Supporting Information

Structure-guided engineering of the regioselectivity of RNA ligase ribozymes

Jason N. Pitt[a,b] & Adrian R. Ferré-D'Amaré[b]*

a. Molecular and Cellular Biology Program, University of Washington, Seattle, WA 98195, USA

b. Howard Hughes Medical Institute and Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle WA 98109-1024, USA

* To whom correspondence should be addressed, e-mail: aferre@fhcrc.org; phone (206) 667-3622; facsimile (206) 667-3331

Supporting Discussion

The 2'-5' RNA World

The role of 2'-5' linked nucleic acids in the RNA world is not clear. The prevalence of 2'-5' linkages resulting from the polymerization of activated mononucleotides in solution (reviewed in [1,2]), their presence in the genomes of parasitic nucleic acids[3] and the role of 2'-5' oligoadenylic acid as an activator in anti-viral defense mechanisms[4], together suggest that 2'-5' linkages may have been important in the RNA world. However, the decreased thermostability of double helices containing 2'-5' linkages[5,6], their sensitivity to hydrolysis[7] and their susceptibility to sequence effects[8] would conceivably be selective barriers against any organism relying on a 2'-5' linked RNA genetic polymer. Nonetheless, the ability of a polymerase ribozyme to catalyze the polymerization of disparate substrates would likely increase the structural and functional diversity available to an RNA based organism. In fact some common RNA motifs are actually more stable when joined by 2'-5' linkages[9]. Protein polymerases are able to copy oligonucleotides that contain 2'-5' linkages[10], indicating that there is no reason, *a priori,* that an RNA based genetic system would have excluded 2'-5' linkages.

Catalytic Mechanism of Ligase Ribozymes

Protein-based nucleic acid polymerases employ a conserved two-metal ion mechanism. One divalent ion activates the attacking hydroxyl, the second coordinates the γ and β phosphates, assisting the pyrophosphate leaving group, and both metal ions shield the excess negative charge of the pentacovalent transition state[11]. Mg^{2+} ions are critical for

activity of both the L1 and the class II ligases, but there are no data to indicate that these ions are directly involved in catalysis. In our class II ligase product-template structures, no metal ions are within 8 Å of the relevant functional groups, suggesting that any catalytic cations, if present, must be brought to the active site through tertiary interactions with peripheral elements outside the substrate-template duplex. In the L1 ligase structure, a well-ordered Mg^{2+} ion is present in the active site, nucleating a critical tertiary interaction in the docked conformer. However, the coordination geometry of this Mg^{2+} is not ideal for activating the nucleophilic 3'-OH of U71 and its role in binding the pyrophosphate leaving group is unknown[12]. The structures of both the L1 and class II ligases are of product complexes of the ribozymes. It is possible that either the affinity or coordination geometry of these putative metal ion binding sites in the precursor state is different from the product state, making it difficult to deduce a mechanism based on the available structures. Nonetheless, both the L1 and the Class II ligases are obligate metalloenzymes requiring Mg^{2+} concentrations much higher than the 10-100 μM dissociation constants of free nucleoside triphosphates for Mg^{2+}, suggesting that the γ and β phosphates coordinate Mg^{2+} in the precursor. In the L1 ligase, were the complete structure of the ribozyme is known, how the nucleophilic hydroxyl is activated remains an open question[12]. In this ribozyme, the non-bridging phosphate oxygen of A39, which coordinates a Mg^{2+} ion , is equidistant (2.8 Å) from the 2' and 3' OHs, but in the precursor it has been predicted to form an activated hydrogen bond to the 3' OH, owing to a well order water molecule that could quench activation of the 2'OH [12]. Alternatively, it is also plausible that it is the Mg^{2+} bound by A39 that directly activates the 3' nucleophile for attack, and the proximity of this ion to the 3'-OH (2.8 Å), compared to the 2'-OH (4.1 Å),

helps explain the strong 3'-5' regioselectivity we observe when U71 is mutated to C (Figure 5 C,D), which should remove many of the hydrogen bonds to the critical water molecule Robertson and Scott predict to quench the activated 2' nucleophile[12].

Modularity of RNAs and Recombinatorial Evolution?

Contemporary genetic systems can rely on heterologous recombination to drive the evolution of novel protein enzymes by the chimeric assembly of disparate genetic elements. The local folding and inherent structural flexibility of RNAs, which results from their chemically similar monomers and stereotypical structural motifs[13], may make recombination an even greater factor in the evolution of new ribozymes[14]. Our experiments relied on this flexibility to aid in the rational engineering of ligase ribozymes, resulting in catalysts with simple changes in regioselectivity. However, in the RNA world, primitive genetic systems likely exploited these same traits, through recombination, to generate the diverse array of ribozymes that would have been required for life.

Table S1. Crystallographic statistics

		Crystal form I			Crystal form II
Data collection					
		Crystal 1		**Crystal 2**	**Crystal 3**
Unit Cell Dimensions					
a, b, c (Å)		31.17,31.17, 169.09		31.55, 31.55, 171.07	30.04, 30.04, 59.67
Space Group		$P4_1$		$P4_1$	$P3_1$
Wavelength (Å)	0.91500	0.92004	0.91981	1.00000	1.00000
Resolution (Å)	30-2.9	30-2.9	30-2.9	30-2.7	30-2.3
Reflections[a] /Redundancy	3541/9.9	2163/5.0	3544/10.0	4565/3.8	2663/5.7
R_{sym} (%)	8.3 (42.6)[b]	6.2 (32.1)	7.3 (39.2)	4.6 (44.2)	4.2 (30.7)
<I>/<σ(I)>	38.8 (5.9)	31.9 (5.0)	43.3 (6.0)	29.4 (2.9)	31.3 (4.6)
Completeness (%)	98.4 (99.4)	98.4 (99.4)	98.5 (99.4)	98.7 (99.6)	99.5 (99.6)
Phasing					
IF_HI/E					
ano	0.71 (0.1)	1.89 (0.72)	2.68 (1.56)		
iso		1.78 (0.75)	2.55 (1.43)		
R_{Kraut} (%)					
ano	3.4 (4.8)	4.2 (12.6)	5.1 (13.7)		
iso		4.3 (13.0)	5.1 (13.4)		
Mean Figure of Merit		0.58 (0.32)			
Refinement					
Atoms (total/water/ions)				999/6/11	466/12/5
R_{work}/R_{free}				21.8/26.8	20.8/23.4
Mean *B*-factor				44.6	43.4
RMSD bond lengths/angles (Å / °)				0.005/1.05	0.005/1.2
Cross-validated σA coordinate error (Å)				0.6	0.4
PDB accession code				3FTM	3FS0

[a] Unique reflections

[b] Values in parentheses are for highest resolution shell.

Table S2. Rates and regioselectivities of ligase ribozyme mutants

	Mutant	Rate (min^{-1})	σ (min^{-1})	R	Regioselectivity (2'-5':3'-5')	Yield 2'-5' %product [b]	σ (%)
Class II ligase a4-11	WT	5.5×10^{-1}	-	0.991	77 : 1 [d]	100	3
	C-1U	2.3×10^{-3}	3×10^{-6}	-	15 : 1	95	3
	C-1G	3.1×10^{-4}	5×10^{-6}	-	-	ND	-
	G49A	2.1×10^{-2}	3×10^{-6}	-	80 : 1	104	7
	G49A:C-1U	1.8×10^{-2}	2×10^{-6}	-	84 : 1	109	2
	G49A:C-1G	7.9×10^{-5}	5×10^{-6}	-	-	ND	-
	G49C	1.6×10^{-4}	7×10^{-6}	-	-	ND	-
	G49C:C-1U	4.1×10^{-4}	7×10^{-6}	-	-	ND	-
	G49C:C-1G	9.1×10^{-2}	-	0.999	-	ND	-
	G1A	3.8×10^{-4}	1×10^{-4}	-	85 : 1	110	4
	A48G	1.4×10^{-4}	9×10^{-5}	-	84 : 1	109	7
	G1A:A48G	1.0×10^{-5}	2×10^{-6}	-	-	ND	-
	G2A	1.1×10^{-1}	-	0.999	30 : 1	98	4
	G2C	3.5×10^{-2}	-	0.987	23 : 1	97	2
	G2U	1.3×10^{-1}	-	0.999	23 : 1	97	0.7
	ΔG2	4.4×10^{-4}	1×10^{-5}	-	2 : 1	67	2
	A3G	6.3×10^{-4}	5×10^{-5}	-	80 : 1	104	0.7
	A3C	8.0×10^{-5}	2×10^{-5}	-	43 : 1	99	6
	A3U	1.6×10^{-4}	3×10^{-6}	-	-	ND	-
	ΔA3	4.5×10^{-5}	3×10^{-5}	-	82 : 1	107	0.7
	G47A	2.0×10^{-5}	8×10^{-6}	-	77 : 1	100	18
	G47C	1.9×10^{-5}	8×10^{-7}	-	77 : 1	100	16
	G47U	9.5×10^{-6}	6×10^{-7}	-	-	ND	-
	G47U:C-1U	1.3×10^{-6}	8×10^{-8}	-	-	ND	-
	ΔG2:C-1U	6.0×10^{-6}	1×10^{-7}	-	1 : 3	27	1
	G47U:ΔG2:C-1U	3.3×10^{-7}	2×10^{-8}	-	< 1 : 6	14	6
	G47U:ΔA3	3.9×10^{-7}	1×10^{-7}	-	-	ND	-
	G47U:ΔA3:C-1U	3.8×10^{-6}	-	0.999	< 1 : 126	0.8	3
Class II substrate-template duplex alone [a]	WT	7.5×10^{-7}	2×10^{-8}	-	-	ND	-
	G47U:ΔA3:C-1U	4.8×10^{-7} [c]	7×10^{-8}	-	-	ND	-
L1 Ligase L1x6c	WT	7.3×10^{-2}		0.994	1 : 525	0.19	0.04
	U71C	3.2×10^{-4}	1×10^{-5}	-	< 1 : 4300	0.023	0.02
	U50G	ND	-	-	< 1 : 4300	0.024	0.02
	U50G:U71C	3.5×10^{-4}	2×10^{-5}	-	-	ND	-
	U50G:A3insertion	5.1×10^{-4}	7×10^{-6}	-	-	ND	-
	U50G:A3insertion:U71C	7.4×10^{-5}	-	0.993	1 : 14	6.6	0.3
L1 substrate-template duplex alone [a]	WT	1.2×10^{-6} [c]	2×10^{-7}	-	-	ND	-
	U50G:A3insertion:U71C	7.9×10^{-7}	8×10^{-8}	-	-	ND	-

ND = Not determined WT = wildtype

[a] Residues 6 - 44 of class II ligase A4-11 replaced by a three base pair stem capped by a GAAA tetraloop and residues 8 – 44 of L1 ligase L1x6c replaced by a GAAA tetraloop.

[b] Percentage yield 2'-5' yields for class II mutants are normalized to that of the wildtype class II ligase a4-11 [15]

[c] The difference between the isolated substrate-template duplexes of the engineered class II and the wildtype L1 ligases is statistically significant. The difference in rates may reflect the increased thermostability conferred by the more stable Watson-Crick duplex 5' of the ligation junction (on the substrate strand) in the L1 substrate-template duplex since the wildtype L1 substrate-template duplex is 2 base pairs longer than the engineered class II substrate-template duplex, which only contains the mutations G47U:ΔA3:C-1U relative to the class II substrate-template duplex.

[d] Reference [15]

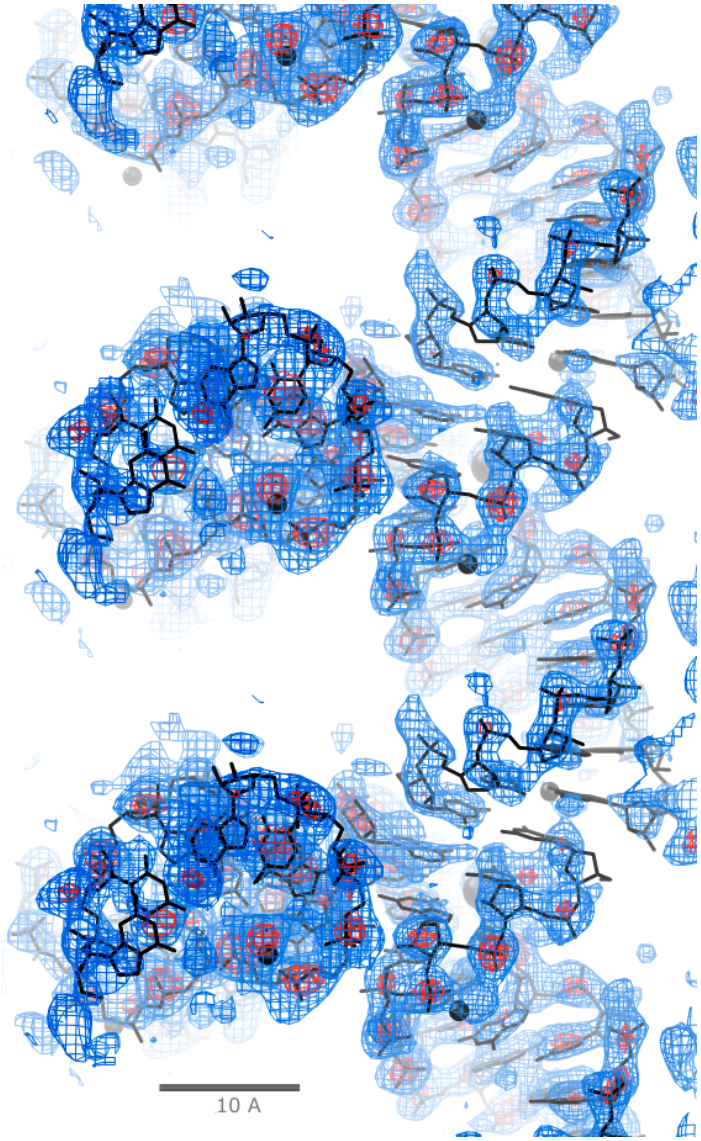

Figure S1. Experimental electron density map Portion of the density-modified experimental electron density map calculated with MAD phases from crystal 1, phase-extended to 2.7 Å using data from crystal 2, and crystal 2 structure factor amplitudes, contoured at 1.2 σ and 4σ (blue and red mesh, respectively). Final refined models from three adjacent asymmetric units are shown in black to display pseudo-infinite helix packing in the crystal.

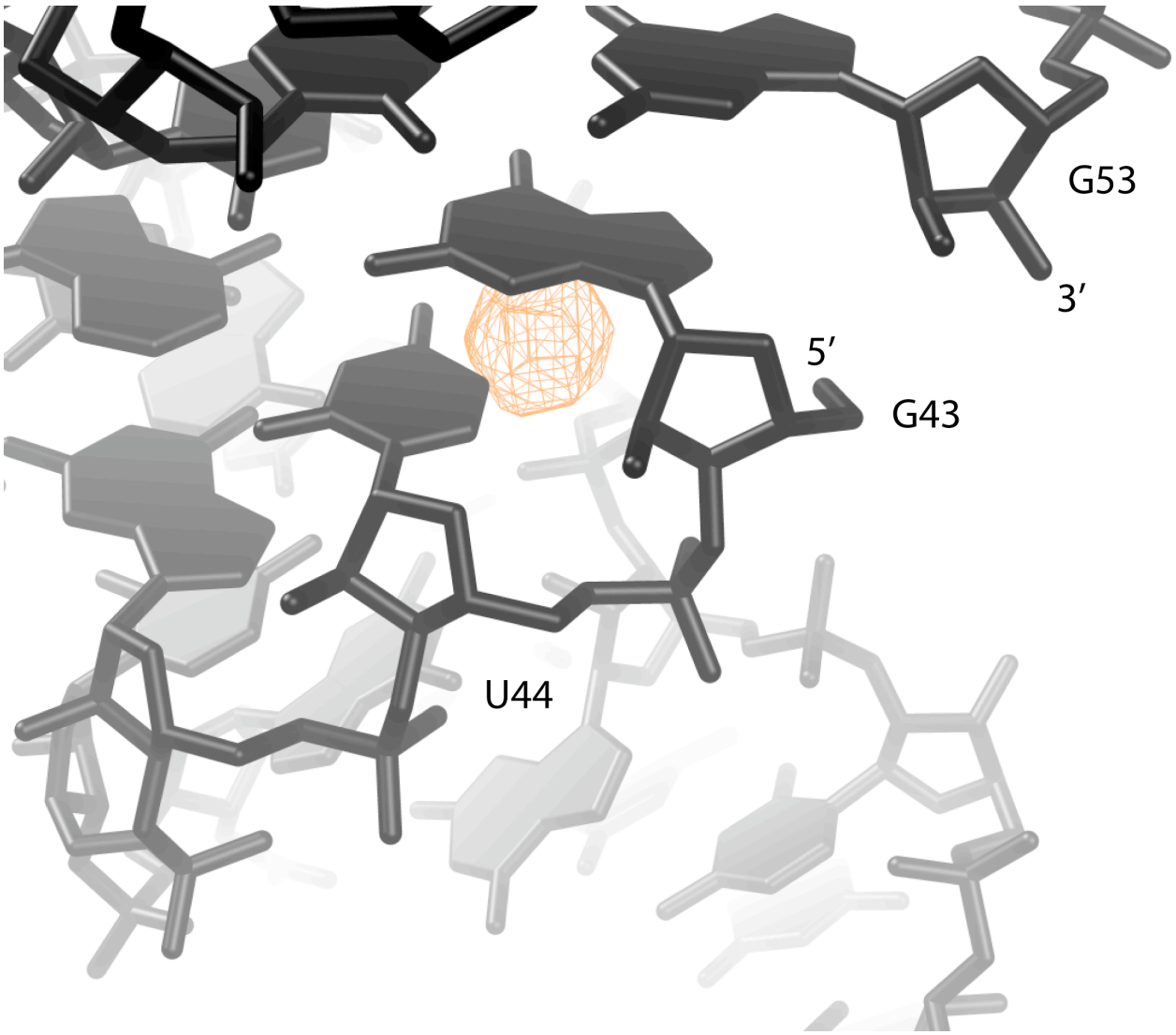

Figure S2. Anomalous difference Fourier map of crystal form I A portion of the anomalous difference Fourier map calculated with amplitudes from the 0.91981 Å dataset of Crystal 1 and the density-modified MAD phases, contoured at 10 σ (orange mesh). Peaks were used to locate 5-bromouridine residues in the pseudo-infinite helix prior to model building.

References for Supporting Information

(1) Orgel, L. E. *Cold Spring Harb Symp Quant Biol* **1987**, *52*, 9-16.

(2) Joyce, G. F. *Cold Spring Harb Symp Quant Biol* **1987**, *52*, 41-51.

(3) Cote, F.; Levesque, D.; Perreault, J. P. *J Virol* **2001**, *75*, 19-25.

(4) Player, M. R.; Torrence, P. F. *Pharmacol Ther* **1998**, *78*, 55-113.

(5) Kierzek, R.; He, L.; Turner, D. H. *Nucleic Acids Res* **1992**, *20*, 1685-1690.

(6) Wasner, M.; Arion, D.; Borkow, G.; Noronha, A.; Uddin, A. H.; Parniak, M. A.; Damha, M. J. *Biochemistry* **1998**, *37*, 7478-7486.

(7) Usher, D. A.; McHale, A. H. *Proc Natl Acad Sci U S A* **1976**, *73*, 1149-1153.

(8) Premraj, B. J.; Raja, S.; Bhavesh, N. S.; Shi, K.; Hosur, R. V.; Sundaralingam, M.; Yathindra, N. *Eur J Biochem* **2004**, *271*, 2956-2966.

(9) Hannoush, R. N.; Damha, M. J. *J Am Chem Soc* **2001**, *123*, 12368-12374.

(10) Lorsch, J. R.; Bartel, D. P.; Szostak, J. W. *Nucleic Acids Res* **1995**, *23*, 2811-2814.

(11) Steitz, T. A. *Nature* **1998**, *391*, 231-232.

(12) Robertson, M. P.; Scott, W. G. *Science* **2007**, *315*, 1549-1553.

(13) Leontis, N. B.; Lescoute, A.; Westhof, E. *Curr Opin Struct Biol* **2006**, *16*, 279-287.

(14) Lehman, N.; Unrau, P. J. *J Mol Evol* **2005**, *61*, 245-252.

(15) Ekland, E. H.; Szostak, J. W.; Bartel, D. P. *Science* **1995**, *269*, 364-370.