SUPPLEMENTARY INFORMATION

Native architecture encodes cooperativity and specificity in RNA folding intermediates

Reza Behrouzi,[1] Joon Ho Roh,[2, 3] Duncan Kilburn,[1] Robert M. Briber,[2] Sarah A. Woodson[1]*
[1]T. C. Jenkins Department of Biophysics, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218 USA, [2]Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742 USA, [3]Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, MD, 20899

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Assays of ribozyme folding. (A) SAXS measurements of ribozyme collapse. Left, representative scattering intensity, I(Q), as a function of momentum transfer, Q, after background subtraction. These data are for 0.4 mg/mL (6.2 μM) L9P6 ribozyme in 20 mM Tris-HCl plus 0-3.25 mM $MgCl_2$ at 37 ˚C. Right, pair distribution functions, P(r) obtained from inversion of the data (see Experimental Procedures). (B-E) Ribozyme activity assay. ^{32}P-labeled RNA substrate (5' rCAUAUCGCC) was cleaved at 37 ˚C in 20 mM Tris buffer (pH 7.5) containing 0.5 mM GTP, 0-15 mM $MgCl_2$, and 6.2 μM WT L-3 *Azoarcus* ribozyme. Products were analyzed by denaturing 20% PAGE. (B) Burst phase of substrate cleavage was determined by monitoring the reaction progress in 15 mM Mg^{2+}. (C) Fraction of product versus time was fit to single or double exponential rate equations to determine the length of the burst phase representing the first turnover. All time-progress curves are normalized to WT maximum cleavage. (D) Substrate cleavage after 20 s at 37 ˚C in 20 mM Tris buffer (pH 7.5) and 0-15 mM $MgCl_2$. (E) Fraction of native ribozyme determined from the burst size in (D), fit to cooperative two state model (Hill equation). The ribozyme was folded 15 min at 50 ˚C (broken lines, open circles) or > 30 min at 37 ˚C (solid line, filled circles) before the addition of substrates at 37 ˚C.

Figure S2. Hydroxyl radical footprinting of native state. Comparison of backbone solvent accessibility hydroxyl radical (Fe-II)-EDTA) cleavage of wild type (black) and mutant ribozymes (red). RNAs were folded in CE+15 mM Mg^{2+} (10 mM sodium cacodylate, 0.1 mM EDTA, 15 mM $MgCl_2$. RNAs were 5' end labeled with ^{32}P and analyzed on 8% sequencing gels. (A) J2/3, A39U; (B) P6, A97U; (C)TH, G125A; (D) L9, A190U. Mutations sites are marked by * next to x axis. Blue shaded regions indicate residues in which the ribose C4' is buried in the crystal structure (Adams et al., 2004a). Residues 72-83 and 100-118 were not quantified due to gel compressions. Backbone protection levels and error bars are the average and standard deviation of 2-4 independent experiments. Band intensities were normalized to the total lane intensity and adjusted for the number of quantified bands in each sample. Relative cleavage values below 1 indicate protected regions.

Figure S3. Mg^{2+} -dependence of folding. (A) U to I_C transition from SAXS. Values of Rg^2 were obtained from *P(r)* distributions based on SAXS data (see Fig. S1), and were fit to a 3-state model (eq. 2) that included an extended (non-native) intermediate I_U. A small 2 Å decrease in R_g above 2 mM Mg^{2+}, which correlates with the transition from I_C to N (Moghaddam et al., 2009), was excluded to improve the accuracy of the fit to the two main collapse transitions between U and I_U and I_U and I_C. Fit parameters are listed in Table S2. (B) I to N transition from ribozyme activity. The fraction native RNA determined from the first turnover (see Fig. 1) was normalized to the maximum cleavage in high Mg^{2+} and fit with a two state Hill model (eq. 1). Data from two or more independent repeats were pooled and fit globally. Fit parameters are listed in Table S2.

Figure S4. Contributions of tertiary interactions to folding stability. Stacked plots showing the folding free energies of the ribozyme variants in different Mg^{2+} concentrations. Top, native state (N), low Mg^{2+} (white) to high Mg^{2+} (blue); bottom, compact intermediates (I_C), low Mg^{2+} (white) to high Mg^{2+} (red). Mg^{2+} concentrations corresponding to 90% population of N (5 mM) and I_C (0.8 mM) are highlighted in green on the top and bottom surfaces, respectively.

Figure S5. Thermodynamic cycles used to calculate ΔG_{link}. (A) Thermodynamic coupling (linkage) between pairs of tertiary interactions (*i* and *j*) was determined from perturbations to folding free energy of the ribozyme using linkage analysis and double mutant cycles (Horovitz and Fersht, 1992; Weber, 1975). Vertical arrows indicate folding steps, for which we measured the free energies experimentally, whereas horizontal arrows indicate the hypothetical process of introducing *i* and *j* mutations. We calculated the linkage energy between *i* and *j* as the difference in the free energy of making mutation *j*, when mutation *i* is present or absent. Colored squares show the transitions that are used in deriving the equations for calculating the linkage free energies (see Extended Experimental Procedures). (B) Mg^{2+}-dependence of cooperativity between tertiary interactions. Stacked plots showing $\Delta\Delta G_{link}$ in different Mg^{2+} concentrations. Top, I_C to N transition, low Mg^{2+} (white) to high Mg^{2+} (blue); bottom, U to I_C transition, low Mg^{2+} (white) to high Mg^{2+} (red).

SUPPLEMENTARY TABLES

Table S1. Description of mutations used in this study

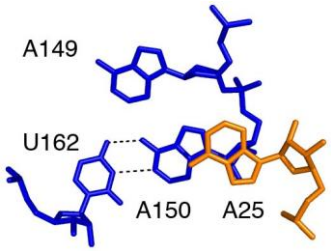
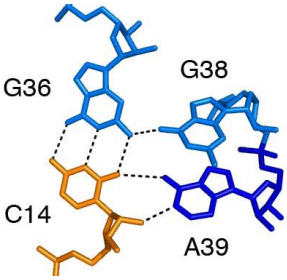
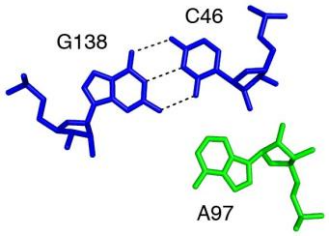
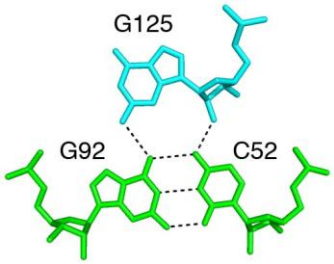
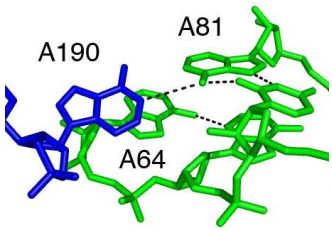
Mutation		Interaction	
L2	A25U	(L2) GAAA – (P8-8a) canonical 11 nucleotide receptor. Removal of this interaction has only a small effect on 3' splice site cleavage (Ikawa et al., 2000). Diagrams of molecular interactions made with PyMOL from pdb *1u6b* (Adams et al., 2004b).	
J2/3	A39U	Tandem stack of minor groove interactions by J2/3 and P2 (Adams et al., 2004a). The greater catalytic turnover of this mutant may originate from altered P1 docking as observed for the *Tetrahymena* ribozyme (Young et al., 1991).	
P6	A97U	A-minor interaction with C46 in P3 (Adams et al., 2004b; Rangan et al., 2003). Perturbations measured by A97U mutation are identical within experimental uncertainty to the effect of changes in Watson-Crick base pairs in the receptor in P4-P6 RNA (Battle and Doudna, 2002) and compatible with stabilization by each hydrogen bond of A-minor motifs by ~0.5 kcalmol^{-1} (Silverman et al., 1999).	
TH	G125A	Base triple with C52-G92 base pair in P4 is part of a universally conserved triple helix in group I introns (Michel et al., 1989; Michel and Westhof, 1990; Tanner and Cech, 1997). The G-C-G triple in the *Azoarcus* ribozyme is a particularly stable configuration (Ikawa et al., 2000).	
L9	A190U	(L9) GAAA – (P5-5a) canonical 11 nt receptor. All group I introns contain an analogous contact that contributes to stability and 3' splice site cleavage (Jaeger et al., 1993); free energy 0 to 2 kcal/mol depending on temperature and Mg^{2+} concentration (Downey et al., 2006).	

Table S2. Mg^{2+}-dependent folding of wildtype and mutant *Azoarcus* ribozymes[a]

RNA	U → I_C, SAXS			I_C → N, Activity		
	C_m	n	ΔG ($kcalmol^{-1}$) (0.8 mM $MgCl_2$)	C_m	n	ΔG ($kcalmol^{-1}$) (5 mM $MgCl_2$)
WT	0.6 ± 0.02	6.7 ± 0.7	-1.3 ± 0.1	2.6 ± 0.1	3.0 ± 0.2	-1.3 ± 0.1
L2	0.6 ± 0.01	4.8 ± 0.4	-0.8 ± 0.1	2.5 ± 0.1	3.7 ± 0.3	-1.5 ± 0.1
J2/3	1.0 ± 0.02	3.7 ± 0.2	0.4 ± 0.1	3.8 ± 0.2	3.6 ± 0.6	-0.5 ± 0.2
P6	0.6 ± 0.01	5.3 ± 0.6	-0.8 ± 0.1	2.5 ± 0.1	3.1 ± 0.3	-1.3 ± 0.2
TH	1.0 ± 0.02	5.1 ± 0.5	0.7 ± 0.1	2.8 ± 0.1	5.1 ± 0.8	-1.7 ± 0.4
L9	0.7 ± 0.02	5.8 ± 0.5	-0.6 ± 0.1	2.8 ± 0.1	6.0 ± 0.6	-2.1 ± 0.2
L2P6	0.6 ± 0.02	4.6 ± 0.2	-0.7 ± 0.1	2.7 ± 0.1	4.0 ± 0.3	-1.5 ± 0.1
L2TH	0.5 ± 0.01	5.3 ± 0.4	-1.7 ± 0.1	2.8 ± 0.0	5.8 ± 0.5	-2.0 ± 0.2
L2L9	0.8 ± 0.03	4.6 ± 0.4	-0.1 ± 0.1	3.3 ± 0.1	5.9 ± 0.4	-1.5 ± 0.1
P6TH	0.4 ± 0.02	4.2 ± 0.5	-1.6 ± 0.1	3.7 ± 0.1	5.4 ± 0.5	-1.0 ± 0.1
L9P6	0.7 ± 0.03	4.1 ± 0.3	-0.3 ± 0.1	3.0 ± 0.1	3.0 ± 0.3	-0.9 ± 0.1
L9TH	0.7 ± 0.01	4.5 ± 0.4	-0.5 ± 0.1	4.6 ± 0.2	2.5 ± 0.2	-0.1 ± 0.1
L2P6TH	0.5 ± 0.02	4.0 ± 0.5	-1.3 ± 0.1	4.1 ± 0.1	4.6 ± 0.4	-0.6 ± 0.1
L2P6L9	0.4 ± 0.02	2.3 ± 0.2	-1.1 ± 0.1	3.4 ± 0.1	4.8 ± 0.4	-1.1 ± 0.1
L2THL9	0.9 ± 0.01	6.2 ± 0.5	0.6 ± 0.1	4.6 ± 0.2	3.4 ± 0.3	-0.2 ± 0.1
P6THL9	0.4 ± 0.02	2.2 ± 0.2	-1.1 ± 0.1	5.7 ± 0.2	3.8 ± 0.4	0.3 ± 0.1
L2P6THL9	0.5 ± 0.02	3.5 ± 0.3	-1.0 ± 0.1	5.7 ± 0.1	4.3 ± 0.3	0.3 ± 0.1

[a] All assays were performed at 37 ˚C using 6.2 μM ribozyme. When available, data from independent repeats were pooled before fitting. Transition midpoints (C_m) and slopes (n) for U to I_C and I_C to N were calculated from two-state transitions applied to populations of I_C and N, respectively (see Methods and Fig. S3). The reported errors were obtained by iterative resampling of residuals and refitting. In U to I_C, the reported C_m and n values correspond to the major collapse transition (Fig 3B and Fig. S3).

EXTENDED EXPERIMENTAL PROCEDURES

Thermodynamic linkage

Thermodynamic linkage (coupling) between two tertiary interactions *i* and *j* were calculated from the folding energies of single and double mutants by the use of a thermodynamic cycle (Fig. S5A). Analogous to the case of binding of multiple ligands (Weber, 1975), thermodynamic linkage ($\Delta\Delta G^{S}_{link}$) between interactions *i* and *j* in the S state (I_C or N) is defined as the difference in the free energy of mutating interaction *i* in the presence (ΔG_i) or absence (ΔG_{ij}) of interaction *j*:

$$\Delta\Delta G^{I}_{link} = \Delta G^{I}_{ij} - \Delta G^{I}_{i} = \Delta G^{I}_{ji} - \Delta G^{I}_{i} \quad (4a)$$

$$\Delta\Delta G^{N}_{link} = \Delta G^{N}_{ij} - \Delta G^{N}_{i} = \Delta G^{N}_{ji} - \Delta G^{N}_{i} \quad (4b)$$

The linkage free energies measured by double mutant cycles ($\Delta\Delta G_{exp}$, Fig S9) are the differences between linkage energies in the final and initial states of each transition:

$$\Delta\Delta G^{I}_{\exp} = (\Delta G^{I}_{ij} - \Delta G^{I}_{i}) - (\Delta G^{U}_{ij} - \Delta G^{U}_{i}) \quad (5a)$$

$$\Delta\Delta G^{N}_{\exp} = (\Delta G^{N}_{ij} - \Delta G^{N}_{i}) - (\Delta G^{I}_{ij} - \Delta G^{I}_{i}) \quad (5b)$$

Assuming that the mutated residues do not interact in the unfolded ensemble, the second term in eq. 5a is zero. Therefore,

$$\Delta\Delta G^{I}_{\exp} = (\Delta G^{I}_{ij} - \Delta G^{I}_{i}) \approx \Delta\Delta G^{I}_{link} \quad (6a)$$

$$\Delta\Delta G^{N}_{\exp} = (\Delta G^{N}_{ij} - \Delta G^{N}_{i}) - \Delta\Delta G^{I}_{link} = \Delta\Delta G^{N}_{link} - \Delta\Delta G^{I}_{link} \quad (6b)$$

These values can be calculated from the folding energies using the thermodynamic cycles shown in Fig. S5A:

$$\Delta\Delta G^{I}_{\exp} = (\Delta G^{U\to I}_{ij} + \Delta G^{U\to I}_{WT} - \Delta G^{U\to I}_{i} - \Delta G^{U\to I}_{j}) = \Delta\Delta G^{I}_{ij} - (\Delta\Delta G^{I}_{i} + \Delta\Delta G^{I}_{j}) \quad (7a)$$

$$\Delta\Delta G^{N}_{\exp} = (\Delta G^{I\to N}_{ij} + \Delta G^{I\to N}_{WT} - \Delta G^{I\to N}_{i} - \Delta G^{I\to N}_{j}) = \Delta\Delta G^{N}_{ij} - (\Delta\Delta G^{N}_{i} + \Delta\Delta G^{N}_{j}) \quad (7b)$$

in which $\Delta\Delta G^{S}_{x}$ designates the free energy perturbations caused by mutation *x* (*i*, *j*, or *ij*) in the I_C or N states at the reference Mg^{2+} concentration in each assay, and

$$\Delta\Delta G^{I}_{x} = \Delta G^{U\to I}_{x} - \Delta G^{U\to I}_{WT} \quad (8a)$$

$$\Delta\Delta G^{N}_{x} = \Delta G^{I\to N}_{x} - \Delta G^{I\to N}_{WT}. \quad (8b)$$

Thus, the linkage free energies in I_C and N (Fig. 5) were obtained from:

$$\Delta\Delta G^{I}_{link} = \Delta\Delta G^{I}_{\exp} \quad (9a)$$

$$\Delta\Delta G^{N}_{link} = \Delta\Delta G^{N}_{\exp} + \Delta\Delta G^{I}_{\exp} \quad (9b)$$

Using the double mutant cycles and a triple mutant (*ijk*), the thermodynamic linkage between interactions *i* an *j* in background of mutant *k* can be obtained (Siegfried and Bevilacqua, 2009), yielding

$$\Delta\Delta G'^{I}_{\exp} = (\Delta G^{U\to I}_{ijk} + \Delta G^{U\to I}_{k} - \Delta G^{U\to I}_{ik} - \Delta G^{U\to I}_{jk}) \quad (10a)$$

$$\Delta\Delta G'^{N}_{\exp} = (\Delta G^{I\to N}_{ijk} + \Delta G^{I\to N}_{k} - \Delta G^{I\to N}_{ik} - \Delta G^{I\to N}_{jk}) \quad (10b)$$

Equations 9a and 9b are used to calculate the actual linkage energy in the mutant ($\Delta\Delta G'_{link}$) from experimental values above (Fig 5).

Linking frequencies of tertiary interactions from coupling free energies

We define linking frequency for a pair of tertiary interaction motifs as the fraction of molecules in the ensemble of RNA structures that contain both motifs. If the two motifs form independently, linking frequency is simply the product of relative saturation of each motif. Therefore, at 50% saturation of each motif, 25% of the molecules contain both interactions, 50% contain only one, and 25% have formed neither of the interactions.

In cases where formation of interaction motifs are coupled (positively or negatively), coupling free energies obtained from double mutant cycles can be used to calculate linking frequencies of the interactions. Consider the hypothetical equilibrium among all possible species in an ensemble of RNA structures:

$$R_i + R_j \xleftrightarrow{K_{link}} R_{ij} + R$$

The *i* and *j* indices indicate different tertiary interaction motifs that in the RNA structure. The free energy change associated with this reaction (with respect to a reference state) is

$$\Delta G = -RT \ln(K_{link}) = \Delta G(R_{ij}) + \Delta G(R) - \Delta G(R_i) - \Delta G(R_j) \qquad (11)$$

This is the same as eq. 7a, which is used to obtain the coupling free energies between pairs of interactions from double mutant cycles. Intuitively, as interactions become more cooperative, they are more likely to exist together in one molecule, and have a higher linking frequency (R_{ij}). At any given fractional saturation of interaction motifs *i* and *j*, f_i and f_j,

$$f_i = R_i + R_{ij} \qquad (12a)$$

$$f_j = R_j + R_{ij}, \qquad (12b)$$

Therefore, knowledge of the coupling free energy (ΔG_{link}) between the two interactions is enough to calculate linking frequency (R_{ij}):

$$\frac{R_{ij}}{R_{total}} = \frac{1}{2}\left(f_i + f_j + \frac{1}{|K-1|} - \sqrt{\frac{1}{(K-1)^2} + (f_i - f_j)^2 + \frac{2(f_i + f_j - 2f_i f_j)}{K-1}}\right); \qquad K \neq 1 \quad (13a)$$

$$\frac{R_{ij}}{R_{total}} = f_i \times f_j; \qquad K=1 \quad (13b)$$

Particularly, at half-saturation of each interaction ($f_i = f_j = 0.5$)

$$\frac{R_{ij}}{R_{total}} = \frac{1}{2}\left(\frac{\sqrt{K}}{1+\sqrt{K}}\right) \quad (14)$$

SUPPLEMENTARY REFERENCES

Adams, P.L., Stahley, M.R., Gill, M.L., Kosek, A.B., Wang, J., and Strobel, S.A. (2004a). Crystal structure of a group I intron splicing intermediate. RNA *10*, 1867-1887.

Adams, P.L., Stahley, M.R., Kosek, A.B., Wang, J., and Strobel, S.A. (2004b). Crystal structure of a self-splicing group I intron with both exons. Nature *430*, 45-50.

Battle, D.J., and Doudna, J.A. (2002). Specificity of RNA-RNA helix recognition. Proc. Natl. Acad. Sci. U. S. A. *99*, 11676-11681.

Downey, C.D., Fiore, J.L., Stoddard, C.D., Hodak, J.H., Nesbitt, D.J., and Pardi, A. (2006). Metal ion dependence, thermodynamics, and kinetics for intramolecular docking of a GAAA tetraloop and receptor connected by a flexible linker. Biochemistry *45*, 3664-3673.

Horovitz, A., and Fersht, A.R. (1992). Co-operative interactions during protein folding. J. Mol. Biol. *224*, 733-740.

Ikawa, Y., Naito, D., Shiraishi, H., and Inoue, T. (2000). Structure-function relationships of two closely related group IC3 intron ribozymes from Azoarcus and Synechococcus pre-tRNA. Nucleic Acids Res. *28*, 3269-3277.

Jaeger, L., Westhof, E., and Michel, F. (1993). Monitoring of the cooperative unfolding of the sunY group I intron of bacteriophage T4. The active form of the sunY ribozyme is stabilized by multiple interactions with 3' terminal intron components. J. Mol. Biol. *234*, 331-346.

Michel, F., Hanna, M., Green, R., Bartel, D.P., and Szostak, J.W. (1989). The guanosine binding site of the Tetrahymena ribozyme. Nature *342*, 391-395.

Michel, F., and Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J. Mol. Biol. *216*, 585-610.

Moghaddam, S., Caliskan, G., Chauhan, S., Hyeon, C., Briber, R.M., Thirumalai, D., and Woodson, S.A. (2009). Metal ion dependence of cooperative collapse transitions in RNA. J. Mol. Biol. *393,* 753-764.

Rangan, P., Masquida, B., Westhof, E., and Woodson, S.A. (2003). Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. Proc. Natl. Acad. Sci. U. S. A. *100,* 1574-1579.

Siegfried, N.A., and Bevilacqua, P.C. (2009). Thinking inside the box: designing, implementing, and interpreting thermodynamic cycles to dissect cooperativity in RNA and DNA folding. Methods Enzymol. *455,* 365-393.

Silverman, S.K., Zheng, M., Wu, M., Tinoco, I.,Jr, and Cech, T.R. (1999). Quantifying the energetic interplay of RNA tertiary and secondary structure interactions. RNA *5,* 1665-1674.

Tanner, M.A., and Cech, T.R. (1997). Joining the two domains of a group I ribozyme to form the catalytic core. Science *275,* 847-849.

Young, B., Herschlag, D., and Cech, T.R. (1991). Mutations in a nonconserved sequence of the Tetrahymena ribozyme increase activity and specificity. Cell *67,* 1007-1019.

Weber, G. (1975). Energetics of ligand binding to proteins. Adv. Protein Chem. *29,* 1-83.