Supporting Online Material for

Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme

David M. Shechner, Robert A. Grant, Sarah C. Bagby, Yelena Koldobskaya, Joseph A. Piccirilli, David P. Bartel*

*To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu

Published 27 November 2009, *Science* **326**, 1271 (2009)
DOI: 10.1126/science.1174676

This PDF file includes

Supplemental Discussion

Regarding General Acid Base Catalysis by C47 N4

Formally, the C47 N4 exocyclic amine could serve as a general acid catalyst according to two models: (**1**) C47 acts from the neutral form, generating the nucleobase anion following proton transfer from N4, or (**2**) the catalytically active form of C47 is a cation that is protonated at N3. N3 would hence serve as an electron sink, favoring a tautomeric shift to the N4-imino form as a proton is transferred from N4 during catalysis. We consider both of these models unlikely. For Model **1**, the neutral form of C (pK_a ~12) is not sufficiently acidic to serve as an effective general acid. For Model **2**, in order to provide general acid catalysis without influencing the pH-rate profile over the range 5.7-8.5, the ribozyme would have to shift the pKa of the C47 cation by at least four pKa units (from 4 to > 8). Furthermore, methylation of N3, which would perturb the pK_a for both N3 and N4, has a negligible effect on catalysis (*S1*), which also suggests that model **2** is not valid.

Known and Novel Structural Motifs in the Class I Ligase

The GNRA-tetraloop-like triloop enclosed between C40 and G44 is one of two familiar structural motifs that cap opposite ends of the P3-P6-P7 domain. L7 (U93-A98) forms a canonical "uridine-turn (U-turn)" motif, observed first in the anticodon loop of yeast phenylalanyl tRNA (*S2*) and subsequently in the active site of the hammerhead ribozyme (fig. S6, (*S3,S4*)). The ubiquity of this motif in diverse RNA species might imply that its role in the ligase is analogous to that of the J3/4 GAA triloop, namely, to provide a short, thermodynamically stable cap at the end of a helical region. This being said, this motif is not required for ligase activity, as the L7 sequence is variable in active ligase and polymerase isolates (*S1, S5–S7*).

A third known motif involves the docking of nucleotides A22–A25 (near the 5´ end of J1/3) into P1. These residues form a structure nearly identical to that of a small viral ribosomal frameshifting pseudoknot (fig. S7, (*S8*)). U23 is of particular interest, as the analogous residue in the viral pseudoknot is an adenosine, yet both are able to share a network of hydrogen bonds with the following cytidine residue and the 2´-hydroxyl of the docking nucleotide (fig. S7B). Supporting the isosteric nature of these interactions,

an adenosine at position 23 is frequently observed among highly active isolates from our most recent ligase selection (*S1*), and the improved polymerase variants isolated by Zaher and Unrau show a sequence bias for adenosines in the region analogous to the 5´ end of J1/3 (*S9*).

The viral pseudoknot-like fold near the 5´ end of J1/3 immediately abuts one copy of a novel substructure we observe twice in the class I ligase, a four-nucleotide motif we call the "A-minor triad" (fig S8). Residues A25–A27 dock into the minor groove of two consecutive base pairs in P1, using a sequence of hydrogen bonds directly mirrored by residues A31–A33 and helix P6 (fig 2B,C, fig. S8A and S9). In both cases, a base that bears a major-groove carbonyl (G28 at the 3´ end of the P1 triad, U34 at the 3´ end of the P6 triad) follows this succession of adenosines. In both cases this fourth base is rotated away from the docking helix, directing its carbonyl downward toward the base triples which preceded it, such that the backbone strand of the motif assumes an ~90° angle. The two incarnations of this motif are superposable (fig. S8B, all-atom RMSD = 1.29Å, without the 5´-phosphate and 3´-(U/G), RMSD = 0.14Å), though only the P6 triad appears to form a binding site for a magnesium ion. However, in superposing the two iterations of the motif, the helices into which they each dock – P1 and P6 – do not align as well, presumably because of local conformational and contextual differences between these helices, and slight differences in the angles with which the triads recognize their cognate binding sites.

With the exception of a single hydrogen bond (between A26 N1 and G-3 N3, and between A32 N1 and G74 N3), the contacts made by an A-minor triad could form irrespective of the sequence into which it docks. Hence, it is possible that this motif represents a simple structural module in which single-stranded RNA docks into the minor groove of a regular A-form duplex, although if so, it is exceedingly rare. A search through known RNA crystal structures revealed only a single additional instance – nucleotides 607-610 of the 16S rRNA (*T. thermophilus* numbering, fig. S8C). Perfect A-minor triads are observed at this position both in *T. thermophilus* (*S10*) and in *E. coli* (*S11*). Note that in the *T. thermophilus* A-minor triad, the three adenosines are followed by a guanosine, whereas in *E. coli* this last residue is a uridine. These observations, in addition to the repeated selection of a uridine or guanosine at the end of each motif in

active ligase or polymerase ribozyme variants (S1, S5, S7, S9), further support the notion that the A-minor triad motif preferentially terminates in a residue bearing a major-groove carbonyl. In addition, the 16S rRNA examples resemble the ligase P6 A-minor triad in that a nonbridging oxygen on the second adenosine binds to a solvent atom, around which the rest of the motif hooks. Hence, the propensity to form a metal-binding site may be a general trait of A-minor triad motifs (although given the ligase P1 triad, not a universal one).

Fidelity and Regioselectivity by the Class I Ligase Ribozyme

The electrophile for the self-ligation reaction is the triphosphoguanosine (G1) at the 5´ terminus of the ribozyme *(S12)*. G1 is deleted in polymerase ribozymes that perform primer extension reactions, its role performed instead by free nucleoside triphosphates (NTPs). Also deleted are the J1/2 residues tethering G1 to the ribozyme core. The crystal structure revealed highly specific contacts to J1/2, which could help orient G1 of the ligase and explain why primer extension with pppGGA, corresponding to G1–A3 of the ligase, is 10^3 fold more efficient than with GTP alone (*S13*). Although primer extension with NTPs is less robust in the absence of J1/2-mediated substrate tethering and positioning, it is nonetheless highly selective for Watson-Crick matching to the template (*S7, S14*). This fidelity might be explained in part by a high-occupancy hydrated magnesium ion bound to C12. C12 pairs to G1 and represents the template nucleotide that pairs with an incoming NTP during primer extension. As modeled, the C12-bound magnesium is held in place by a series of outer-sphere contacts with the O2 of C12 and the 2´-hydroxyls of C12 and U48 (Fig. 3B, table S3). Analogous contacts could form irrespective of the identity of the template nucleotide, and might help constrain the template residue relative to the active site. Such a constraint could disfavor primer extension by mismatched NTPs, which would pair with a geometry placing the α-phosphate outside the line of nucleophilic attack. Moreover, the 2´-hydroxyl of C47, which is located within 2.8–3.1Å of both the 2´- and 3´-hydroxyls of G1, could provide additional sequence-independent contacts favoring a Watson-Crick match to position 12 (Fig. 3B,C).

Although many RNA ligase ribozymes have been selected *in vitro*, the class I ligase is in the minority of species that employ the substrate 3´-hydroxyl as a nucleophile. The crystal structure of another such enzyme – the L1 ligase – has also been solved (*S15*). Given the structural evidence, it was hypothesized that this ribozyme achieves regioselectivity via an inhibitory water-mediated contact to the substrate 2´-hydroxyl, which is proposed to quench any alkoxide species that might develop on the 2´-position. Binding this quenching water was proposed to require that the nucleophile be situated within a G•U wobble pair. We observe no analogous interactions in the class I ligase structure; in fact, there is a remarkable dearth of interactions made to the A-1 nucleophile in general. Although we cannot preclude the possibility that the interactions responsible for regioselectivity have been lost during the course of catalysis due to a local conformation change, our favored model is that selection arises purely as a consequence of Watson-Crick pairing. Adenosine -1 and G1 are positioned in canonical base pairs within a continuously stacked A-form duplex, which naturally places the A-1 3´-hydroxyl proximal to the active site, orienting it for in-line attack on the G1 5´-α-phosphate. Reorienting A–1 so as to place its 2´-hydroxyl at the active site, however, would require gross deviation from A-form geometry, which would be energetically disfavored and could perturb the active site.

Recent crystallographic and biochemical work on the class II ligase – which almost exclusively catalyzes the formation of a 2´-5´ phosphodiester linkage (*S6*) – seems to further support the notion that regioselectivity is largely determined by the local sequence context situating the nucelophile and electrophile (*S16*). By examining the L1 and class II crystal structures, the authors were able to alter the regioselectivity of each ribozyme simply by changing the sequence of the base pairs abutting the ligation junction. Interestingly, in the L1 ligase, conversion of the G•U wobble pair orienting the nucleophile to a Watson-Crick pair has little perceptible effect on the regioselectivity, calling into question the proposed wobble-specific water-quenching mechanism. For both ribozymes, the most efficient conversion to the opposite selectivity was observed when mutations at the ligation junction were augmented by mutation of the base pair downstream of the ligation junction, implying that the positioning of the electrophile by adjoining bases plays a role in determining regiospecificity. In the class I ligase,

however, sequences downstream of the ligation junction do not play an essential role for regiospecificity. Primer extension with exogenous GTP, which occurs without downstream sequences with which to position the electrophile, maintains selection for the 3´ hydroxyl nucleophile (*S14*).

Also of note is that the L1 and class II ligases both position their electrophiles via a *trans* Hoogsteen-sugar A•G base pair, not via Watson-Crick pairing. This is identical to the G2-A11 pair observed in the class I ligase. It is interesting to consider that, had the register by which J1/2 docks into P2 been offset by one base – aligning A11 with G1 rather than with G2 – the class I ligase might have emerged as a 2´-5´ selective enzyme as well.

Materials and Methods

U1A Expression and Purification

The U1A A1-98 Y31H/Q36R double mutant (*S17, S18*) was expressed from the plasmid p11U1ADb. To construct this plasmid, the protein-coding sequence from a vector provided by Adrian Ferré-D'Amaré and Jennifer Doudna was inserted into the pET11a (Invitrogen) plasmid, thereby generating a vector that expressed more efficiently than its predecessor. For protein expression, a saturated culture of *E. coli* BL21(DE3) cells transformed with p11U1ADb was diluted 1:1000 into LB supplemented with 100 μg/mL ampicillin and grown in baffled flasks at 37º C with vigorous shaking. Protein overexpression was induced by the addition of 0.5 mM IPTG when cultures reached an OD_{600} of 0.7. Growth was continued at 37º C in the same fashion for another 2.5 hours, and cells were harvested by centrifugation and stored at –80º C before use. Pellets from a two-liter culture were resuspended in 40 mL of the lysis buffer described in (*S19*) and lysed by two passages through a French press at 4º C. Thereafter, purification was performed essentially as in (*S17*), following a detailed protocol provided by Kaihong Zhou of Jennifer Doudna's laboratory. Column fractions from the final purification step were assayed individually for the presence of contaminating ribonuclease by overnight incubation with ^{32}P body-labeled RNA and visualized by gel electrophoresis and

phosphorimaging. Fractions lacking nuclease were pooled, dialyzed into storage buffer (*S17*), concentrated to 15 g/L and stored at –80º C before use.

Crystallization Constructs

The crystals producing the native dataset were of an improved class I ligase (*S1*) (Fig. 1A), modified at the end of P5 with four additional base pairs terminating in the U1A-binding loop (Fig. 1B). DNA representing the unligated form of this ribozyme was subcloned into pUC19 (New England Biolabs) under a T7 transcription promoter, followed by a genomic hepatitis delta virus (HDV) self-cleaving ribozyme (*S20*) and an *EarI* restriction site, yielding the plasmid p307HU. The HDV sequence cleaves itself from the ribozyme 3´ terminus, thereby producing a homogenous ribozyme 3´ end. The relevant sequence of the insert was

GCGTAATACGACTCACTATA**G**GAACACTATACTACTGGATAATCAAAGACAAATCTGCC
CGAAGGGCTTGAGAACATACCC*ATTGCACTCCG*GGTATGCAGAGGTGGCAGCCTCCGGT
GGGTTAAAACCCAACGTTCTCAACAATAGTG**A**GGCCGGCATGGTCCCAGCCTCCTCGCT
GGCGCCGGCTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACCC
AC

in which the highlighted section indicates the T7 promoter, bold nucleotides denote the 5´ and 3´ ends of the ligase, underlined nucleotides are those that have been engineered to facilitate crystallization, and the italicized residues are the U1A-binding loop.

A second construct, pH307HP, was used to transcribe RNA representing the post-ligation product of the ribozyme. To ensure a homogenous 5´ terminus with the 5´-hydroxyl resembling that of the synthetic substrate, the transcript began with a hammerhead (HH) self-cleaving ribozyme (*S20*), which excises itself from the ribozyme. The relevant sequence of the insert was

GCGTAATACGACTCACTATAGGGAGATTCCTACTGGACTGATGAGTCCGTGAGGACGAA
ACGGTACCCGGTACCGTC**T**CCAGTAGGAACACTATACTACTGGATAATCAAAGACAAAT
CTGCCCGAAGGGCTTGAGAACATACCC*ATTGCACTCC*GGGTATGCAGAGGTGGCAGCCT
CCGGTGGGTTAAAACCCAACGTTCTCAACAATAGTG**A**GGCCGGCATGGTCCCAGCCTCC
TCGCTGGCGCCGGCTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGG
GACCCAC

annotated as above. Prior to use as templates for *in vitro* transcription, plasmids were digested with *EarI* nuclease.

RNA Synthesis and Purification

T7 *in vitro* transcription reactions were as described (*S5*), in volumes of 3–10 mL, with 20 μg/mL *EarI*-linearized template plasmid. After 2.5 hours, reactions were quenched by the addition of EDTA, extracted with phenol and chloroform, then ethanol precipitated. Pellets were resuspended in water and desalted with RNase-free P30 gel-filtration spin-columns (BioRad). To ensure complete processing of the self-cleaving ribozyme(s), the effluent was heated and cooled (85º C and 37º C, 5 minutes each), brought to one-fifth the original transcription volume in HDV Buffer (30 mM Tris, pH 7.4, 10 mM $MgCl_2$, 200 mM KCl), and incubated for 45 minutes at 37º C. These reactions were quenched with EDTA, ethanol precipitated, and resuspended in a minimal volume of water.

RNA that had been transcribed as the unreacted ligase species (from the p307HU template) was diluted with urea (~6 M final urea concentration), and loaded onto preparative-scale denaturing 6% polyacrylamide TBE/Urea gels (33 cm width, 42 cm height, 4.76 mm spacers). Following electrophoresis, bands were visualized by UV-shadowing and excised, and RNA was isolated either by electro-elution into 0.5X TBE (Elutrap System, Schleicher and Schuell), or by passive elution into 300 mM NaCl at 4º C for two days, harvesting material after each day. RNA was ethanol precipitated and resuspended in water to a final concentration of 5 μM. To form the product species, this material was preincubated (80º C for 5 minutes, 22º C for 10 minutes) then reacted with the substrate oligonucleotide 5´-UCCAGUA- 3´ (Dharmacon), initiating the reaction with simultaneous addition of buffer, salts and substrate (final concentrations, 1 μM ligase, 2 μM substrate, 50 mM MES, pH 6.0, 10 mM $MgCl_2$, 50 mM KCl, 1 μM EDTA). After 30 minutes at 22º C, the reaction was quenched with EDTA, ethanol precipitated and resuspended in ~6 M urea. To enable separation of the ligated product (137 nt) from unreacted ribozyme (130 nt), material was loaded onto multiple preparative 6% polyacrylamide TBE/urea gels (same dimensions as above, 0.5 mg RNA/gel), and subjected to 12-15 hours of electrophoresis. Product was visualized and purified as for the unligated RNA. Following the final precipitation, pellets were air-dried to remove

residual ethanol, resuspended in minimal water, filtered to remove any particulate matter (0.22μm Centrex filters, Schleicher and Schuell), and stored at –20º C.

RNA used to prepare derivative crystals was transcribed as the ligated product species from the pH307HP plasmid and treated to remove the 2´-3´ cyclic phosphate left by cleavage of the HDV ribozyme. (This material was prepared in the course of exploring an alternate crystal form for which crystallization was sensitive to the state at the 3´ terminus.) Following HDV processing, RNA was desalted, diluted to 10 μM in water and heat/cooled (85º C and 37º C, 5 minutes each). The cyclic phosphate was removed by T4 polynucleotide kinase (PNK) treatment, using "method iii" described in (*S21*), except that RNA was diluted to a final concentration of 5 μM, and PNK (New England Biolabs) was at 0.5 U/μL. After six hours at 37º C, the reaction was quenched with EDTA, extracted with phenol and chloroform, and ethanol precipitated. The pellet was resuspended in water, and material from up to 5mL of initial transcription was separated on a single preparative-scale denaturing 6% polyacrylamide TBE/urea gel. After electrophoresis, purification was identical to that described above for ligated p307HU-transcribed material.

To inspect for full removal of the 2´-3´ cyclic phosphate, we developed an assay appropriate for larger RNAs, which cannot be evaluated using the gel-mobility-shift assays useful for shorter RNA species. Following T4 PNK treatment, 25 pmol of RNA was desalted and incubated at 2.5 μM final concentration with 750 U of Yeast Poly(A) Polymerase (USB), in the manufacturer's supplied buffer, supplemented with 1 mM ATP. After 30 minutes at 37º C, reactions were quenched with EDTA, extracted with phenol, diluted with urea, boiled and separated on a diagnostic-scale 6% polyacrylamide TBE/urea gel. RNA molecules terminating in a *cis*-diol were extended by poly(A) addition to produce longer species, whereas RNA not treated with PNK did not shift (fig. S11A). As expected, this assay was sensitive to the covalent state of both the 2´ and 3´ hydroxyls: opening the cyclic phosphate by acid-catalyzed hydrolysis (resulting in a mixture of 2´- and 3´ linear mono-phosphates (*S22*)) produced RNAs that were unsuitable substrates for poly(A) extension, whereas further alkaline-phosphatase treatment (Roche) to remove the linear phosphate (*S22*) allowed poly(A) extension of nearly all material (fig. S11B).

Crystallization

Approximately 200 μM RNA was heated (80º C, 5 minutes) and cooled (22º C, 10 minutes) in water, and then mixed with annealing buffer (final composition, 5 mM MES, pH 6.0, 10 mM $MgCl_2$, 1 mM DTT), and incubated at 22º C for an additional 15 minutes. U1A was added at a 1:1 molar ratio of RNA:protein to bring the final complex concentration to 5.2 g/L. This mixture was incubated at 22º C for 45–60 minutes, mixing periodically, then centrifuged at 13,000×g for 1 minute prior to setting up crystallization experiments.

Initial crystals were obtained by hanging drop vapor diffusion, mixing 1 μL of the ligase:U1A complex sample with 0.5 μL mother liquor consisting of 20–30% (v/v) 2-Methyl-2,4-pentane-diol (MPD), 50 mM sodium cacodylate (pH 6.0–7.0), 30-40 mM magnesium acetate and 1 mM spermine, and equilibrating over 0.6 mL of this same precipitation mixture at 20º C. Crystals appeared over 2–3 days, and grew to full size (~70 μm per side) within a week, most often with an intractable habit (inseparable clusters or stacks of plates). Only ~1% experiments performed in this fashion yielded usable crystals.

A systematic screen of additive compounds (Hampton Research) revealed that addition of 50 mM KCl to the crystallization mixture slightly diminished the degree of crystal clustering, whereas 100 mM KCl ablated nucleation altogether. We therefore exploited these properties, using microseeding to obtain morphologically tractable, diffraction-quality crystals more reproducibly. Seed-producing crystal clusters were grown by mixing 0.5 μL Ligase:U1A complex with an equal volume of 22–26% (v/v) MPD, 50 mM sodium cacodylate (pH 6.0), 40 mM magnesium acetate, 50 mM KCl and 1 mM spermine, and equilibrating over 0.6 mL of this precipitation mixture. Drops containing viable crystals were stabilized by bringing them to 36% MPD, keeping all other buffer components (including those derived from the RNA-annealing and U1A-storage buffers) isotonic. Crystal clusters were crushed using a Seed-Bead kit (Hampton Research), and the resultant seed suspensions were serially diluted in the same stabilization buffer. Diffraction-quality crystals were obtained by using these serially diluted seed stocks directly as precipitant under conditions that would otherwise prohibit

crystal nucleation: 0.7 μL of Ligase:U1A complex was prepared in annealing buffer supplemented with 100 mM KCl, mixed with 0.5 μL of the seed-stock dilution, and were equilibrated the drops over 0.6 mL wells of 22–26% MPD. Using this method, roughly one out of every three drops produced at least one high-quality crystal (50–100 μm in all dimensions) within a week. For cryoprotection, drops bearing suitable crystals were brought to 30% (v/v) MPD, again keeping all other buffer compositions isotonic, and incubated against a well of 30% MPD for 2–24 hours before mounting in a nylon loop and plunging directly into liquid nitrogen.

Early crystallization efforts started with RNA that had been transcribed from p307HU and had undergone the self-ligation reaction. It was this RNA that produced the crystal used in our native data set. We subsequently found that RNA transcribed from pH307HP produced crystals under identical conditions, and that crystal-to-crystal isomorphism was observed independent of the RNA-production strategy. We therefore used the RNA transcribed from pH307HP, as this strategy involved only one gel purification step and gave higher yields. Derivative crystals were prepared by first cryo-stabilizing native crystals at 30% (v/v) MPD for 2 hours as above, and then soaking them for 16–24 hours in otherwise isotonic solutions containing 5mM cobalt hexamine (Sigma), or iridium hexamine (gift of Robert Batey). In an effort to maximize isomorphism, crystals soaked in each of these compounds were derived from the same parental seed stock. Crystals were harvested directly from the soaking solution with a nylon loop and plunged directly into liquid nitrogen.

Data Collection and Processing

Native and derivative data sets were collected at NE-CAT beamlines 8-BM and 24-ID at the Advanced Photon Source (APS), respectively, the latter aided by the technical advice of K.R. Rajashankar. This work is based upon research conducted at the Northeastern Collaborative Access Team (NE-CAT) beamlines of the APS, supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

All data were integrated and scaled using the HKL2000 software suite (*S23*). Initial phases were obtained from a single isomorphous replacement with anomalous scattering (SIRAS) experiment, using a cobalt hexamine-soaked crystal as a "pseudo-native," and an iridium hexamine-soaked crystal as a derivative, similar to the method described in (*S24*) (table S1). Nineteen initial iridium sites were found using SHELXD, and a preliminary round of solvent flattening was performed in SHELXE (*S25*); both processes were aided by the hkl2map graphical interface (*S26*). The resultant initial electron density map (fig. S1B) contained clear RNA A-form density, as well as conspicuous proteinaceous features from U1A.

Model Building and Refinement

Two monomers corresponding to the U1A protein bound to its cognate loop and a seven base-pair stem (analogous to P5 in the crystallization construct) were located the original experimental density using the program MOLREP (*S27*) in the CCP4 program suite (*S28*), using a search model derived from the hairpin ribozyme-inhibitor crystal structure (*S29*) (PDB ID: 1HP6). Further rounds of solvent flattening were performed in the program SHARP (*S30*), using σA-weighted, combined phases from the initial heavy atom sites and this partial molecular replacement solution, iteratively adjusting the solvent content by hand and inspecting the quality of the resultant electron density map. This yielded modestly improved maps, restoring some phosphate and base-density which had been truncated by the initial SHELXE output.

An initial model was built into this experimental 3.36Å density by real-space refinement in the program COOT (*S31*), starting with 3–4-nucleotide stretches of A-form polycytidine and building into areas with clear RNA-backbone density (*S15*). Each round of poly(C) model building was followed by a round of restrained, individual-site, isotropic Atomic Displacement Parameter (ADP) refinement with automated bulk-solvent correction (termed the "individual ADP" strategy in PHENIX (*S32*)). Refinement was against experimental structure factor amplitudes and phases, using a maximum likelihood target. Given the limited resolution of the data, target geometric weights ("wxc_scale" in PHENIX) were set to a relatively restrictive setting of 0.05 and B-factors were initially refined per-residue. The "wxu_scale" was kept constant at 1.0 during all refinement

rounds (*S32*). In early stages, ADP refinement was followed by simulated annealing (5000K to 300K in 100K steps), also implemented in PHENIX. The two monomers were built individually and refined without the use of noncrystallographic symmetry (NCS) averaging. Once ~60% of the RNA sequence had been placed in this fashion, experimental phases were discarded and ADP refinement with automated bulk solvent correction continued using model-based phases and native amplitudes at 2.99Å, now allowing refinement of individual atomic B-factors. Inspection of the $|F_{obs}| - |F_{calc}|$ difference maps from early rounds of refinement at higher resolution resulted in the unambiguous assignment of individual bases as purine or pyrimidine, and in many cases guanosines could be distinguished from adenosines. This, along with the position of the U1A protein and P5 stem, allowed establishment of the RNA sequence register. In later stages of refinement, NCS averaging was used for residues 20–28, where the electron density map was noisiest, and the resulting B-factors the highest (fig. S2).

Parallel experiments using a TLS refinement in addition to ADP refinement failed to improve R_{free} values and the quality of $2|F_{obs}|-|F_{calc}|$ maps, and revealed no additional features in $|F_{obs}| - |F_{calc}|$ difference maps. Hence, the final structure factors and model were calculated entirely without the use of TLS refinement.

After placement of all 137 RNA nucleotides for each ligase monomer and all but the first 6-7 protein residues in each U1A protomer, solvent atoms were placed into peaks of 4σ or greater in the resultant $|F_{obs}| - |F_{calc}|$ difference map and further refined; this process of refinement and solvent placement was reiterated until the R_{free} value reached a local minimum. Owing to the resolution limit, magnesium and water atoms could not be individually refined in a chemically sensible manner, so partially- or fully hydrated magnesium clusters were treated as individual monomers during refinement, having defined model residues with ideal bond lengths and geometry. In the final rounds of refinement, the bond distances (2.07Å) and angle (90°) between backbone phosphate nonbridging oxygens at positions A31 and A32 were fixed with respect to the magnesium ion they coordinate (Fig. 2B, 3A); all other clusters were unrestrained relative to RNA or protein. Side-chain atoms for a number of the U1A residues have been removed in cases for which density in our final $2|F_{obs}| - |F_{calc}|$ maps was apparent for the backbone only.

Despite having not enforced global NCS averaging, the two copies of the RNA are nearly identical: excluding residues in the U1A loop, their all-atom RMSD is 0.92Å.

The final model is in excellent agreement with a large body of biochemical data accumulated on the ligase (*S1*). In addition, per-residue solvent accessibilities for all RNA C4´ carbons (calculated using AREAIMOL in the CCP4 program suite, (*S28*)) correlate with observed Fe•EDTA solvent protection data at these positions (fig. S12, (*S33*)), further supporting the validity of our model.

Simulated annealing OMIT maps (Fig 3B, D) were calculated in PHENIX, without additional ADP refinement. Simulated annealing composite OMIT maps (fig. S1C) were calculated in CNS (*S34*), with blocks of 5% unit cell volume omitted from each calculation. Though not quoted in table S1, the R_{work} and R_{free} values calculated in CNS are in good agreement with those calculated in PHENIX (23.0 and 25.7%, respectively).

Real space R-values, σA and Luzzati estimated coordinate errors (table S1) were calculated in CNS. The lengths of potential hydrogen bonds (table S2) and metal ion contacts (table S3) were measured without modification of our model. However, in order to measure hydrogen bond angles (table S2), hydrogens were added automatically using MolProbity (*S35*); where necessary, 2´-hydroxyls were rotated by hand into the optimal hydrogen-bonding orientation. Structural alignments (figs. S3, S6, S7 and S8) were made using the least-squares all-atom alignment function in COOT (*S31*). All structure figures were made with the program PyMOL (*S36*).

Biochemical Experiments

Mutant ribozymes were generated from the U1A-modified construct, p307HU (QuickChange mutagenesis kit, Stratagene). For kinetic assays, body-labeled RNA was heated and cooled (80° C, 22° C, 5 minutes each) in water, and the reaction was initiated by the simultaneous addition of buffer, salts and an excess of oligonucleotide substrate (final concentrations, 1 μM ligase, 2 μM substrate, 50 mM Tris, pH 7.0, 10 mM $MgCl_2$, 50 mM KCl, 1.0 μM EDTA). Reactions were incubated at 22°C, and aliquots were removed at time points and quenched by mixing into a two-fold excess of gel loading buffer (8 M urea, 120 mM EDTA, trace xylene cyanol and bromphenol blue). Samples

were boiled and separated on 6% polyacrylamide TBE/urea sequencing gels, imaged and quantified by phosphorimaging (Fujifilm BAS-2500). For each time point, the fraction product was measured as F_P = Product/(Product + Reactant), and fitted to the curve:

$$F_P(t) = F_M(1 - e^{-k_{OBS}t}),\ t = \text{time}$$

with F_M (the maximum fraction reacted), and k_{OBS} (the observed rate constant) treated as unknowns. Fits were performed using the least squares method, implemented in KaleidaGraph (Synergy Software).

Nucleotide analog interference mapping (NAIM) was performed as described (*S1, S37*), starting with a pool of randomly phosphorothioate-modified RNAs and allowing them to label themselves with ^{32}P-labeled substrate RNA, under solvent and time constraints that selected for only the most active molecules.

Fab-Ligase Crystal Structure

Selection and analysis of the ligase Fab, as well as crystallization and structure solution of it complexed with the ligase (*S38*) shall be presented elsewhere.

Supplemental References and Notes:

S1. S. C. Bagby, N. H. Bergman, D. M. Shechner, C. C. Yen, D. P. Bartel, *RNA* **15**, published online (2009).
S2. G. J. Quigley, A. Rich, *Science* **194**, 796 (1976).
S3. H. W. Pley, K. M. Flaherty, D. B. McKay, *Nature* **372**, 68 (1994).
S4. M. Martick, W. G. Scott, *Cell* **126**, 309 (2006).
S5. E. H. Ekland, D. P. Bartel, *Nucleic Acids Res* **23**, 3231 (1995).
S6. E. H. Ekland, J. W. Szostak, D. P. Bartel, *Science* **269**, 364 (1995).
S7. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel, *Science* **292**, 1319 (2001).
S8. L. Su, L. Chen, M. Egli, J. M. Berger, A. Rich, *Nat Struct Biol* **6**, 285 (1999).
S9. H. S. Zaher, P. J. Unrau, *RNA* **13**, 1017 (2007).
S10. B. T. Wimberly *et al.*, *Nature* **407**, 327 (2000).
S11. B. S. Schuwirth *et al.*, *Science* **310**, 827 (2005).
S12. D. P. Bartel, J. W. Szostak, *Science* **261**, 1411 (1993).
S13. M. E. Glasner, C. C. Yen, E. H. Ekland, D. P. Bartel, *Biochemistry* **39**, 15556 (2000).
S14. E. H. Ekland, D. P. Bartel, *Nature* **382**, 373 (1996).
S15. M. P. Robertson, W. G. Scott, *Science* **315**, 1549 (2007).
S16. J. N. Pitt, A. R. Ferré-D'Amaré, *J Am Chem Soc* **131**, 3532 (2009).
S17. A. R. Ferré-D'Amaré, J. A. Doudna, *J Mol Biol* **295**, 541 (2000).
S18. C. Oubridge, N. Ito, C. H. Teo, I. Fearnley, K. Nagai, *J Mol Biol* **249**, 409 (1995).
S19. P. Pognonec, H. Kato, H. Sumimoto, M. Kretzschmar, R. G. Roeder, *Nucleic Acids Res* **19**, 6650 (1991).
S20. S. R. Price, N. Ito, C. Oubridge, J. M. Avis, K. Nagai, *J Mol Biol* **249**, 398 (1995).
S21. H. Schürer, K. Lang, J. Schuster, M. Morl, *Nucleic Acids Res* **30**, e56 (2002).
S22. J. D. Pata, B. R. King, T. A. Steitz, *Nucleic Acids Res* **30**, 4855 (2002).
S23. Z. Otwinowski, W. Minor, *Methods Enzymol.* **276**, 307 (1997).
S24. J. C. Cochrane, S. V. Lipchock, S. A. Strobel, *Chem Biol* **14**, 97 (2007).
S25. G. M. Sheldrick, *Acta Crystallogr A* **64**, 112 (2008).
S26. T. Pape, T. R. Schneider, *J. Appl. Cryst.* **37**, 843 (2004).
S27. A. Vagin, A. Teplyakov, *J. Appl. Cryst.* **30**, 1022 (1997).
S28. Collaborative computation project number 4, *Acta Crysallogr. D. Biol. Crystallogr.* **50**, 760 (1994).
S29. P. B. Rupert, A. R. Ferré-D'Amaré, *Nature* **410**, 780 (2001).
S30. E. DeLaFortelle, G. Bricogne, *Methods. Enzymol.* **276**, 472 (1997).
S31. P. Emsley, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **60**, 2126 (2004).
S32. P. D. Adams *et al.*, *Acta Crystallogr D Biol Crystallogr* **58**, 1948 (2002).
S33. N. H. Bergman, N. C. Lau, V. Lehnert, E. Westhof, D. P. Bartel, *RNA* **10**, 176 (2004).
S34. A. T. Brünger *et al.*, *Acta Crystallogr D Biol Crystallogr* **54**, 905 (1998).
S35. I. W. Davis *et al.*, *Nucleic Acids Res* **35**, W375 (2007).
S36. W. L. Delano, *http://pymol.sourceforge.net*, DeLano Scientific (2002).
S37. S. A. Strobel, *Curr Opin Struct Biol* **9**, 346 (1999).
S38. The coordinates of the ligase complexed with the antibody fragment are available (PDB ID: 3IVK).
S39. A. R. Ferré-D'Amaré, K. Zhou, J. A. Doudna, *Nature* **395**, 567 (1998).
S40. S. D. Gilbert, R. P. Rambo, D. Van Tyne, R. T. Batey, *Nat Struct Mol Biol* **15**, 177 (2008).

S41. C. Lu *et al.*, *Nat Struct Mol Biol* **15**, 1076 (2008).
S42. R. T. Batey, S. D. Gilbert, R. K. Montange, *Nature* **432**, 411 (2004).
S43. A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, *Nature* **441**, 1167 (2006).
S44. R. K. Montange, R. T. Batey, *Nature* **441**, 1172 (2006).
S45. A. Serganov, L. Huang, D. J. Patel, *Nature* **458**, 233 (2009).
S46. C. E. Dann, 3rd *et al.*, *Cell* **130**, 878 (2007).
S47. A. Serganov, L. Huang, D. J. Patel, *Nature* **455**, 1263 (2008).
S48. M. P. Robertson *et al.*, *PLoS Biol* **3**, e5 (2005).
S49. H. Shi, P. B. Moore, *RNA* **6**, 1091 (2000).
S50. T. Hainzl, S. Huang, A. E. Sauer-Eriksson, *RNA* **11**, 1043 (2005).
S51. J. S. Pfingsten, D. A. Costantino, J. S. Kieft, *Science* **314**, 1450 (2006).
S52. A. S. Krasilnikov, X. Yang, T. Pan, A. Mondragón, *Nature* **421**, 760 (2003).
S53. A. S. Krasilnikov, Y. Xiao, T. Pan, A. Mondragón, *Science* **306**, 104 (2004).
S54. J. H. Cate *et al.*, *Science* **273**, 1678 (1996).
S55. A. Serganov *et al.*, *Nat Struct Mol Biol* **12**, 218 (2005).
S56. H. Xiao, T. E. Edwards, A. R. Ferré-D'Amaré, *Chem Biol* **15**, 1125 (2008).
S57. H. Xiao, H. Murakami, H. Suga, A. R. Ferré-D'Amaré, *Nature* **454**, 358 (2008).
S58. S. Rudisser, I. Tinoco, Jr., *J Mol Biol* **295**, 1211 (2000).
S59. P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, T. A. Steitz, *Proc Natl Acad Sci USA* **98**, 4899 (2001).

Table S1. Crystallographic statistics for the class I ligase product

Crystal	Native	Ir-Hex	Co-Hex
Data Collection			
Space group	P1	P1	P1
Cell dimensions			
a,b,c (Å),	59.56, 70.91, 70.41	59.52, 71.65, 71.30	59.59, 71.10, 70.94
α,β,γ (°)	100.30, 103.89, 99.61	101.14, 104.16, 99.72	100.58, 104.13, 99.71
Wavelength (Å)	0.9795	1.1053	1.1053
Resolution (Å)[a]	50-2.99 (3.11-2.99)	50-3.60 (3.73-3.60)	50-3.50 (3.63-3.50)
R_{merge} (%)[a]	7.3 (22.5)	12.1 (18.1)	8.4 (20.6)
Mean I/σ(I)[a]	20.2 (5.3)	11.8 (6.4)	16.7 (5.3)
Unique Observations[a]	21,093 (2,306)	12,353 (1,183)	13,188 (1116)
Completeness (%)[a]	98.7 (87.6)	97.3 (91.0)	96.0 (79.2)
Redundancy[a]	3.6 (3.0)	3.7 (3.1)	6.8 (4.5)
SIRAS Phasing			
Number of heavy-atom sites		19	
Phasing power			
Isomorphous (acentrics)[a]		1.506 (0.474)	
Anomalous(acentrics)[a]		0.754 (0.295)	
R_{Cullis} (%)			
Isomorphous (acentrics)[a]		68.6 (97.2)	
Anomalous(acentrics)[a]		82.2 (99.6)	
Mean Figure of Merit[a]		0.339 (0.125)	
Refinement			
R_{Work} (%)[b]	20.18		
R_{Free} (%)[b]	23.35		
Number of atoms			
RNA	5844		
Protein	1475		
Mg^{2+}	34		
Water	192		
Mean B-factors(Å^2)			
RNA	79.3		
Protein	70.2		
Mg^{2+}	88.8		
Water	77.0		
Mean Real-Space R-factors (%)			
RNA	15.4		
Protein	16.9		
Mg^{2+} + Water	8.6		
R.m.s. deviations			
Bond distances (Å)	0.002		
Angles (°)	0.795		
Estimated coordinate error (Å)[c]:			
Luzzati[d]	0.41 (0.42)		
σA[d]	0.44 (0.50)		
Maximum Likelihood	0.33		

[a]Values in parenthesis refer to the highest resolution shell. [b]5% of all data were excluded from refinement for R_{Free} calculations. [c]Luzzati and σA as calculated in CNS (*S34*); Maximum likelihood as calculated in PHENIX (*S32*). [d]Numbers in parenthesis refer to the cross-validation test set.

Table S2. Lengths and estimated angles for non-solvent-mediated, non-Watson-Crick hydrogen bonds, grouped by secondary structure. All quoted values are the average of observations for each of the two RNA copies in the asymmetric unit; errors report the difference between individual monomer values and the average. Distances are those between hydrogen bond donor and acceptor nuclei. Angles are those centered on the hydrogen atom participating in each bond (*i.e.* the angle: Donor─H•••Acceptor), as modeled in MolProbity *(S35)*.

Residues (Atoms)		Distance (Å)	Angle (°)
J1/2:			
Guanosine 2			
(N2)	A11 (N7)	3.28 ± 0.17	163.70 ± 9.90
(N3)	A11 (N6)	3.07 ± 0.26	143.30 ± 9.20
(O2')	A11 (N6)	2.98 ± 0.25	166.55 ± 6.25
Adenosine 3			
(N1)	C113 (O2')	2.78 ± 0.13	161.90 ± 4.50
(N3)	G46 (N2)	3.34 ± 0.15	130.80 ± 1.50
(O2')	G36 (N3)	2.74 ± 0.02*	155.75 ±12.35
Adenosine 4			
(Pro-Rp O)	A114 (O2')	2.91 ± 0.00	157.50 ± 5.80
J1/3:			
Adenosine 22			
(N1)	C-6 (O2')	3.27 ± 0.01	162.70 ± 0.10
(N3)	G18 (N2)	3.20 ± 0.00	133.70 ± 0.10
(O2')	G18 (N3)	3.00 ± 0.00	144.10 ± 0.00
(O2')	G18 (O2')	1.67 ± 0.00	146.10 ± 0.08
Uridine 23			
(Pro-Rp O)	G17 (N2)	3.19 ± 0.00	157.85 ± 0.15
(O4)	G17 (O2')	3.70 ± 0.00	149.50 ± 0.30
Cytidine 24			
(N4)	G17 (O2')	2.66 ± 0.00	112.70 ± 0.10
Adenosine 25			
(N1)	U16 (O2')	3.65 ± 0.00	166.05 ± 9.55
(N6)	U16 (O2)	3.70 ± 0.00	152.80 ± 0.00
Adenosine 26			
(N1)	G-3 (N2)	3.15 ± 0.00	154.15 ± 0.15
(N6)	G-3 (N3)	3.17 ± 0.00	157.45 ± 0.15
(N6)	G-3 (O2')	3.66 ± 0.00	133.10 ± 0.10
Adenosine 27			
(N1)	G-3 (O2')	3.88 ± 0.00	127.65 ± 0.95
Adenosine 29			
(N1)	G72 (O2')	2.66 ± 0.01	147.80 ± 3.60
(N3)	G72 (N2)	3.43 ± 0.04	168.75 ± 1.15
(N6)	A71 (O4')	3.15 ± 0.00	135.20 ± 0.50
(O2')	C85 (O2')	3.77 ± 0.01	137.80 ±15.30
Cytidine 30			
(N4)	C47 (O4')	2.97 ± 0.03	105.00 ± 0.00
(O2')	C47 (N4)	3.60 ± 0.12	134.00 ± 0.60
(O4')	A73 (O2')	3.49 ± 0.01	111.35 ± 1.95
Adenosine 31			
(N1)	U84 (O2')	2.74 ± 0.00	146.60 ± 1.00
(N6)	U84 (O2)	3.02 ± 0.01	160.00 ± 1.20
Adenosine 32			
(N1)	G74 (N2)	3.35 ± 0.05	159.10 ± 0.60
(N6)	G74 (N3)	3.35 ± 0.06	159.00 ± 3.20
(N6)	G74 (O2')	3.46 ± 0.07	112.80 ± 1.40
Adenosine 33			
(N1)	G74 (O2')	2.80 ± 0.09	157.55 ±13.05

Residues (Atoms)		Distance (Å)	Angle (°)
J3/4:			
Guanosine 41			
(N2)	G44 (pro-Sp O)	2.76 ± 0.06	120.60 ±14.40
(O2')	A43 (N7)	3.04 ± 0.21	159.40 ±11.70
Guanosine 44			
(O2')	U76 (O2')	2.80 ± 0.23	158.25 ±15.05
Guanosine 45			
(O2')	C5 (pro-Rp O)	3.09 ± 0.48	136.55 ±10.35
Cytidine 47			
(N4)	G1 (pro-Sp O)	3.06 ± 0.01	126.85 ± 0.65
(O2')	G1 (O2')	2.85 ± 0.02	148.15 ± 0.55
(O2')	G1 (O3')	3.11 ± 0.04	127.35 ± 5.65
J5/6:			
Adenosine 71			
(N1)	C86 (O2')	2.67 ± 0.02	155.60 ±12.70
(N3)	G105 (N2)	3.65 ± 0.01	110.12 ± 4.12
(O4')	G72 (O2')	3.27 ± 0.00	119.45 ± 0.15
L7:			
Uridine 93			
(O2)	A98 (N6)	3.82 ± 0.90	165.90 ± 0.40
Uridine 94			
(N3)	A97 (pro-Rp O)	3.28 ± 0.45	157.85 ± 7.65
(O2')	A96 (N7)	2.88 ± 0.09	168.61 ± 6.49

Table S3. Potential ligands and bond distances for partially hydrated magnesium ions described in the text. Listed values are the average of observations for each of the two RNA copies in the asymmetric unit; errors report the difference between individual monomer values and the average.

	Potential Ligands	Distance (Å)
"P6 A-Minor Triad" $Mg^{2+}\cdot 4H_2O$[a]		
Inner Sphere Ligands		
	C30 (pro-Rp O)	2.10 ± 0.01
	A31 (pro-Rp O)	2.09 ± 0.00
Outer Sphere Ligands		
	A32 (N7)	2.83 ± 0.07
	A33 (N7)	3.17 ± 0.31
	N33 (N6)	3.34 ± 0.26
	U34 (O4)	3.31 ± 0.31
"Template" $Mg^{2+}\cdot 6H_2O$[b]		
Outer Sphere Ligands		
	G1 (N2)	2.83 ± 0.04
	G2 (O2´)	3.24 ± 0.12
	C12 (O2)	2.58 ± 0.00
	C12 (O2´)	2.79 ± 0.02
	U13 (O4´)	3.10 ± 0.34
	U48 (O2´)	3.04 ± 0.28

[a]See Figs. 2B, 3A, 3C. [b]See Fig. 3B.

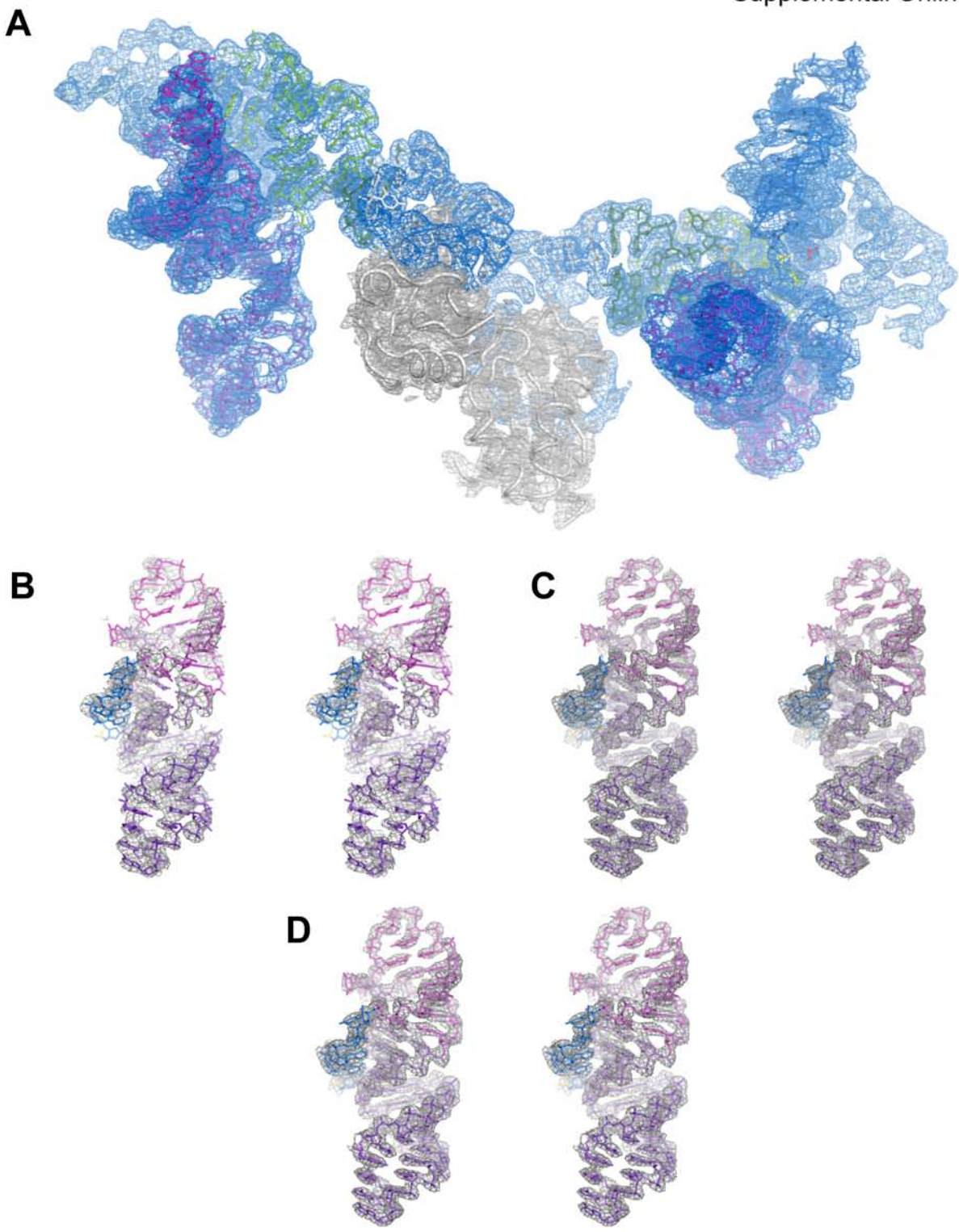

Fig. S1. The model and accompanying electron density maps. (**A**) View of the complete contents of the asymmetric unit, highlighting a crystallographic dimer between copies of U1A. This is juxtaposed with the final $2|F_{OBS}| - |F_{CALC}|$ electron density map, drawn in blue surrounding the RNA and in gray surrounding U1A. The two ligase monomers are rendered in sticks; the protein is shown as a cartoon backbone trace; colors are as in Fig 1. The electron density map is contoured at 1.2σ. (**B**-**D**) Wall-eyed stereo depictions of the 3′-end of J1/3 docking into the P3-P6-P7 domain, juxtaposed with (**B**) the initial 3.36Å solvent-flattened experimental electron density map, contoured at 1σ; (**C**) a simulated-annealing composite-OMIT map, contoured at 1σ; (**D**) the final refined 3.0Å $2|F_{OBS}| - |F_{CALC}|$ map, contoured at 1.4σ.

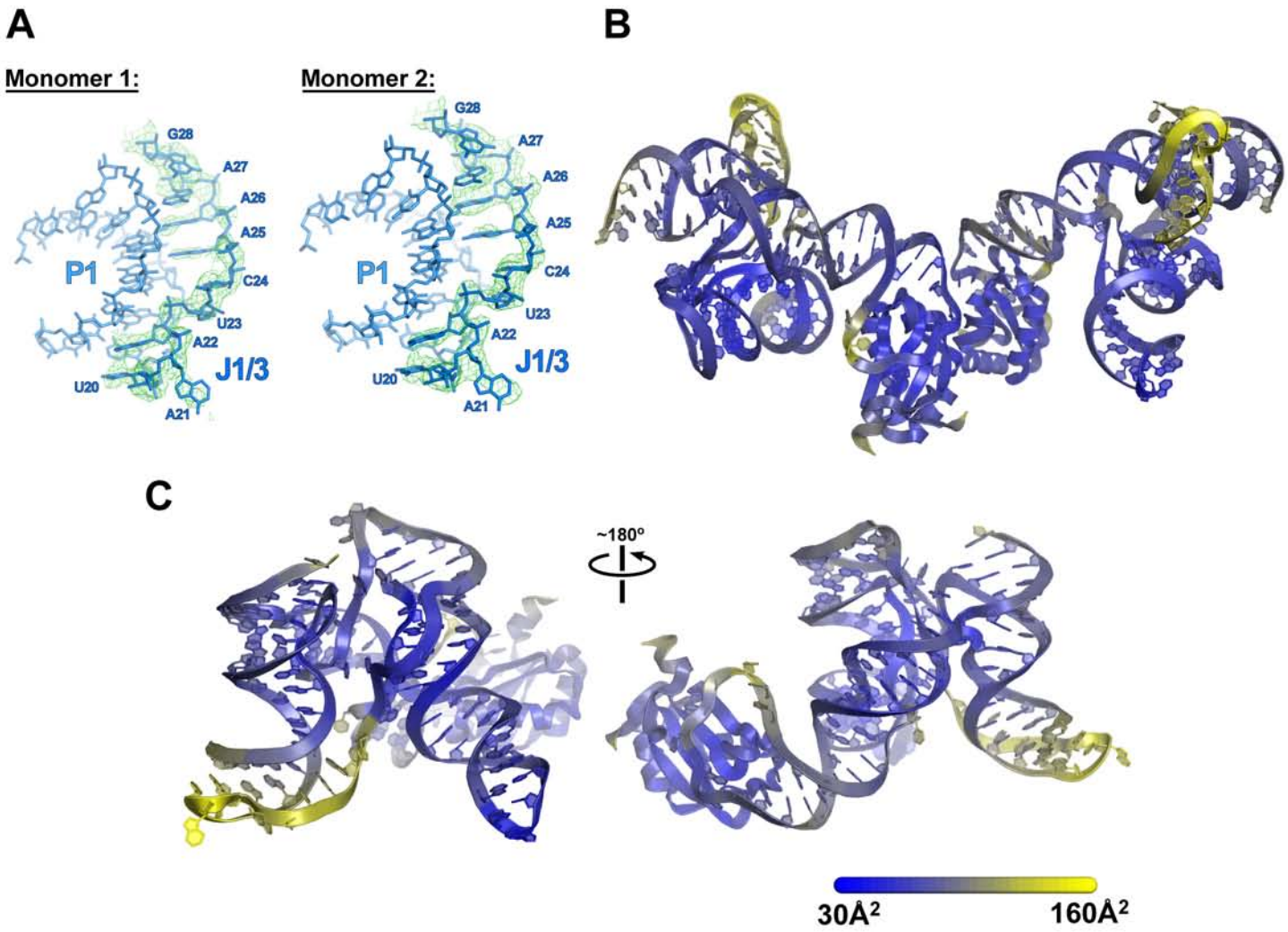

Fig. S2. Disorder in the 5´-half of J1/3, and distribution of B-factors in the refined ligase model. (**A**) Views of the first nine residues of J1/3 (U20-G28) for each of the copies in the asymmetric unit. Green mesh corresponds to the final $2|F_{OBS}| - |F_{CALC}|$ electron density map, contoured at 1σ. Density in this region was the weakest observed in any part of the asymmetric unit. (**B**) Global view of the complete contents of the asymmetric unit, as in fig. S1A, colored by final B-factor. (**C**) Two views of a single Ligase-U1A complex, separated by ~180° rotation. The left-hand view corresponds to that of Fig. 1C. The temperature factor scale bar (bottom) is the same for both **B** and **C**.

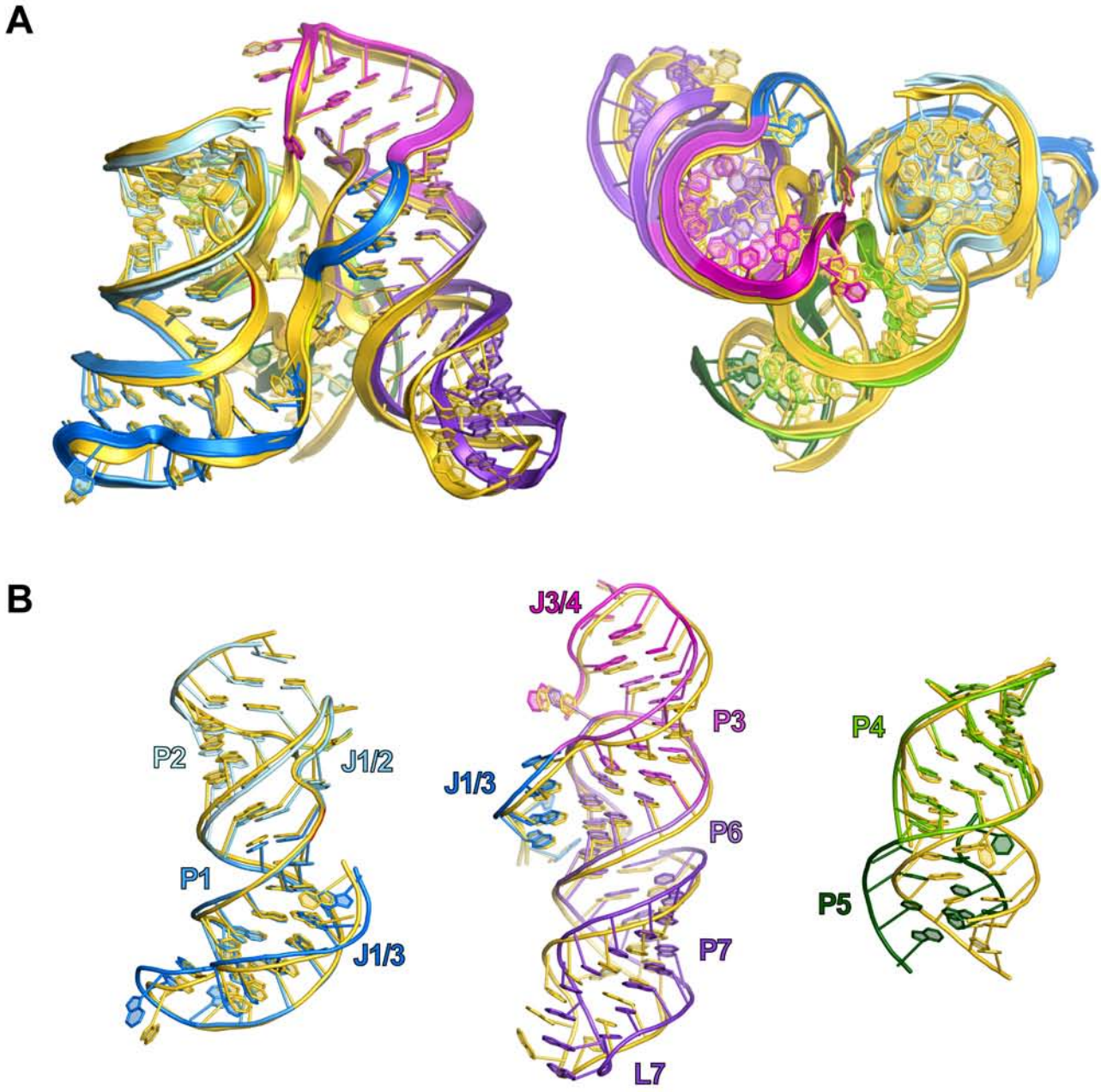

Fig S3. Superposition of class I ligase structures independently solved from different crystalline environments. (**A**) Two views of an all-atom alignment between the U1A-bound structure (colored and oriented as in Fig 1C,D) and the 3.15Å structure obtained in complex with a ligase-specific antibody selected through phage-display (gold) *(S38)*. Both protein modules bind to L5, which was omitted during the alignment and has been removed from view. Each crystal form contains two copies in its asymmetric unit. Because the four possible pair-wise alignments between each monomer of each crystal form were nearly indistinguishable, an alignment is shown for only one copy from each crystal. Note that the greatest points of divergence between the two models are found in regions involved in crystal-contact formation: P5 (which binds to the protein crystallization modules) and L7 (which is involved in an RNA-protein crystal contact in the U1A-bound structure). (**B**) The relative orientation of the three domains varies slightly between the two crystal forms. The two structures were aligned by the P1-P2 domain alone, so as to highlight differences in the relative fine positioning of the three helical domains. Each domain is show in isolation: P1-P2 (*left*), P3-P6-P7 (*middle*) and P4-P5 (*right*).

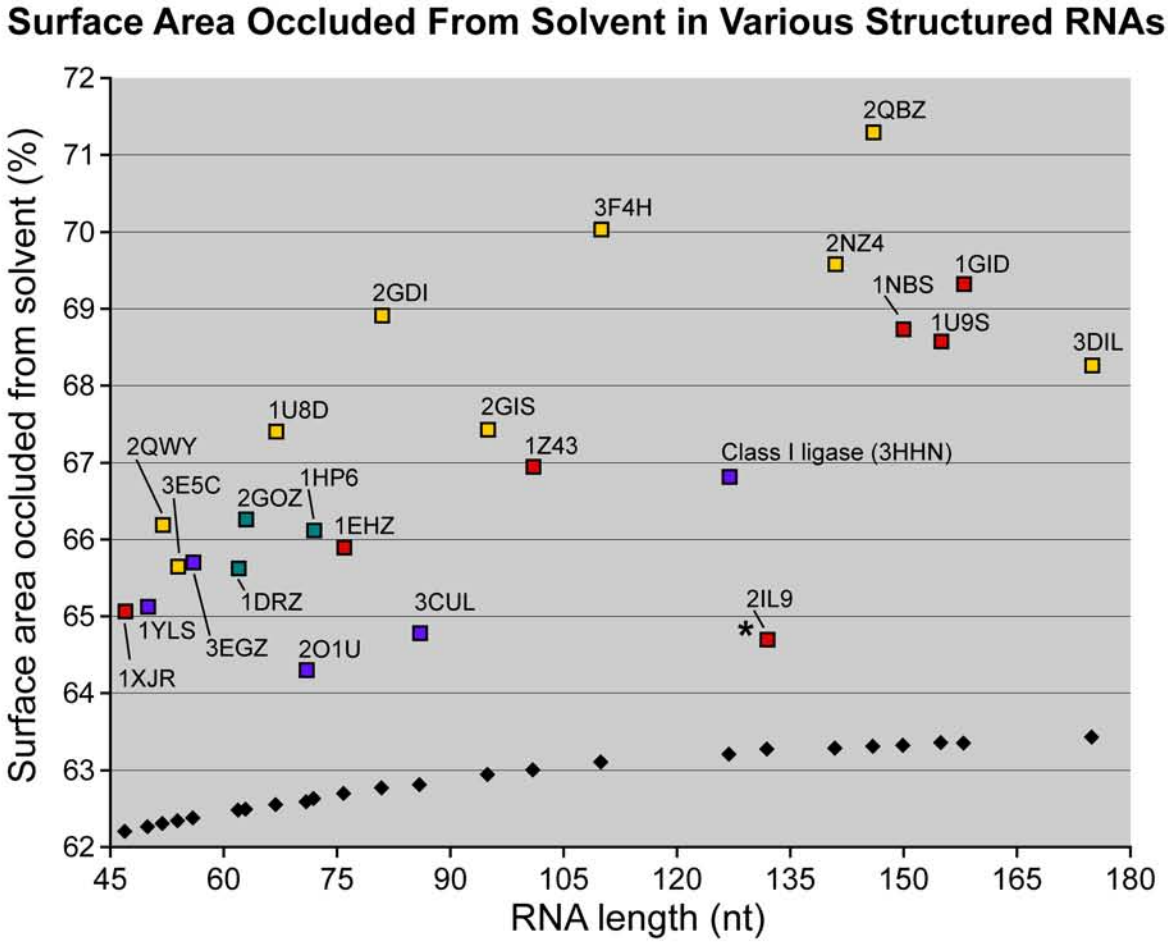

Fig. S4. Occlusion of surface area from solvent in known RNA structures. The total solvent-accessible surface area for the present structure (class I ligase, PDB ID: 3HHN) and 23 known RNA structures in the range of 61-174 nt was computed using the program AREAIMOL *(S28)*. The percent surface area buried from solvent exposure (Θ_{OCC}) was calculated by comparing this total surface area (A_{OBS}) by that of a theoretical freely-joined RNA chain of equal length (A_{FREE}) using the following equation: $\Theta_{OCC} = 100 \times (A_{FREE} - A_{OBS})/A_{FREE}$. For each structure, calculations were performed on a monomer derived from the available PDB file, without bound solvent atoms; for all RNA species using the U1A binding module to facilitate crystallization (including the class I ligase), the protein and its cognate loop were replaced *in silico* with a GNRA tetraloop. A water probe radius of 1.4 Å was used. For comparison, the same calculation was performed on perfect A-form RNAs of the same total length (black diamonds). The data have been subdivided into four classes, as follows. In teal are small, natural self-cleaving ribozymes: hepatitis delta (PDB ID: 1DRZ, *(S39)*), hammerhead (2GOZ, *(S4)*) and hairpin (1HP6, *(S29)*). In gold are the GlmS ribozyme (2NZ4, *(S24)*) and riboswitches: S-adenosylmethionine (SAM) type II (2QWY, *(S40)*), SAM type III (3E5C, *(S41)*), guanosine (1U8D, *(S42)*), thiamine pyrophosphate (2GDI, *(S43)*), SAM type I (2GIS, *(S44)*), flavin mononucleotide (3F4H, *(S45)*), metal ions (2QBZ, *(S46)*) and lysine (3DIL, *(S47)*). In red are other natural structured RNAs or domains: the SARS stem-loop II motif (1XJR, *(S48)*), yeast tRNAPhe (IEHZ, *(S49)*), the SRP S-domain (1Z43, *(S50)*), the *Plautia stali* intestinal virus IRES ribosome-binding domain (2IL9, *(S51)*), B-type (1NBS, *(S52)*) and A-type (1U9S, *(S53)*) RNase P substrate domains and the T. thermophila group I intron P4-P6 domain (1GID, *(S54)*). In dark blue are in vitro-selected RNAs: the Diels-Alderase ribozyme (1YLS, *(S55)*), the tetracycline aptamer (3EGZ, *(S56)*), the "docked" form of the L1 ligase ribozyme (2O1U, *(S15)*), the aminoacyl synthetase ribozyme ("Flexizyme"), in complex with its substrate (3CUL, *(S57)*), and the class I ligase ribozyme (3HHN). The IRES domain (marked with an asterisk) crystallized as a domain-swapped pseudo dimer with long, flexible, disordered regions between monomers; hence, a portion of surface area that may be occluded from solvent may have been artificially exposed when we generated a monomer model for solvent accessibility calculations.

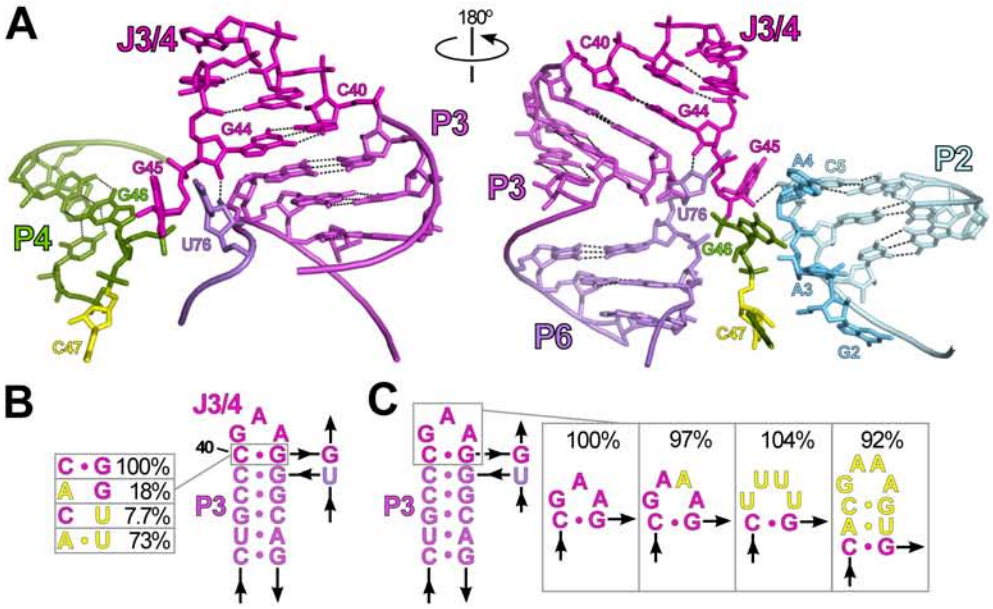

Fig. S5. Inter-domain contacts involving J3/4. (**A**) Two views of J3/4, showing the triloop enclosed by C40:G44 and highlighting interactions involving G45. For clarity, P2 has been removed from the left view and P4 has been removed from the right view. Dotted black lines denote hydrogen bonds. (**B**) Biochemical support for the role of the C40:G44 Watson-Crick pair. Relative self-ligation rates of ribozymes with the indicated substitutions (yellow) are on the left. (**C**) Expansion and replacement of the J3/4 triloop. Relative self-ligation rates of ribozymes with the indicated substitutions (yellow) are on the right.

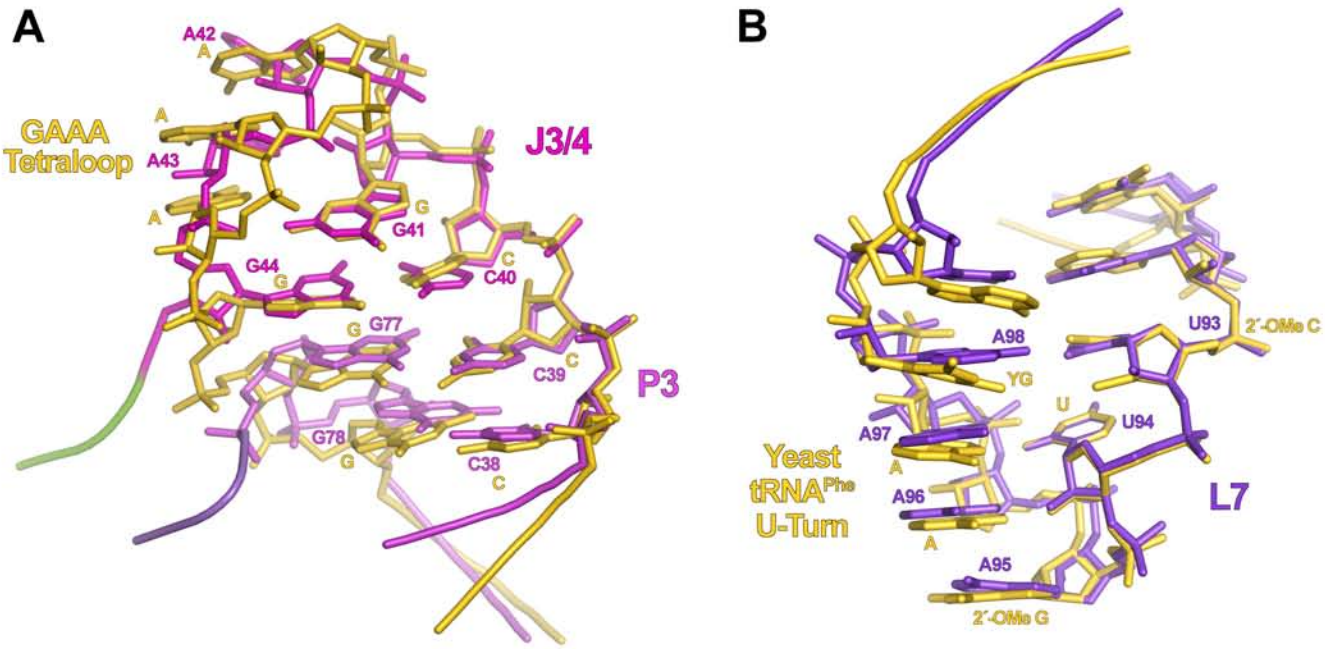

Fig. S6. The P3-P6-P7 domain is capped by two familiar RNA structural motifs. (**A**) J3/4 forms a GNRA tetraloop-like triloop. C38 and C39 were used to align the last base pairs of P3 with analogous residues in the solution structure of a GAAA tetraloop (PDB ID 1F9L, *(S58)*), shown in gold. Note that, in addition to superposing the C40-G44 base pair, G41 aligns with the analogous residue in the GNRA tetraloop. (**B**) L7 forms a classic uridine-turn (U-turn). U93 and U94 were used to align L7 with 2´-O-Methyl C32 and U33 of the anticodon loop of yeast $tRNA^{Phe}$(PDB ID 1EHZ, *(S49)*), shown in gold. For clarity, the aliphatic group of wybutosine 37 ("YG") has been removed from view.

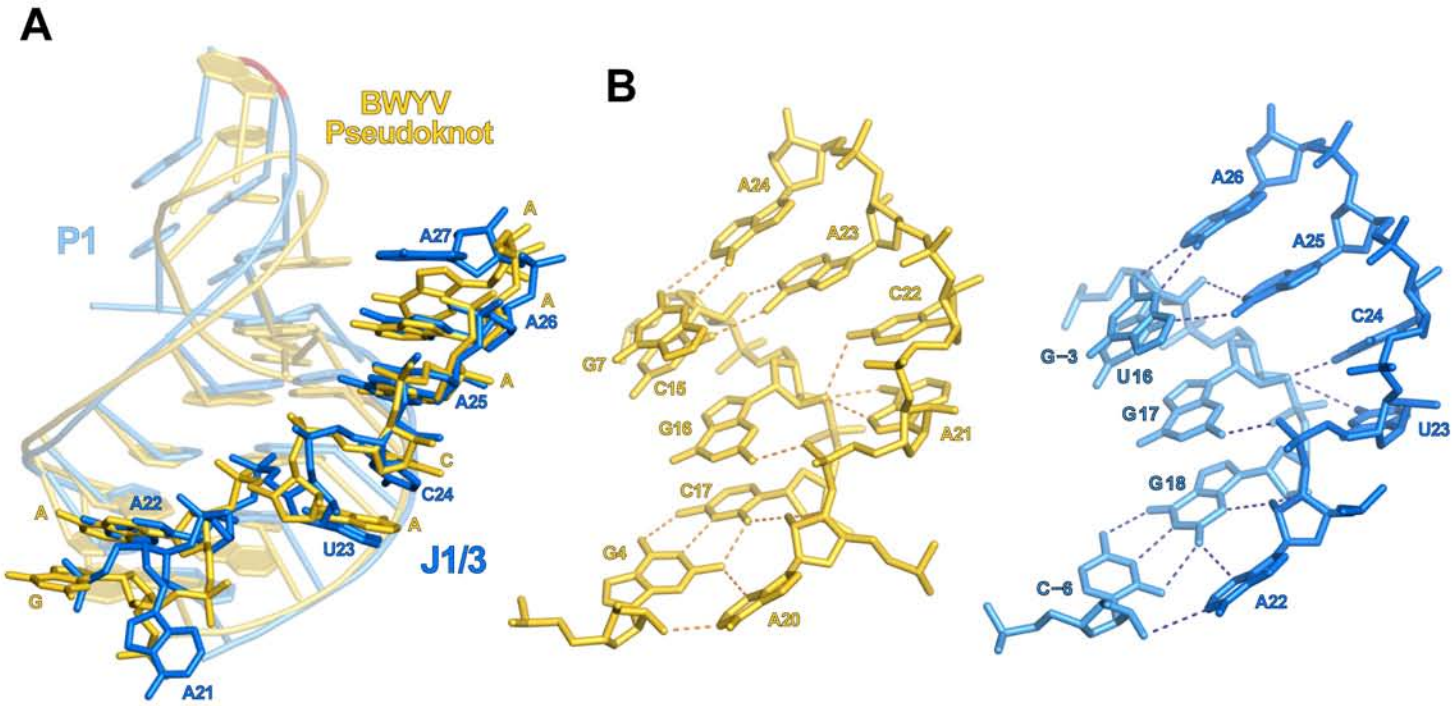

Fig. S7. The 5′ end of J1/3 mimics a viral ribosomal frame-shifting pseudoknot. (**A**) All-atom alignment between ligase residues A22–A26 in the 5′ half of J1/3 (colored as in Fig. 1C) and the beet western yellow virus (BWYV) ribosomal frame-shifting pseudoknot residues A20–A24 (gold, *(S8)*). P1 of the ligase and the remainder of the pseudoknot structure are shown as ghosted cartoons. Note the divergence of residues immediately abutting the alignment (ligase A21 and pseudoknot G19; ligase A27 and pseudoknot A25). (**B**) Side-by-side comparison between the two motifs (colored as in **A**); the BWYV pseudoknot is on the left, the class I ligase on the right. The first residue of this motif makes a type I A-minor interaction with a base-pair in the receptor helix (fig. S9). The phosphate of the following residue (ligase U23; pseudoknot A21) makes a charge-dipole interaction with the guanosine N2 in the receptor duplex. Ligase residues U23–C24 (pseudoknot residues A21–C22) make joint hydrogen bonds to the 2'-hydroxyl in the docking strand helix, for which U23 uses its exocyclic carbonyl and the pseudoknot A21 uses its N7 and N2. Ligase A25–A26 (pseudoknot residues A23–A24) form the first two base-triples of the A-minor triad motif (Fig. 2C, fig. S8 and S9). For clarity, in this second base-triple only the nucleobase for the receptor guanosine (ligase G–3; pseudoknot G4) is shown.

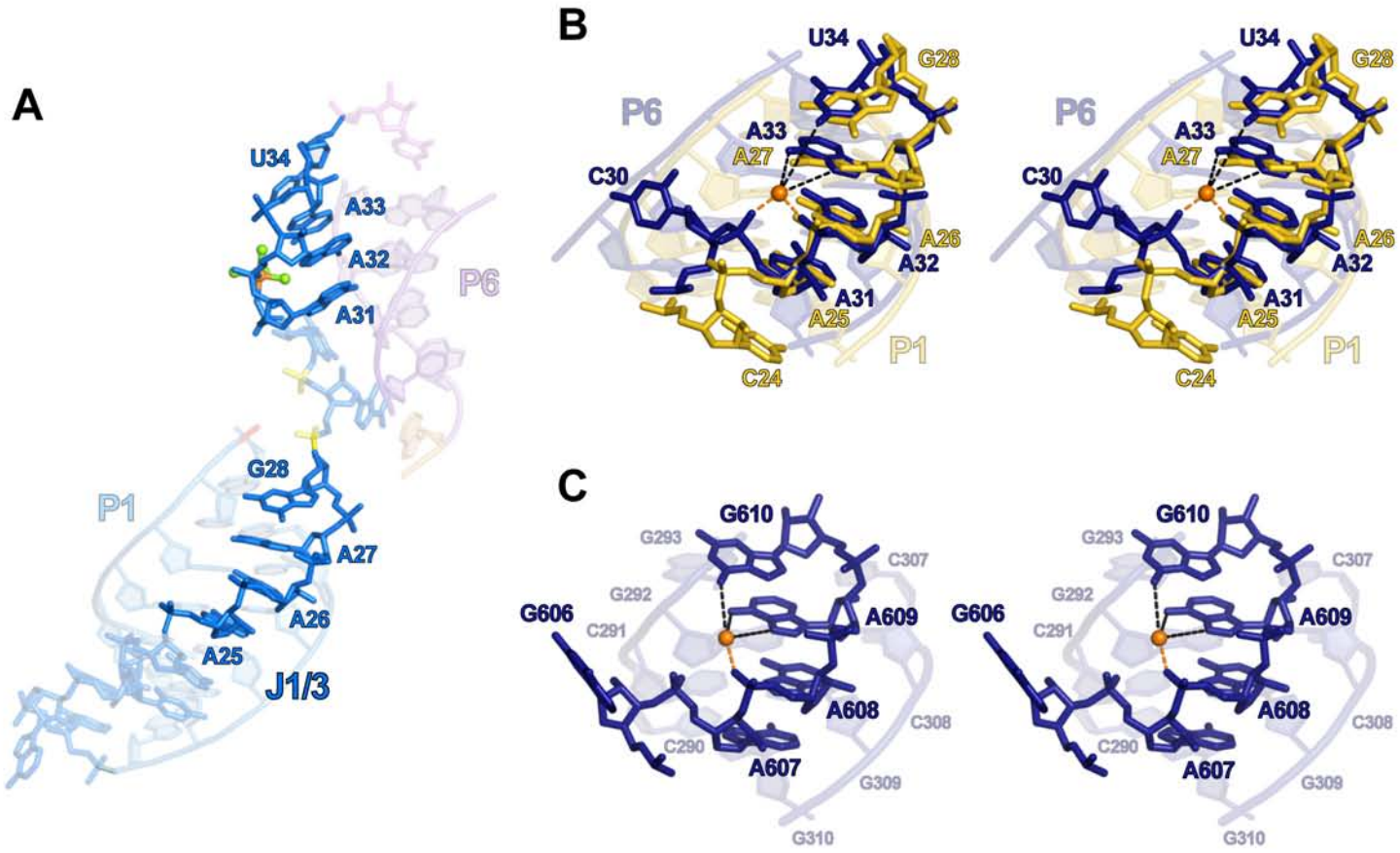

Fig. S8. The "A-minor triad" motif. (**A**) Overview of the path of J1/3, which binds in the minor grooves of P1 and P6. The view is rotated ~60° relative to that in Figure 2B. Residues comprising each of the two iterations of the A-minor triad are highlighted and numbered. (**B**) Wall-eyed stereogram of an all-atom alignment between the two motifs in J1/3. The helices bound by each are ghosted. The 5´-motif (bound to P1) is in gold; the 3´-motif (bound to P6) is in navy. The inner-sphere-coordinated magnesium ion bound by the 3´-motif is in orange. Orange dashes denote inner-sphere contacts to this metal; black dashes denote the closest approach between moieties that are proposed to make outer-sphere metal contacts and the metal center (Fig. 2B, 3A). For clarity, waters have been removed. Note both the divergence in the positions of the preceding cytidine residue (C24 and C30), and the proximity of the G28 and U34 major-groove carbonyls to one another. (**C**) An A-minor triad in the T. thermophilus 16S ribosomal RNA *(S10)*. A solvent atom (of unassigned chemistry in the PDB) is shown as an orange sphere; dashes to it are drawn in analogy to those in **B**. Note that the first base pair into which this motif docks is a C:G pair (C291:G309), whereas both those observed in the ligase are A:U pairs (A−4:U16 and A73:U84), demonstrating that recognition of this position is independent of the identity of the Watson-Crick base pair.

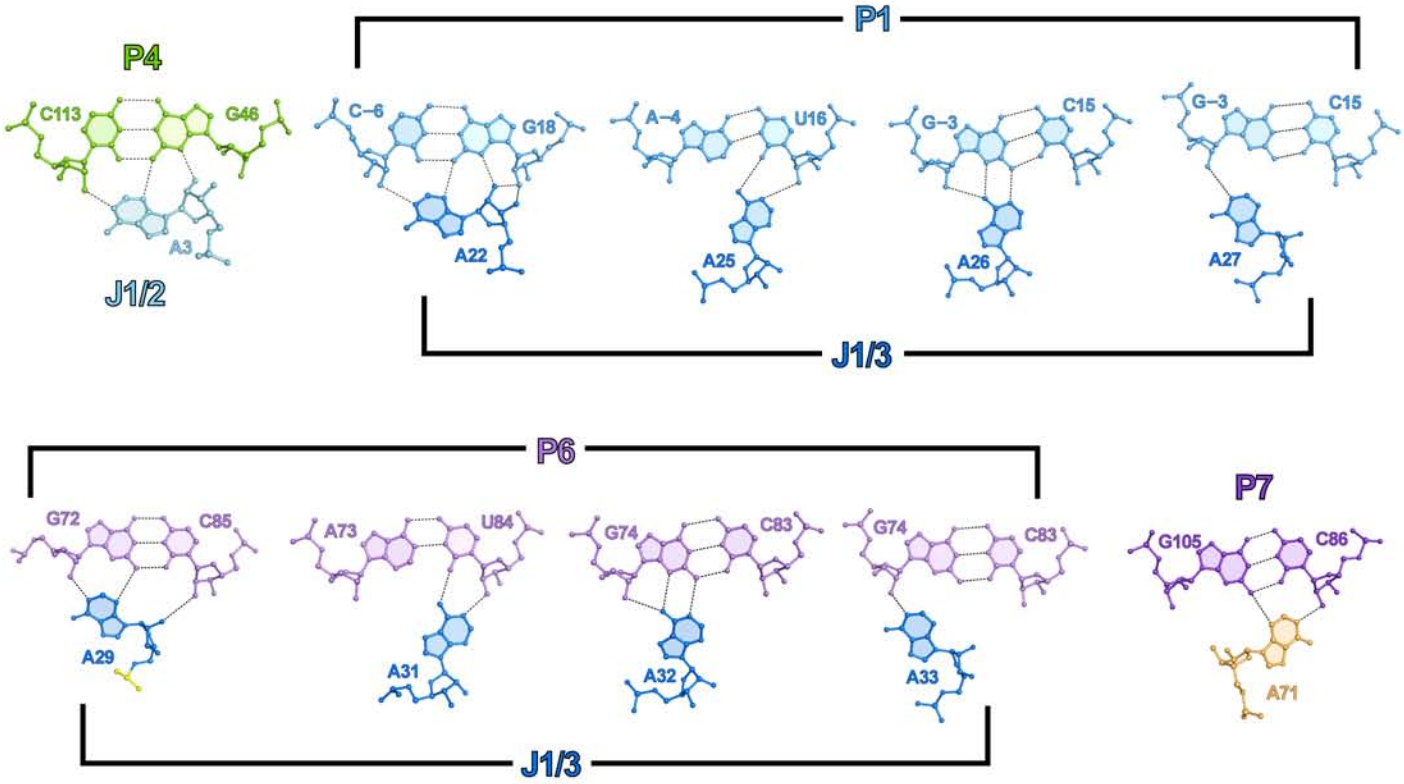

Fig. S9. Base triples observed in the class I ligase structure, each mediated through an adenosine residue. Note that the packing of residues A31 through A33 into adjacent A:U and G:C base pairs is accomplished using the same series of interactions as those employed by A25 through A27 packing into the same sequence. The remaining triples are type I A-minor interactions, although the angle at which A71 packs into C86:G105 precludes it from forming a canonical hydrogen bond with its 2´-hydroxyl *(S59)*.

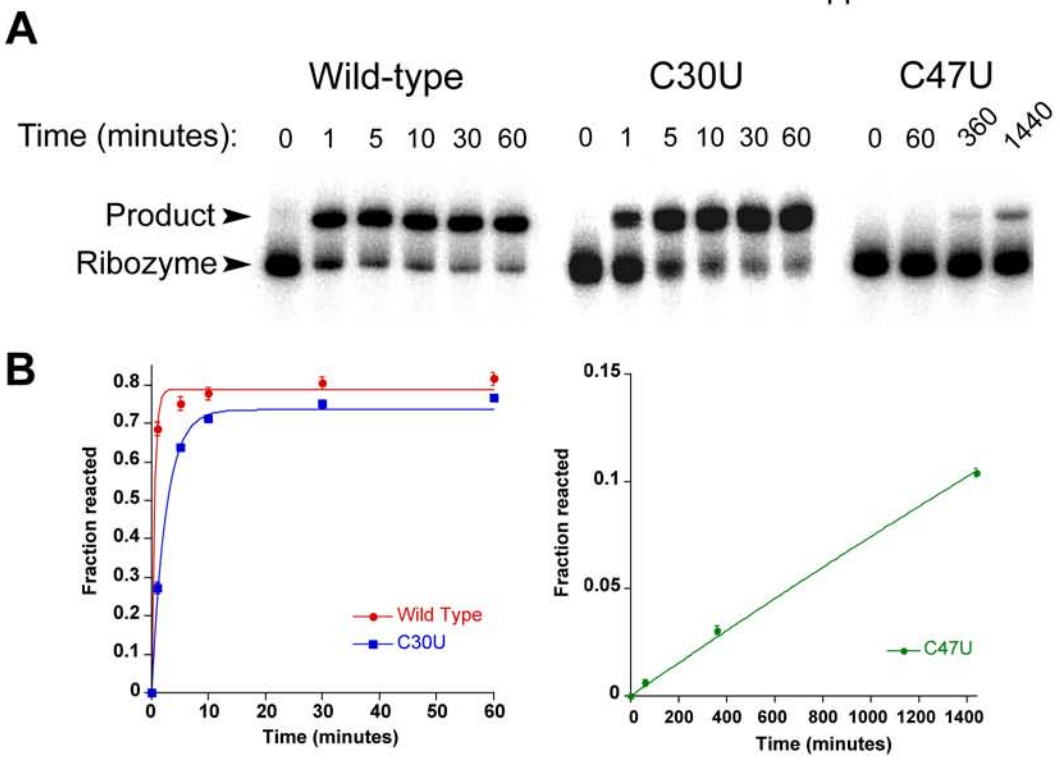

Fig. S10. Kinetics of C47U and C30U ligase active-site mutants. (**A**) Typical self-ligation time courses of ^{32}P body-labeled wild-type, C30U and C47U mutant ligase ribozymes. (**B**) Plots for reaction time courses. Error bars indicate the standard deviation for three replicate experiments.

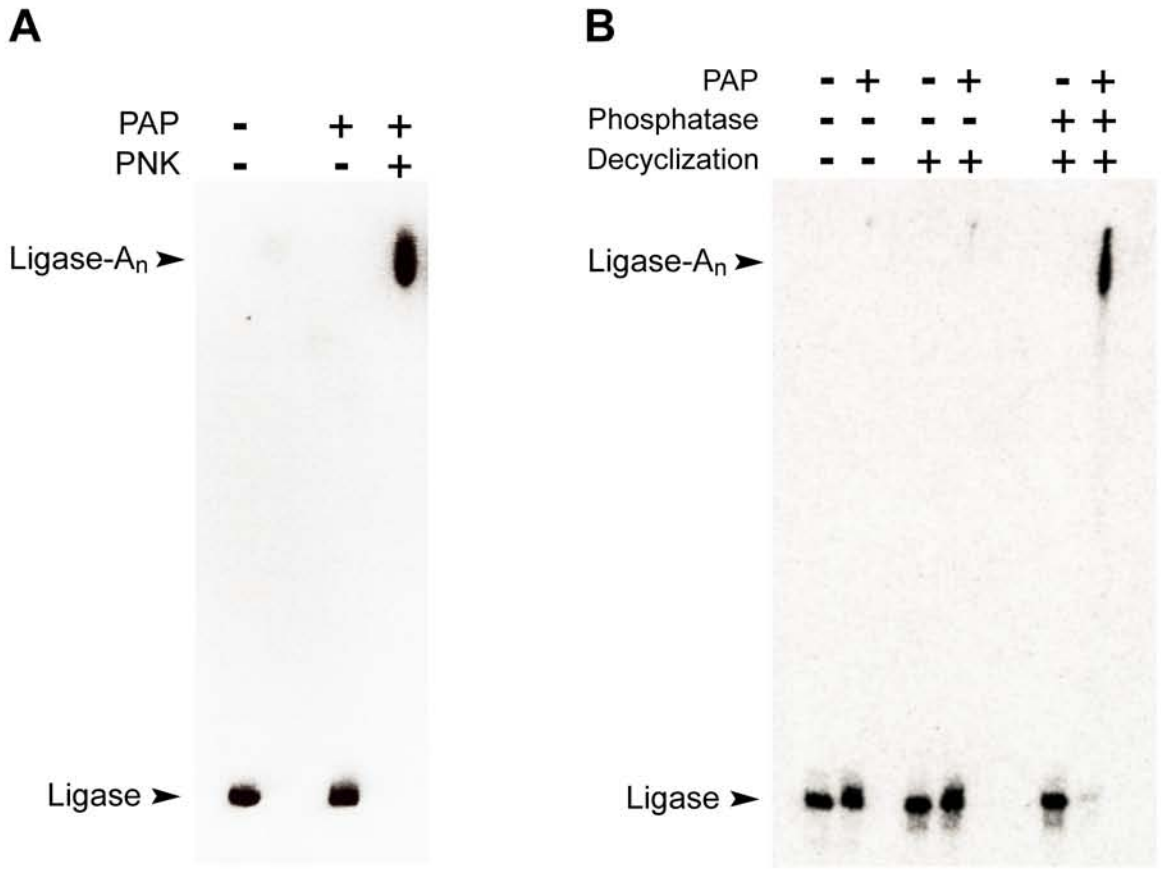

Fig. S11. The covalent state of the ribozyme 3´ terminus, probed using poly(A) polymerase (PAP). (**A**) Removal of the terminal 2´-3´ cyclic phosphate by T4 polynucleotide kinase (PNK *(S21)*) generates a substrate suitable for extension by PAP. Shown is a denaturing polyacrylamide gel stained with ethidium bromide. Ligase–A_n denotes the 3´-polyadenylated ligase. (**B**) PAP extension requires full removal of the terminal phosphate. Neither untreated RNA bearing a 2´-3´cyclic phosphate (left lanes), nor acid-decyclized RNA bearing a mixture of 2´- and 3´-linear monophosphates (center lanes, *(S22)*) were extended by PAP, whereas subsequent removal of the terminal phosphates by alkaline phosphatase treatment (right lanes) generated a substrate for PAP extension. RNA tested here was ^{32}P body labeled but was otherwise identical to that used in **A**.

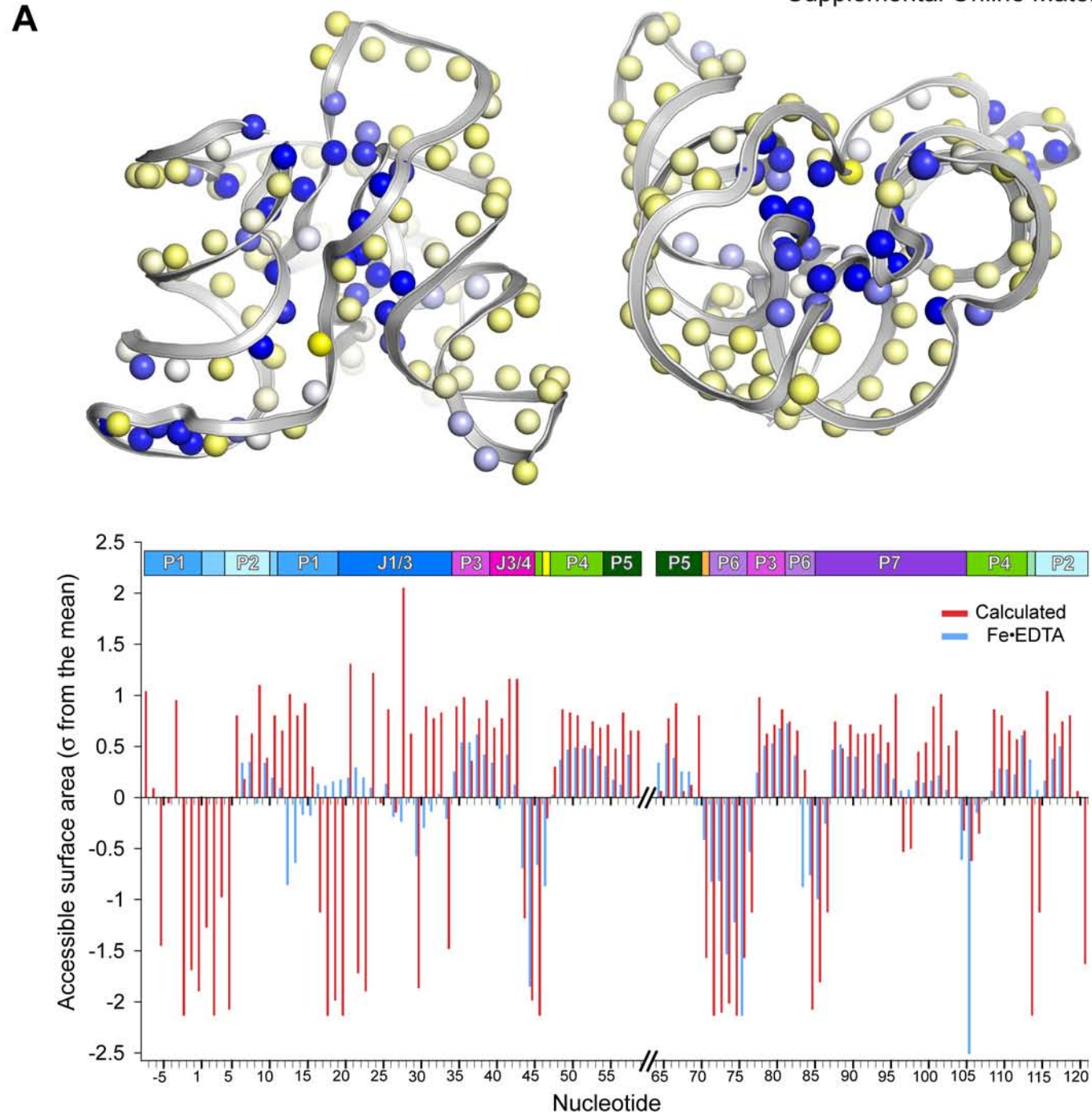

Fig. S12. Correlation between calculated and observed solvent accessibilities. (**A**) The calculated solvent-accessible surface area for C4´ carbons (spheres), generated with the program AREAIMOL *(S28)* using a water probe of 1.4Å. These data are projected onto two views of the ligase corresponding approximately to those in Fig. 1C and Fig. 1D. Positions protected from solvent are in blue; those accessible to solvent are in yellow. (**B**) Plot of the AREAIMOL calculated output (red) compared with the observed solvent protection factors derived from Fe•EDTA hydroxyl radical probing (blue) *(S33)*. Both data sets have been plotted as deviations from the mean, in standard deviation units: (value - mean)/σ, as in *(S52)*. Fe•EDTA protection factors were multiplied by -1, as positive protection factors imply occlusion from solvent. For the AREAIMOL data, mean = 7.19 Å^2, σ = 3.4Å^2 ; for the Fe•EDTA probe data, mean = 1.13, σ = 0.58. Segments spanning residues -7 to 6 and 118 to 121 were not probed by Fe•EDTA. The residues in J1/3 (which interact with the P1 helix in the crystal structure) differ in the construct subjected to Fe•EDTA probing. A lack of correlation in this region may be attributed to this change in sequence, less precise modeling (as suggested by the high temperature factors, fig. S2), or both.