioinformatics, P. D. B. http://www.rcsb.org/pdb/statistics/holdings.do
15. Doudna JA, Cate JH. RNA structure: crystal clear? Curr Opin Struct Biol. 1997; 7:310–316. [PubMed: 9204271]
16. Golden BL, Kundrot CE. RNA crystallization. Journal of Structural Biology. 2003; 142:98. [PubMed: 12718923]
17. Edwards AL, Garst AD, Batey RT. Nucleic Acid and Peptide Aptamers. 2009; 1
18. Ke A, Doudna JA. Crystallization of RNA and RNA-protein complexes. Methods. 2004; 34 :408. [PubMed: 15325657]
19. Ferre-D'Amare AR, Zhou K, Doudna JA. A general module for RNA crystallization. Journal of Molecular Biology. 1998; 279:621. [PubMed: 9641982]
20. 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. Nucl Acids Res. 2004; 32:1638–1646. [PubMed: 15007109]
21. Hobartner C, et al. Syntheses of RNAs with up to 100 Nucleotides Containing Site-Specific 2'-Methylseleno Labels for Use in X-ray Crystallography. Journal of the American Chemical Society. 2005; 127:12035. [PubMed: 16117544]
22. Keel AY, Rambo RP, Batey RT, Kieft JS. A General Strategy to Solve the Phase Problem in RNA Crystallography. Structure. 2007; 15:761. [PubMed: 17637337]
23. Robertson MP, Scott WG. A general method for phasing novel complex RNA crystal structures without heavy-atom derivatives. Acta Crystallogr D Biol Crystallogr. 2008; D64:738–744. [PubMed: 18566509]
24. Ferre-D'Amare AR, Doudna JA. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. Journal of Molecular Biology. 2000; 295:541. [PubMed: 10623545]
25. Iwata S, Ostermeier C, Ludwig B, Michel H. Structure at 2.8 A resolution of cytochrome c oxidase from Paracoccus denitrificans. Nature. 1995; 376:660. [PubMed: 7651515]
26. Koide S. Engineering of recombinant crystallization chaperones. Current Opinion in Structural Biology. In Press, Corrected Proof (2009).[AU: Is updated information available for this reference?].
27. Uysal S, et al. Crystal structure of full-length KcsA in its closed conformation. Proc Natl Acad Sci U S A. 2009; 106:6644–6649. [PubMed: 19346472]

28. Dutzler R, Campbell EB, Cadene M, Chait BT, MacKinnon R. X-ray structure of a ClC chloride channel at 3.0[thinsp][angst] reveals the molecular basis of anion selectivity. Nature. 2002; 415:287. [PubMed: 11796999]
29. 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. 2008; 17:1175–1187. [PubMed: 18445622]
30. Koide S. Engineering of recombinant crystallization chaperones. Curr Opin Struct Biol. 2009; 19:449–457. [PubMed: 19477632]
31. Carter PJ. Potent antibody therapeutics by design. Nat Rev Immunol. 2006; 6:343. [PubMed: 16622479]
32. Birtalan S, et al. The Intrinsic Contributions of Tyrosine, Serine, Glycine and Arginine to the Affinity and Specificity of Antibodies. Journal of Molecular Biology. 2008; 377:1518. [PubMed: 18336836]
33. Sidhu, SS.; Lowman, HB.; Cunningham, BC.; Wells, JA.; Jeremy Thorner, SDEaJNA. Methods in Enzymology. Vol. 328. Academic Press; 2000. p. 333
34. Laird-Offringa, IA.; Belasco, JG.; John, NA. Methods in Enzymology. Vol. 267. Academic Press; 1996. p. 149
35. Ye J-D, et al. Synthetic antibodies for specific recognition and crystallization of structured RNA. Proceedings of the National Academy of Sciences. 2008; 105:82–87. %R. 10.1073/pnas. 0709082105
36. Ekland EH, Szostak JW, Bartel DP. Structurally complex and highly active RNA ligases derived from random RNA sequences. Science. 1995; 269:364–370. [PubMed: 7618102]
37. Bagby SC, Bergman NH, Shechner DM, Yen C, Bartel DP. A class I ligase ribozyme with reduced Mg2+ dependence: Selection, sequence analysis, and identification of functional tertiary interactions. RNA. 2009; 15:2129–2146. [PubMed: 19946040]
38. Fellouse FA, Wiesmann C, Sidhu SS. Synthetic antibodies from a four-aminoacid code: a dominant role for tyrosine in antigen recognition. Proc Natl Acad Sci U S A. 2004; 101:12467–12472. [PubMed: 15306681]
39. Fellouse FA, et al. Molecular Recognition by a Binary Code. Journal of Molecular Biology. 2005; 348:1153. [PubMed: 15854651]
40. Bagby SC, Bergman NH, Shechner DM, Yen C, Bartel DP. A class I ligase ribozyme with reduced Mg2+ dependence: Selection, sequence analysis, and identification of functional tertiary interactions. RNA. 2009; 15:2129–2146. [PubMed: 19946040]
41. 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. 1999; 7:361–370. [PubMed: 10196124]
42. Desmyter A, et al. Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology. J Biol Chem. 2002; 277:23645–23650. [PubMed: 11960990]
43. Shechner D, et al. Crystal structure of the catalytic core of an RNA polymerase ribozyme. Science. 2009; 326:1271–1275. [PubMed: 19965478]
44. Eigenbrot C, Randal M, Presta L, Carter P, Kossiakoff AA. 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. 1993; 229:969–995. [PubMed: 8095303]
45. Kossiakoff AA, Koide S. Understanding mechanisms governing protein-protein interactions from synthetic binding interfaces. Curr Opin Struct Biol. 2008; 18:499–506. [PubMed: 18638552]
46. Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci U S A. 1996; 93:13–20. [PubMed: 8552589]
47. Lunde BM, Moore C, Varani G. RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol. 2007; 8:479. [PubMed: 17473849]
48. Messias AC, Sattler M. Structural basis of single-stranded RNA recognition. Acc Chem Res. 2004; 37:279–287. [PubMed: 15147168]
49. CCP4 PISA Webserver. http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html

50. Bergman NH, Lau NC, Lehnert V, Westhof E, Bartel DP. The three-dimensional architecture of the class I ligase ribozyme. RNA. 2004; 10:176–184. [PubMed: 14730016]

51. Chao JA, Patskovsky Y, Almo SC, Singer RH. Structural basis for the coevolution of a viral RNA-protein complex. Nat Struct Mol Biol. 2008; 15:103–105. [PubMed: 18066080]

52. Hogg JR, Collins K. RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. RNA. 2007; 13:868–880. [PubMed: 17456562]

53. Keryer-Bibens C, Barreau C, Osborne HB. Tethering of proteins to RNAs by bacteriophage proteins. Biol Cell. 2008; 100:125–138. [PubMed: 18199049]

54. Piekna-Przybylska D, Liu B, Fournier MJ. The U1 snRNA hairpin II as a RNA affinity tag for selecting snoRNP complexes. Methods Enzymol. 2007; 425:317–353. [PubMed: 17673090]

55. Fellouse FA, et al. High-throughput Generation of Synthetic Antibodies from Highly Functional Minimalist Phage-displayed Libraries. Journal of Molecular Biology. 2007; 373:924. [PubMed: 17825836]

56. Kabat EA, Wu TT. Attempts to locate complementarity-determining residues in the variable positions of light and heavy chains. Annals of the New York Academy of Sciences. 1971; 190:382–393. [PubMed: 5290024]

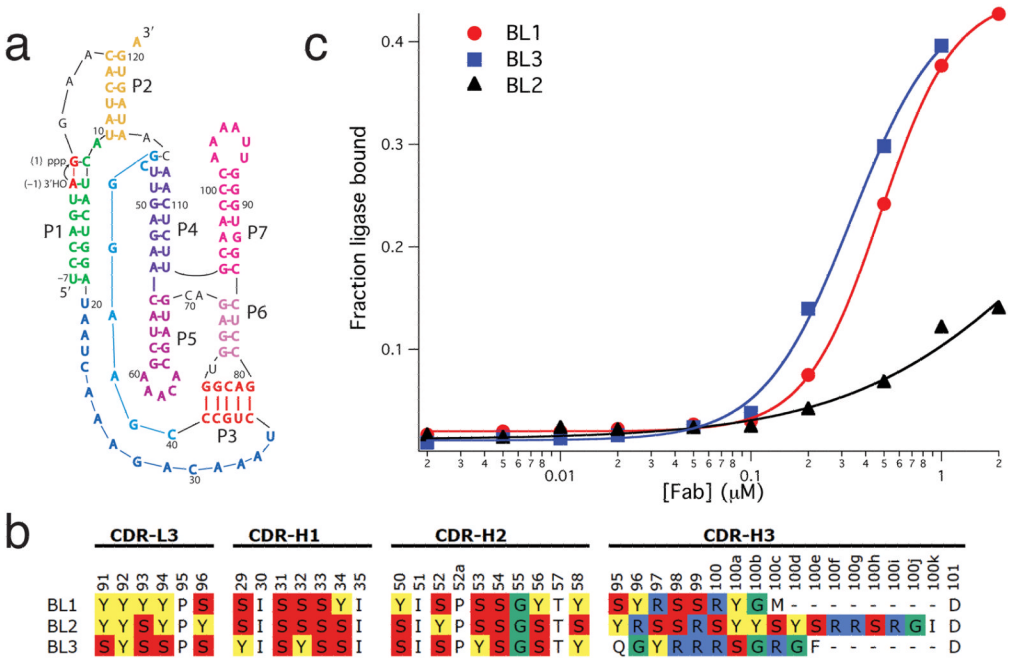

Figure 1.

Selection of class I ligase-binding Fabs from the YSGR Fab super library. (**a**) Secondary structure of the ligase ribozyme and its substrate[40]. The arrow indicates the self- ligation reaction, in which the substrate 3′-terminal hydroxyl (nucleotide A-1, red) attacks the alpha phosphate of the ribozyme 5′-terminal triphosphate (nucleotide G1, red). (**b**) CDR sequences of ligase-binding Fabs selected from the YSGR super library. Red = serine, yellow = tyrosine, green = glycine, blue = arginine. Residues are numbered according to the Kabat system.[56] (**c**) Nitrocellulose binding assay reveals low-micromolar K_d's for Fab-class I ligase binding. BL1 K_d = 478 nM; BL2 K_d > 2000 nM; BL3 K_d = 338 nM. Conditions: 50 mM Tris pH 7.6, 10.1 mM $MgCl_2$, 0.1 mM EDTA, 150 mM NaCl, 0.5 mg/mL heparin, trace ^{32}P-labeled class I ligase product RNA, 2 nM to 2 μM Fabs.

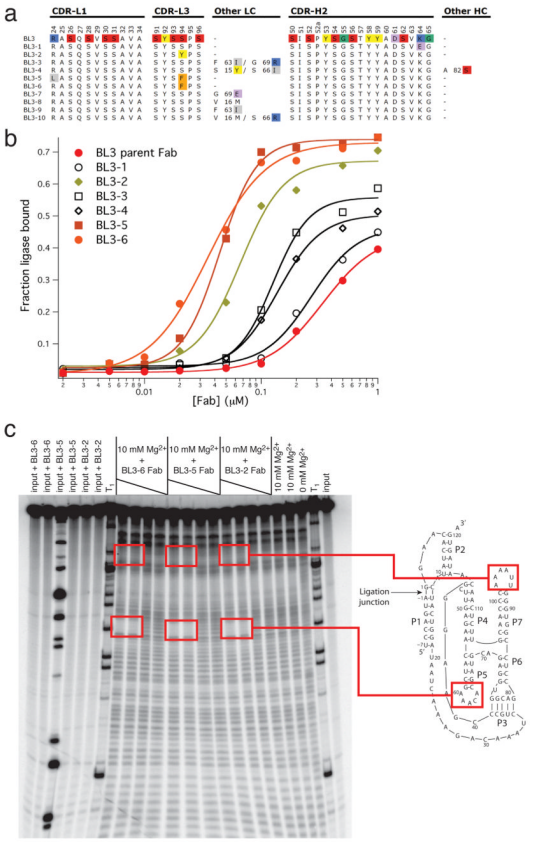

Figure 2.

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's: 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. Conditions: Nitrocellulose filter binding assay conditions identical to those in Fig. 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, likely reflects the faster dissociation of weaker-binding Fabs in the context of the nitrocellulose filter binding assay. Accordingly, we used the nitrocellulose filter binding assay only as a convenient, qualitative method to assay Fab-RNA affinity. As an independent method to assess BL3 Fab K_d's 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. 1x PBS pH 7.3, 10mM $MgCl_2$, 0.5 mg/mL heparin, trace ^{32}P-self-labeled ligase product.

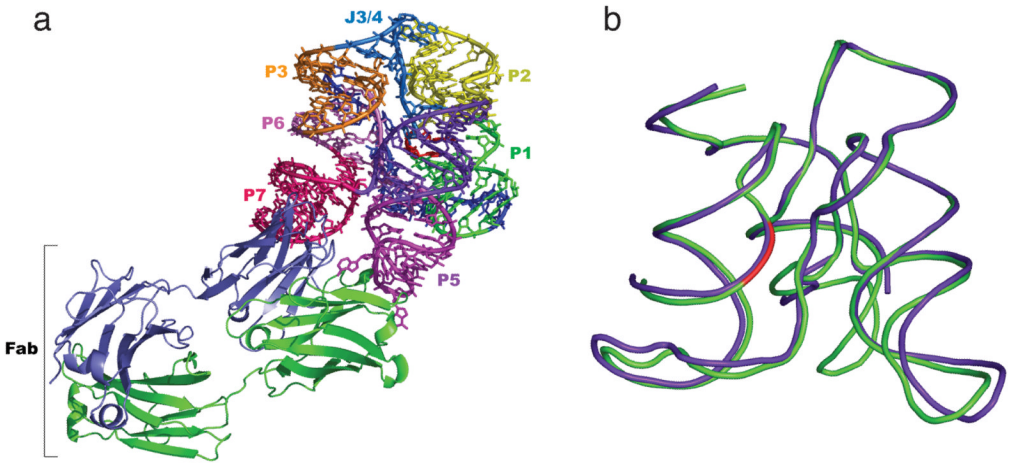

Figure 3.

Crystal structure of the Fab-BL3-6-ligase complex. (a) Crystal structure of the BL3-6 Fab-ligase complex at 3.1Å resolution. (b) Overlay of the Fab-ligase structure (green) with the U1A-ligase structure[43] (blue); red indicates the ligation junction. Nucleotides in the U1A-binding loop have been omitted from view. All-atom RMSD = 1.26 Å, omitting the residues in L5.

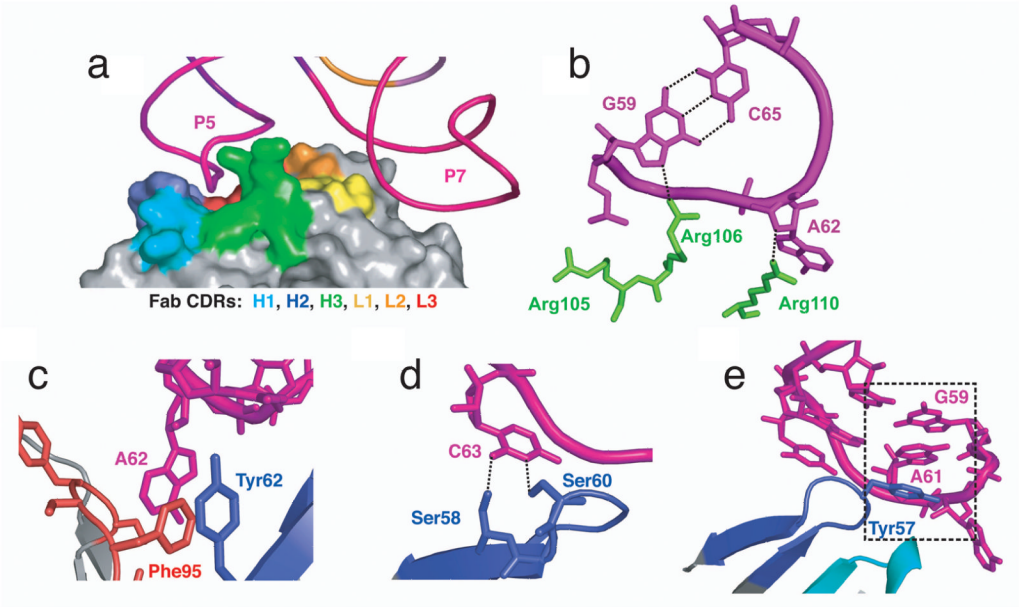

Figure 4.

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. (**f**) summary of BL3-6 CDR residues forming contacts to P5. Note: Fab CDRs are numbered according to PDB residue numbers (top line) and the Kabat antibody numbering system (bottom line).[56]

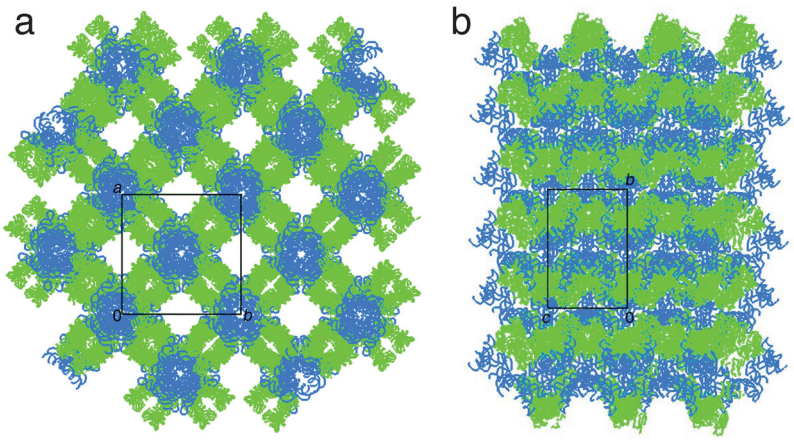

Figure 5.
Fab-ligase crystal packing. (**a**) Crystal packing along the a-b plane. (**b**) Crystal packing along the b-c plane. Green = Fab, blue = ligase.

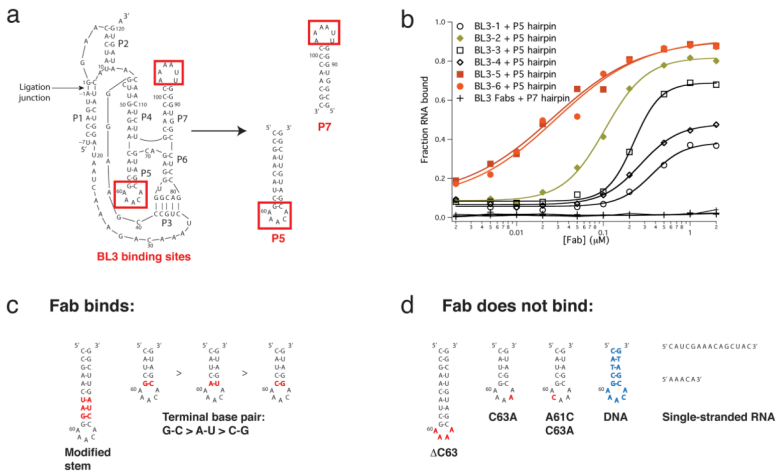

Figure 6.

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. Fab-class I ligase binding sites are indicated by red boxes. (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; and BL3-4, K_d = 270 nM; BL3-5, K_d = 26 nM; BL3-6, K_d = 28 nM. Also plotted are results for the P7 hairpin incubated with BL3 Fabs. Conditions: Nitrocellulose filter binding assay conditions identical to those in Fig. 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.

Table 1

Data collection and refinement statistics (molecular replacement)

	BL3-6 – ligase
Data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	207.5, 206.5, 135.9
α, β ,γ (°)	90, 90, 90
Resolution (Å)	50.0-3.1 (3.15–3.10)*
R_{sym}	0.104 (0.557)
$I/\sigma I$	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	49591
R_{work}/R_{free}	20.8/22.5 (31.7/34.5)
No. atoms	12299
Protein	6604
RNA	5554
Ligand/ion	141
Water	—
B-factors	68.09
Protein	63.25
RNA	83.57
Ligand/ion	71.20
Water	—
R.m.s. deviations	
Bond lengths (Å)	0.006
Bond angles (°)	1.323

*Number of crystals: 1

*Values in parentheses are for highest-resolution shell.