The Structural Basis of Ribozyme-Catalyzed RNA Assembly

Michael P. Robertson and William G. Scott*
The Center for the Molecular Biology of RNA and
Department of Chemistry and Biochemistry
Robert L. Sinsheimer Laboratories
University of California, Santa Cruz

Supplementary materials

Additional materials, including interactive displays and electron density maps, may be found at the following website:

http://xanana.ucsc.edu/L1/

1

RNA Preparation:

The transcription template for the L1X6c construct was created by PCR amplification of a previous crystallization construct, L1X2, with primers that left the internal region of the ribozyme intact but which altered the 5' and 3' ends by introducing a GAAA tetraloop on stem A to link the 3' end of the ribozyme to its ligation substrate. Flanking the template on the upstream side was the T7 RNA polymerase promoter sequence and an EcoRI restriction site. On the downstream side of the template a BsmAI restriction site was inserted to facilitate the proper run-off transcription product. Further downstream, a BamHI restriction site was introduced which, together with the upstream EcoRI site was used to clone the construct into a pUC19 vector. The L1X6c-pUC19 expression construct was transformed into *E. coli* DH5α for amplification and the plasmid was harvested from a saturated culture using standard alkaline lysis techniques. Two milligrams of the purified plasmid was digested for 6 hours at 55°C with 50 units of BsmAI (New England Biolabs) in a 2 mL reaction. The cut plasmid was extracted with 1:1 phenol/chloroform and chloroform, ethanol precipitated, and 1 milligram was used in a 5 mL transcription with T7 RNA polymerase. Approximately 50% of the transcribed RNA autoligated in the transcription reaction and was isolated by virtue of the mobility shift of the circular product on denaturing polyacrylamide gel electrophoresis. The remaining unligated linear material was also gel isolated but proved incapable of auto-ligation, presumably a consequence of having incorrect 3' ends due to untemplated (n+1) addition products by T7 RNA polymerase (*1*). Following gel isolation and elution, the circular L1X6c RNA was desalted and concentrated to approximately 10 mg/mL using Centricon YM-3 centrifugal filter devices (Millipore).

Crystallization and Structural Solution:

RNA was heated 3 minutes at 70°C to denature and then cooled to room temperature at which time buffer and salts were added to a final concentration of 100mM NaCl, 60mM MgCl$_2$, and 30mM Tris pH 7.6 with the RNA concentration at 5 mg/mL. 2 microliters of the RNA mixture was combined in a hanging drop vapor diffusion apparatus with 2

microliters of reservoir solution (1.8M Li2SO4, 10mM MgSO4, 50mM sodium cacodylate pH 6.0 (Natrix#12, Hampton Research, (*2*))) over a reservoir of 700 microliters and incubated at room temperature. Crystals grew as plate-like clusters initially over a period of weeks to months and subsequently in less than 24 hours by micro-seeding into an equilibrated drop using a pug whisker. Crystals were cryoprotected by transferring into 2M sodium malonate (pH 7.0) and incubating for up to 1 hour at room temperature, then flash frozen in liquid N_2. Crystals diffracted to 2.6 Å in 12.6 keV syncrotron radiation (SSRL BL9-1) with space group $P2_1$ and cell dimensions a=45, b=100, c=72; $\alpha=\gamma=90$, $\beta=104$. Although isomorphous replacement and MAD phasing experiments were attempted, the best preliminary phases were obtained with the molecular replacement program PHASER(*3*) in CCP4(*4, 5*) run in automatic mode.

Crystal Structure Determination:

The structure of the L1 ligase ribozyme is a novel 71 nucleotide RNA fold having significant regions of non-A-form helical structure. Two crystallographically independent molecules possessing two distinct tertiary structural arrangements comprise the asymmetric unit, so that there is not a simple local 2-fold symmetry operator that can generate one molecule given the coordinates of the other. Despite the lack of noncrystallographic symmetry, we were able to determine the structure of the entire 142 nucleotide RNA complex without heavy-atom phasing, using instead only model A-form helical fragments for molecular replacement.

Outline of the procedure used to solve the structure:

1. Using the predicted secondary structure (based on mfold and phylogenetic comparisons) shown in Figure 1c, and the coordinates for a GAAA tetraloop found in the mimimal hammerhead ribozyme structure(*6*), we created ideal A-form helices and helical stem-loops within the program COOT(*7*) (Crystallographic Object-Oriented Toolkit). These consisted of an eight base-pair helix capped with a GAAA tetraloop (20

nucleotides in all), two three base-pair helices, each capped with a GAAA tetraloop (two stem-loops of 10 nucleotides each), and a helix having six base-pairs (a helix with 12 nucleotides in all).

2. Using phenix.xtriage(*8*), we determined the crystal space group and unit cell to be consistent with having one, two or three 71-nucleotide ribozymes in the asymmetric unit. Accordingly, we ran three (automated) independent molecular replacement trials using PHASER(*3, 9*). The "probe" consisted of one, two or three sets of the four helical model fragments generated in step 1 (i.e., the probe models contained 52, 104 and 156 nucleotides, respectively). The most promising molecular replacement solution was obtained with the 52 nucleotide model consisting of four fragments, despite the fact that this represented only 36% of the contents of the asymmetric unit. The 104 and 156 nucleotide probes yielded solutions that possessed major steric clashes between the various helical fragments and appeared to have associated electron density maps that were degraded compared to the 52 nucleotide solution, and these other solutions were not pursued.

3. Using the 52 nucleotide solution obtained with the four helical fragments described in step 1 as an initial partial solution, we were able to iteratively improve the electron density map by adding in single copies of the six base-pair model A-form helix. It quickly became apparent that although a sensible solution had not been obtained with respect to the detailed placement of the known sequence, the solution with respect to the generic helical structure of the molecule was sound. Each iterative improvement was followed by manual model adjustment and REFMAC(*10*) positional refinement within COOT, and individual nucleotides involved in steric clashes or that occupied ill-defined electron density were deleted from the model prior to the subsequent rounds of molecular replacement in PHASER. In this way, about 85% of the model was constructed, at which point unmodeled but easily recognizable structural features, such as missing GNRA tetraloops, appeared in 2Fo-Fc electron density maps and were modeled into the density manually.

4. After about 90% of the nucleotides had been provisionally placed in well-defined electron density, the entire model was converted to poly-C within coot, refined again in REFMAC, and then used to generate an "experimental" phase set.

5. The initial poly-C model was then discarded, and the corresponding "experimental" phase set was then subjected to solvent-flattening (using "solvent flipping") with CNS(*11*) in an attempt to minimize the prior model bias. The electron density map generated by solvent-flattening in CNS was then imported into COOT, and skeletonized. The 2.6 Å solvent-flattened map was of significantly higher quality than that generated by the initial poly-C model, and was of sufficiently high quality that it allowed unambiguous tracing of the phosphodiester backbones of the two monomers. In addition, the density for the nucleotide bases enabled purines to be distinguished from pyrimidines, and in many cases adenines from guanines, without reference to the known sequence. This therefore permitted an unambiguous assignment of the sequence to the electron density, and initial models for each of the two crystallographically independent 71-nucleotide ribozymes were constructed by each of the two coauthors independently and were found to be in agreement.

6. Several rounds of model building and adjustment with the actual sequence, followed by calculation of additional composite-omit maps, permitted identification of every nucleotide in molecule Q and all but two residues in molecule P. Refinement was performed in CNS(*11*) and Refmac(*10*) as well as within COOT(*7*), and figures were prepared using PyMol(*12*). The positions of 20 of the 22 uridines were well-ordered, and these were used to estimate the positions of 20 bromine atoms in a MAD dataset that was unsuccessfully used for MAD phasing (see below).

Confidence checks:

Two additional calculations were performed to check the veracity of the structural model and its sequence.

1. Composite-omit map

Using CNS, a composite-omit 2Fo-Fc map was generated in which a different 5% of the model was omitted for each of 20 simulated annealing refinements in which the initial starting temperature was set to 4,500K in an attempt to minimize model bias. Four complementary stereo views of the resulting composite-omit map are shown in Figure 1S of the supplementary materials, and a PyMOL(*12*) saved session that permits the reader to download and examine the composite omit map, as well as the final 2Fo-Fc refined structural map, is available from the author's web site: http://xanana.ucsc.edu/L1

2. 5-BrU difference Fourier

We also obtained a Br MAD dataset, but our attempts to find the Br sites in order to solve the structure using experimental MAD (or SIR) phases were not successful. The derivative RNA was prepared using 5-BrUTP instead of UTP in the transcription mixture, potentially incorporating 11 Br atoms per ribozyme molecule (and thus 22 Br sites per crystallographic asymmetric unit). Either due to photolysis or partial incorporation, the Br data proved to be quite weak and unusable for phasing. However, when the coordinates of the native 2.6 Å structure were re-refined using the absorption peak data set to 3 Å resolution, a sigmaA-weighted difference Fourier using the Br derivative data and the re-refined native structure revealed clear difference density for most of the Br sites immediately adjacent to the C5 positions on the various uridines. (The exceptions corresponded to the disordered residues discussed in the main text). Most of these peaks were above 5σ. An example is shown in Figure 2S, and the data processing statistics for the Br peak dataset are reported in Table 1S.

Table 1S: Data processing statistics for 5-BrU-L1 Ligase RNA

```
                                          Overall  OuterShell

Low  resolution   limit                    100.00        3.16
High  resolution   limit                     3.00        3.00

Rmerge                                       0.087       0.384
Rmeas  (within  I+/I-)                       0.120       0.527
Rmeas  (all  I+ & I-)                        0.137       0.552
Rpim  (within  I+/I-)                        0.083       0.360
Rpim  (all  I+ & I-)                         0.071       0.285
Fractional   partial   bias                 -0.018      -0.185
Total  number   of  observations            46020        6799
Total  number   unique                      12694        1846
Mean((I)/sd(I))                                8.8         3.0
Completeness                                  99.6        99.6
Multiplicity                                   3.6         3.7

Anomalous   completeness                      98.7        99.0
Anomalous   multiplicity                       1.8         1.8
DelAnom  correlation   between   half-sets   0.435      -0.015
Mid-Slope   of  Anom  Normal  Probability    1.193
```

Figure Captions

Figure 1S: Wall-eyed stereo views of a simulated annealing composite-omit 2Fo-Fc map prepared in CNS, in which 20 simulated annealing refinements, each omitting a different 5% of the structure, were calculated starting with an annealing temperature of 4,500K to minimize model bias. The figure was prepared in PyMOL(*12*).

Figure 2S: Wall-eyed stereo views of a sigma-A-weighted Fo-Fc map at 3 Å resolution (white mesh), in which Fo is the 5BrU-modified RNA, and Fc and the phases are calculated using the model obtained from the native 2.6 Å resolution dataset and subsequently refined with the 5BrU data. The RNA is shown in cyan, except for the Uridine residues, which are highlighted in magenta. **Figure A** gives a global view of the two ribozymes in the crystallographic asymmetric unit, in which the map is contoured at 3.5 rmsd, and **Figure B** shows a close-up of the region including the ligation site of Molecule Q, in which the map is contoured at 4 r.m.s.d. The figure was prepared in PyMOL(*12*).

Figure 3S: Determination of the ligation rate of an intermolecular ligation ribozyme analogue to Lx6c. **Figure A** shows the sequence used for the ligation assay. **Figure B** is a graph of the natural log of the amount of unligated substrate remaining vs. time. The ligation rate is the absolute value of the slope of the plot, which is linear, assuming pseudo-first-order kinetics, and is 0.0301 turnovers per minute or 1.8 turnovers per hour.

References:

1. J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res* **15**, 8783 (Nov 11, 1987).
2. W. G. Scott *et al.*, *J Mol Biol* **250**, 327 (Jul 14, 1995).
3. A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, *Acta Crystallogr D Biol Crystallogr* **61**, 458 (Apr, 2005).
4. M. D. Winn, *J Synchrotron Radiat* **10**, 23 (Jan 1, 2003).
5. N. Collaborative Computational Project, *Acta Crystallographica* **D50**, 76 (1994).
6. W. G. Scott, J. T. Finch, A. Klug, *Cell* **81**, 991 (Jun 30, 1995).
7. P. Emsley, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **60**, 2126 (Dec, 2004).
8. P. D. Adams *et al.*, *J Synchrotron Radiat* **11**, 53 (Jan 1, 2004).
9. R. J. Read, *Acta Crystallogr D Biol Crystallogr* **57**, 1373 (Oct, 2001).
10. A. A. V. a. E. J. D. G. N. Murshudov, *Acta Crystallographica* **D53**, 240 (1997).
11. A. T. Brunger *et al.*, *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**, 905 (Sep 1, 1998).
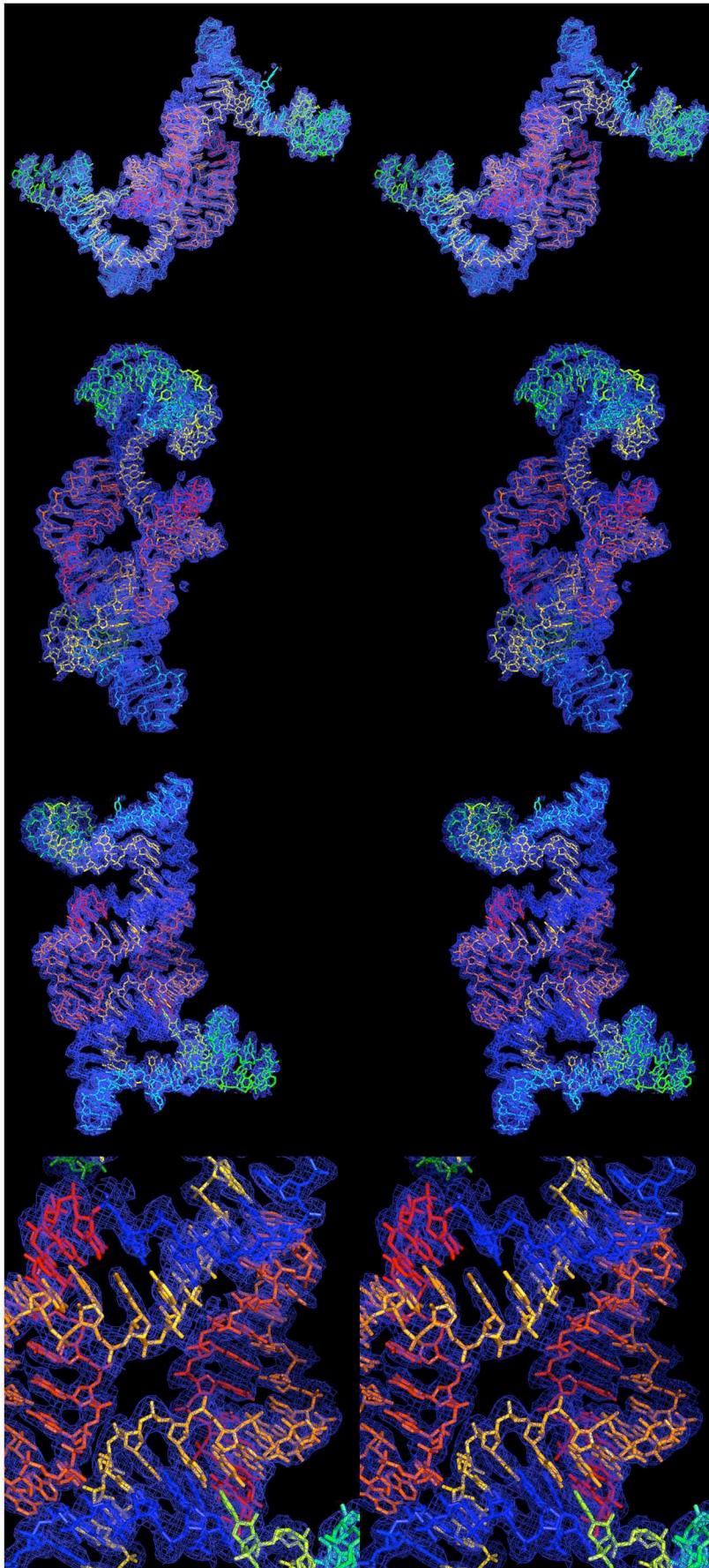12. W. L. DeLano, (2003). http://pymol.sourceforge.net
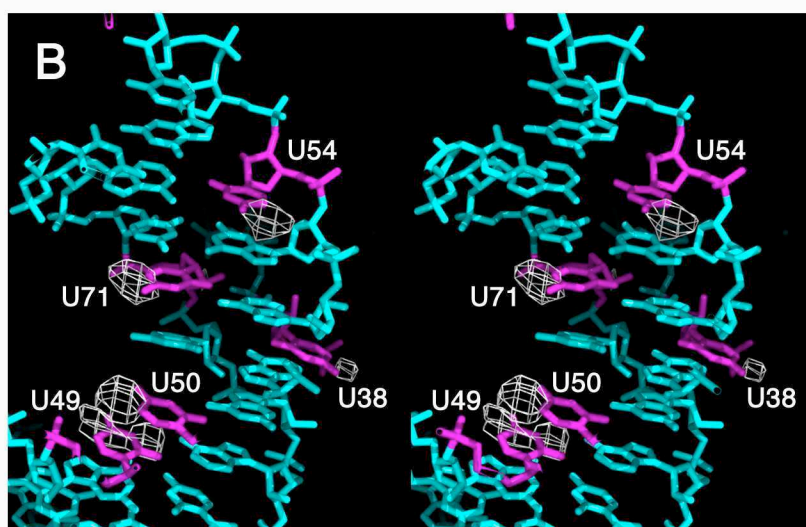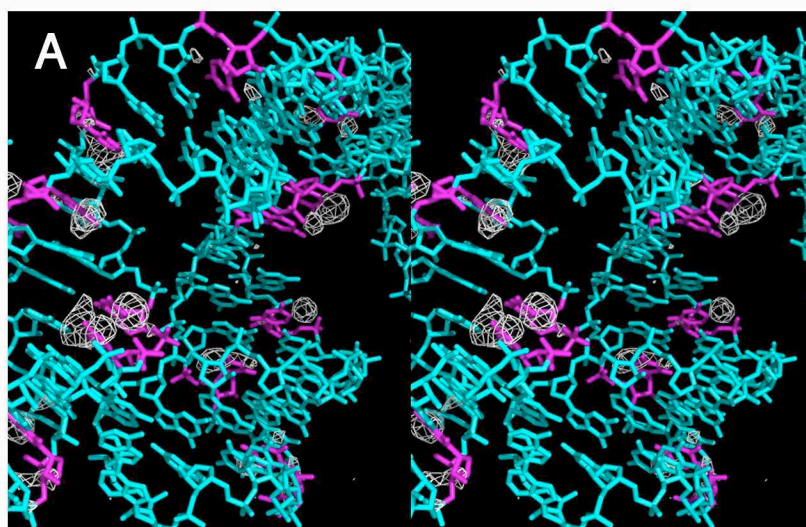
Figure 1S

Figure 2S

A

5' **g c g a c t g g a c a t c a c g a g** 3'   (dA)$_{22}$ 5'

3' c g c u g a c c u g u a g u g c u c   U

                70   60

                                    C — G
                                    G — C
                                    U — A
                                    G — C
Stem A   G ● U 3'
                                    A ◆ g −1   ← ligation site
                              50— U ● g
                                    U — a $\ln$ (substrate)
                                    G — c
                                    G — c
                                    A — u
                                    G — c
                                    G
                              40
    Stem C   C   U
          A A G C U   U A G A   C A
        A | | | ◆   ◆ | | |
        A G C G A   A G C U   G C — g
      30   C C   U   C — g
                        20
                                    G — c −10
                              a   g
                                a a

                    Stem B
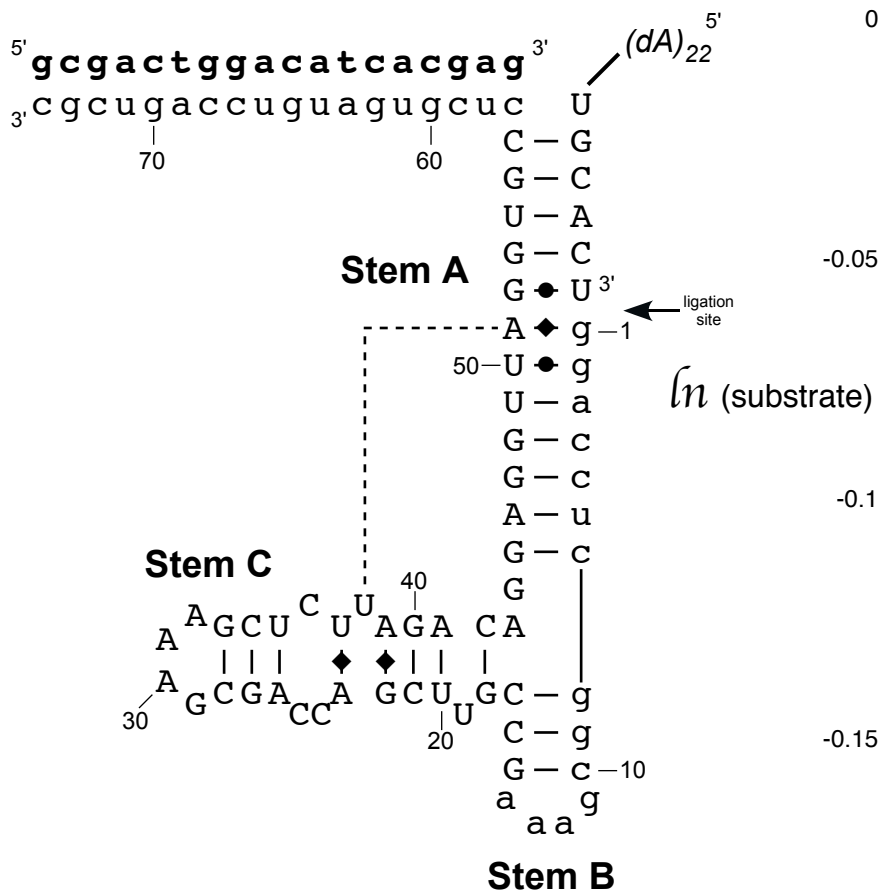
B

ribozyme ligation rate

—— y = -0.030126x   R= 0.99991

ligation rate = 1.8/hr

minutes

Figure 3S