J|A|C|S ARTICLES

Published on Web 02/16/2009

Structure-Guided Engineering of the Regioselectivity of RNA Ligase Ribozymes

Jason N. Pitt[†,‡] and Adrian R. Ferré-D'Amaré*[,‡]

Molecular and Cellular Biology Program, University of Washington, Seattle, Washington 98195, and Howard Hughes Medical Institute and Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109-1024

Received August 28, 2008; E-mail: aferre@fhcrc.org

Abstract: Ribozyme-catalyzed RNA synthesis is central to the RNA world hypothesis. No natural RNA polymerase ribozymes have been discovered. However, ribozymes that catalyze the requisite chemistry, generating a new phosphodiester through attack of a terminal hydroxyl of an RNA on the α-phosphate of a triphosphate-activated oligonucleotide, have been isolated by *in vitro* selection. These experiments often yield ribozymes that generate 2′−5′ phosphodiesters rather than conventional 3′−5′ linkages. We have determined crystal structures of the duplex formed by the template segment of a representative 2′−5′ RNA ligase ribozyme, the class II ligase, and its ligation product. The structures reveal a product−template duplex with a G·A pair at the ligation junction. This sheared pair is flanked on one side by a minor groove-broadening wedge comprised of two unpaired nucleotides. The reported structure of an independently isolated 3′−5′ ligase ribozyme, the L1 ligase, shows a product−template duplex that shares the G·A pair with the class II ligase. However, this G·A pair is flanked by G·U wobbles, rather than an unpaired wedge. We demonstrate that these structural differences of the substrate−template duplexes are largely responsible for the divergent regioselectivity of the two ribozymes, independent of their catalytic moieties, by constructing chimeras. The L1 ligase with a class II substrate−template duplex shows a 30-fold increase in 2′−5′ bond synthesis, while the class II ligase with an L1 substrate−template duplex produces 3′−5′ bonds exclusively. These results demonstrate how local geometry inherent to the substrate−template duplexes controls the regioselectivity of ribozyme-catalyzed RNA ligation reactions.

Introduction

Several ligase ribozymes that catalyze a chemical transformation analogous to a single step of RNA polymerization (Figure 1A) have been isolated by *in vitro* selection.[1−5] In these reactions, a ribose OH group carries out a nucleophilic attack on the α-phosphate of either a nucleotide triphosphate or an RNA with a 5′ triphosphate (for polymerases and ligases, respectively). Because the 3′-terminal ribose of an RNA bears two free OH groups, ligases and polymerases can, in principle, make either 2′−5′ or 3′−5′ phosphodiester bonds. Selections for polymerase-like ligases often yield ribozymes that strongly favor production of the 2′−5′ linked regioisomer.[2,6−8] In cells, 2′−5′ bonds are found in the lariat RNAs resulting from nuclear pre-mRNA and group II intron splicing. Linear oligonucleotides comprised exclusively of 2′−5′ bonds are made by oligoadenylate synthetase (OAS), a key player in the interferon response (see reviews in refs 9, 10). Remarkably, OAS is a close structural homologue of DNA polymerase β, a template- and primer-dependent enzyme (reviewed in ref 11). Thus, closely related enzyme active sites can synthesize either 2′−5′ or 3′−5′ linked nucleic acids. One of the underlying motivations for the *in vitro* selection of RNA ligase ribozymes is that they may be evolved or engineered into template-dependent RNA replicases[12,13] (reviewed in ref 14). A detailed understanding of the structural basis of 2′-OH vs 3′-OH regioselectivity is important for such a goal.

The class II ligase ribozymes are a family of RNAs that catalyze formation of 2′−5′ phosphodiester bonds with ∼5500-fold preference over 3′−5′ bonds and achieve a rate enhancement over background ligation of $\sim 10^{10}$.[7] Deletion analysis by Ekland et al. demonstrated that the duplex formed by the template segment of the ribozyme and the two RNA substrates (hereafter "substrate−template duplex"; Figure 1B) can by itself achieve $\sim 10^3-10^4$ rate enhancement and produces 2′−5′ linkages with 4.5:1 regioselectivity over 3′−5′ linked products, compared to 1:72.5 for a Watson−Crick duplex.[7] Thus, the substrate−template duplex alone appears to be responsible for

[†] University of Washington.
[‡] Fred Hutchinson Cancer Research Center.

(1) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411–1418.
(2) Landweber, L. F.; Pokrovskaya, I. D. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 173–178.
(3) Robertson, M. P.; Ellington, A. D. *Nat. Biotechnol.* **1999**, *17*, 62–66.
(4) Ikawa, Y.; Tsuda, K.; Matsumura, S.; Inoue, T. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13750–13755.
(5) Rogers, J.; Joyce, G. F. *RNA* **2001**, *7*, 395–404.
(6) Wang, Y.; Silverman, S. K. *J. Am. Chem. Soc.* **2003**, *125*, 6880–6881.
(7) Ekland, E. H.; Szostak, J. W.; Bartel, D. P. *Science* **1995**, *269*, 364–370.
(8) Rogers, J.; Joyce, G. F. *Nature* **1999**, *402*, 323–325.
(9) Silverman, R. H. *Cytokine Growth Factor Rev.* **2007**, *18*, 381–388.
(10) Player, M. R.; Torrence, P. F. *Pharmacol. Ther.* **1998**, *78*, 55–113.
(11) Martin, G.; Keller, W. *RNA* **2007**, *13*, 1834–1849.
(12) Johnston, W. K.; Unrau, P. J.; Lawrence, M. S.; Glasner, M. E.; Bartel, D. P. *Science* **2001**, *292*, 1319–1325.
(13) Zaher, H. S.; Unrau, P. J. *RNA* **2007**, *13*, 1017–1026.
(14) McGinness, K. E.; Joyce, G. F. *Chem. Biol.* **2003**, *10*, 5–14.

10.1021/ja8067325 CCC: $40.75

a 326-fold increase in the amount of 2′−5′ linked product formed, compared to a canonical Watson−Crick duplex.

We solved structures of the duplex formed by the template segment of the ribozyme and the ligation product (hereafter "product−template duplex") embedded between arbitrary base pairs that facilitated crystallization, in two different crystal forms (see Table S1 in the Supporting Information and Experimental Section). Crystal form I contained two copies of the duplex in the asymmetric unit and diffracted to 2.7 Å, while crystal form II contained one duplex in the asymmetric unit and diffracted to 2.3 Å. Although we were unable to obtain useful crystals of the unstable substrate−template duplex, our structures contain 2′−5′ linkages at the site of ligation and provide three crystallographically independent views of this functionally key moiety of the ribozyme. Our structures and the recently reported structure of the L1 ligase,[15] a ribozyme that regioselectively catalyzes 3′−5′ bond formation, provide the first opportunity to compare and contrast, at the atomic level, the active sites of RNAs that catalyze the same chemical transformation but with opposite regioselectivity. Our comparative analyses suggest that the geometry of the nucleotides immediately adjacent to the ligation junction largely controls ligase ribozyme regioselectivity. We have tested this prediction by carrying out structure-guided engineering of the regioselectivity of the class II and L1 ligases, without recourse to *in vitro* selection methods.

Experimental Section

RNA Synthesis. Full-length ribozymes were transcribed *in vitro*. Transcription reactions contained 30 mM Tris-HCl pH 8.1, 2 mM spermidine, 25 mM $MgCl_2$, 10 mM DTT, 0.01% (v/v) Triton X-100, 5 mM each of ATP, CTP, GTP, and UTP, 0.1 g/L T7 RNA polymerase, 1 unit/mL *Escherichia coli* inorganic pyrophosphatase[16] (Sigma-Aldrich, St. Louis, MO), and 45% (v/v) of a polymerase chain reaction containing PCR products with 3′-terminal 2′-methoxyribose residues to limit 3′ end heterogeneity.[17] Transcription reactions to produce body-labeled ribozymes contained 0.17 mM α-^{32}P GTP and 2.5 mM concentrations of the other NTPs. After 4 h at 310 K, reactions were stopped by adjusting them to contain 45% (v/v) formamide and 0.5X TBE. RNAs were purified by denaturing urea−PAGE, passively eluted overnight at 277 K into water, filtered, and precipitated in 0.02 g/L glycogen (Roche, Basel, Switzerland), 0.3 M sodium acetate pH 4.5, and 66% (v/v) ethanol. Pellets were washed with 70% (v/v) ethanol, dried, resuspended in DEPC-treated water, and quantified by either UV spectrometry or scintillation counting.

RNA oligonucleotides used for crystallization or as substrates for biochemical analyses were chemically synthesized and were purchased from either Dharmacon (Lafayette, CO) or the W.M. Keck Foundation Biotechnology Resource Laboratory at Yale University (New Haven, CT). The 3′-(OtBDS)-2′-phosphoramidite used to synthesize the desired 2′−5′ linkage was purchased from Chemgenes (Wilmington, MA). Oligonucleotides from Dharmacon were deprotected following the manufacturer's instructions. Keck Laboratory oligonucleotides were deprotected in 1 M tetrabutylammonium fluoride in THF overnight at 295 K. The reaction was quenched by adding Tris pH 7.4 to a final concentration of 500 mM, concentrated under vacuum, and desalted by size exclusion chromatography on Sephadex G25 using DEPC-treated water as the eluent. Oligonucleotides were lyophilized, resuspended in DEPC-treated water, and quantified by UV spectrometry at 260 nm. Extinction coefficients for each oligonucleotide are given below.

Crystallization and Data Collection. Oligonucleotides for crystallization were annealed by heating for 2 min to 333 K at a concentration of 0.2 mM in a buffer containing 5.0 mM Tris pH 7.5 and 0.05 mM EDTA, followed by cooling at a rate of 2 K min^{-1} to 295 K. Exact stoichiometry for crystallization stocks was determined by titration of one oligonucleotide against the other and monitoring complex formation by native gel electrophoresis at 283 K. Crystal form I (Table S1) contained the oligonucleotides CCAGUC*GGAACA (ϵ = 120 400 M^{-1} cm^{-1}), where the asterisk represents a 2′−5′ linkage, and GXGUGAGGCUG, where X is 5-bromouridine in crystal 1 (ϵ = 106 300 M^{-1} cm^{-1}) and uridine in crystal 2 (ϵ = 109 500 M^{-1} cm^{-1}). Crystal 1 was grown by vapor diffusion at 295 K by mixing equal volumes of a solution containing 0.2 mM annealed oligonucleotides in 15 mM Tris pH 7.5, 10 mM $MgCl_2$, 0.05 mM EDTA and a reservoir solution comprising 15% (w/v) PEG 8000, 50 mM sodium cacodylate pH 6.5, 200 mM magnesium acetate and 200 mM KCl. Crystal 2 was grown in the same manner as crystal 1, but with a reservoir solution that lacked KCl. Crystals 1 and 2 both grew as two fused single crystals. The crystals were bisected with a nylon loop prior to transfer to a 5 μL drop of cryoprotectant solution comprised of the reservoir solution supplemented with 15% (v/v) glycerol. Crystal 1 was immediately flash frozen in liquid N_2 after transfer to the cryoprotectant; crystal 2 was allowed to dehydrate for 10 min at 295K in the cryoprotectant prior to freezing. Crystal form II (Table S1) contained the oligonucleotides CCAGUC*GGAAC (ϵ = 106 600 M^{-1} cm^{-1}), where the asterisk represents a 2′−5′ linkage, and GGUGAGGCUG (ϵ = 99 800 M^{-1} cm^{-1}). Crystal 3 was grown by vapor diffusion at 295 K by mixing equal volumes of a solution containing 0.2 mM annealed oligonucleotides, 15 mM Tris pH 7.5, and 10 mM $MgCl_2$ and a reservoir solution comprising 25% (w/v) PEG 4000, 200 mM Li_2SO_4, and 100 mM Tris pH 8.5. The crystal was removed from its mother liquor and directly flash-frozen by plunging into liquid nitrogen. Diffraction intensities were collected at beamlines 5.0.1 and 5.0.2 of the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory, and reduced using the HKL package.[18] Intensity statistics are shown in Table S1.

Structure Determination and Refinement. Two bromine sites were located and phases calculated using MAD data from Crystal 1 with CNS[19] (see Table S1). Density modification by solvent flipping and histogram matching produced an electron density map that allowed the RNA model to be built using O[20] (Figure S1, Supporting Information). The location of the bromine atoms unambiguously identified helical register prior to building (Figure S2, Supporting Information). Rounds of manual rebuilding, simulated annealing, energy minimization, and individual *B*-factor refinement, restraining sugar pucker, base planarity, and noncrystallographic symmetry, produced the current model (R_{free} 26.8%). The structure of crystal form II was determined using amplitudes from crystal 3 by molecular replacement with the crystal form I structure containing the *anti* conformation of G47, using the program EPMR.[21] The top molecular replacement solution had a correlation coefficient of 0.83 and a crystallographic *R*-factor of 0.41. One round of refinement using simulated annealing, energy minimization, and individual *B*-factor refinement produced a model with a free *R*-factor of 26%. The crystal 3 structure was further refined using rounds of manual rebuilding (repositioning G47 to the *syn* conformation), simulated annealing, energy minimization, and individual *B*-factor refinement while restraining sugar pucker and base planarity. Final crystallographic and refinement statistics are shown in Table S1. Alternatively modeling the 2′−5′ phos-

(15) Robertson, M. P.; Scott, W. G. *Science* **2007**, *315*, 1549–1553.
(16) Rupert, P. B.; Ferré-D'Amaré, A. R. *Methods Mol. Biol.* **2004**, *252*, 303–311.
(17) Kao, C.; Zheng, M.; Rudisser, S. *RNA* **1999**, *5*, 1268–1272.
(18) Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307.
(19) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr. D: Biol. Crystallogr.* **1998**, *54*, 905–921.
(20) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. *Acta Crystallogr. A* **1991**, *47*, 110–119.
(21) Kissinger, C. R.; Gehlhaar, D. K.; Fogel, D. B. *Acta Crystallogr. D: Biol. Crystallogr.* **1999**, *55*, 484–491.

phodiester bond as a 3′−5′ bond resulted in a 5σ peak in the $|F_o| - |F_c|$ map and a 0.3% increase in R_{free} following refinement. Figures were prepared using UCSF Chimera.[22]

Structure−Function Analysis and Ribozyme Engineering. All assays were performed in triplicate. Single time point measurements were performed under the following conditions: 0.01 mM ribozyme and 100 nM ^{32}P-labeled oligonucleotide substrate, in a buffer containing 30 mM Tris pH 7.4, 200 mM KCl, 60 mM $MgCl_2$, and 0.6 mM EDTA, incubated for 17 h at 295 K, stopped by the addition of 1 volume of a solution of 50% (v/v) formamide, 1X TBE, and 200 mM EDTA, and separated on 20% 8 M urea−PAGE. Gels were analyzed on a Typhoon phosphorimager (GE Healthcare, Piscataway, NJ) and quantified with ImageQuant software (GE Healthcare). Activity was defined as the mole fraction of substrate converted to product divided by time (minutes). Multiple time point experiments were performed for chimeric ligases under the following conditions: 1 μM ribozyme, 5 nM ^{32}P-labeled oligonucleotide substrate, 30 mM Tris pH 7.4, 200 mM KCl, 60 mM $MgCl_2$, and 0.6 mM EDTA. Reactions were initiated by the simultaneous addition of buffer and substrate to the ribozyme. Reactions were incubated at 295 K, and aliquots were taken at multiple time intervals, stopped, and analyzed as above. The reported rate is the slope of the linear portion of a graph of the fraction of substrate converted to product against time.

For regioselectivity assays, oligonucleotide substrates contained a 5′ biotin group, and ribozymes were body-labeled with α-^{32}P GTP. Excess body-labeled ribozyme was allowed to react for up to 96 h at 295 K in a solution containing 1.75 nmol of biotinylated substrate and a buffer containing 30 mM Tris pH 7.4, 200 mM KCl, 60 mM $MgCl_2$, and 0.6 mM EDTA. Excess substrate was removed before immobilization on streptavidin beads by ultrafiltration using a YM-10 Microcon centrifugal concentrator (Millipore, Billerica, MA) and washing with 7 M urea, followed by multiple washes with water at 313 K in a heated microcentrifuge. Biotinylated products were captured on 1 mg streptavidin magnetic beads (Dynal M-270, Invitrogen, Carlsbad, CA) pretreated following the manufacturer's instructions, concentration of 5 g/L, washed at 315 K with binding buffer, DEPC water containing 0.1% Triton X-100, and twice with DEPC water. Beads were then resuspended in 50 μL of a solution containing 40 mM ammonium acetate pH 4.5, 50 μg of carrier RNA (yeast total RNA or a synthetic oligonucleotide containing a 2′−5′ linkage), and 3 units RNase T2 (see below) and incubated for 1 h at 310 K with agitation. Due to supply shortage at the time of these experiments, RNase T2 was purified from Takadiastase (Fluka, Buchs, Switzerland) as described previously.[23] Beads were captured, and the supernatant containing digested RNA was phenol/chloroform extracted, dried under vacuum, and resuspended in 1 μL of DEPC water. Digestion products were analyzed by 2D thin-layer chromatography[24] on 20 × 20 cm PEI cellulose plates (Selecto Scientific, Suwanee, GA) that had been prerun in DI water, washed in absolute methanol, and dried. Digested RNA was spotted on the bottom corner, 1.5 cm from both edges. Plates were dried, washed with absolute methanol, and then dried again. To increase throughput, plates were rolled into tubes, secured with rubber bands, and developed in a flat-bottomed bucket with a screw-sealed lid containing an O-ring. A dozen plates could be run simultaneously in each bucket. Development in the first dimension was in 2.5 M acetic acid, pH-adjusted to 3.5 with ammonium hydroxide, and continued until the buffer was 1 cm from the top edge of the plate. Plates were washed in absolute methanol for 20 min and dried. For the second dimension, plates were rotated 90° and developed in buffer Sb,[24] which is a solution made by adding 444 g of ammonium sulfate and 24 g of disodium EDTA to 600 mL of deionized water. The pH of the resulting solution is adjusted to 3.5 with solid ammonium hydrogen sulfate. Plates where developed until the buffer reached 1 cm from the top of the plate, air-dried, and analyzed using a phosphoimager as above. For the L1 ligase, regioselectivity was determined as a percentage yield of the 2′−5′ linked product as a fraction of the total product, calculated as the measured counts of the 2′−5′ linked dinucleotide divided by the expected number of counts per phosphate. Counts per phosphate were the total counts of Ap, Cp, Up, Gp, and 2′−5′ dinucleotide divided by the expected number of radioactive phosphates per ribozyme, based on sequence. Our results were consistent with the previously published value for the regioselectivity of the class II ligase, 77:1 (2′−5′:3′−5′),[7] and our reported ratios are based on this and our observed 2′−5′ product yields. Unlike CpGp and UpGp 2′−5′ linked dinucleotides, CpAp dinucleotides could not be resolved in this system from Gp. Reactions which contained CpAp dinucleotides, resulting from G1A mutations through the use of an A initiating promoter,[25] were instead analyzed by denaturing electrophoresis in 20% (w/v) acrylamide 19:1 acrylamide:bisacrylamide, 1X TBE, and 7 M urea on a 420 mm × 350 mm × 0.3 mm gel at 318 K. Gels were covered with plastic wrap and placed over a TLC plate containing fluorescent indicator and bands for control 2′−5′ linked dinucleotide monophosphates, and 3′ nucleotide monophosphates were identified by UV shadowing and marked with a pen. Gels were then quantified using a phosphoimager as above.

Results and Discussion

Structure of the Class II Ligase Product−Template Duplex. The class II ligase product−template duplex comprises a non-Watson−Crick paired segment 3′ of the ligation junction, flanked on both sides by standard A-form helices. Three features characterize the non-Watson−Crick segment. First, G1, the nucleotide that bears the triphosphate in the precursor state, forms a Hoogsteen•sugar-edge pair with A48 (Figure 1C). This sheared pairing is stabilized by cross-strand stacking of A48 on G2 (Figure 1D). Second, G2 and A3 do not form any base-pairing interactions across the helix and project into the minor groove, forming a dinucleotide spacer, or minor-groove-widening "wedge". This wedge forms an extensive π-stacking interaction resulting from A48 stacking on top of G2 and A3. Third, G47 projects into the major groove, where it does not make base-pairing interactions (Figures 1E,F). In our structures, the conformation of G47 is variable. In one structure, its nucleobase adopts the *syn* conformation, making a pair of hydrogen bonds from its exocyclic amine to nonbridging phosphate oxygens of G47 and U46 (Figure 1E), overtwisting the helix and allowing for coplanar pairing of A48 and G1. In the other structures, G47 is in the *anti* conformation and makes a pair of hydrogen bonds to the 2′-OH of G1 and N7 of G2 (Figure 1F). Analysis of the crystal contacts in our two crystal forms did not suggest any obvious influence of crystal packing on the orientation of the G47 glycosidic bond. We therefore conclude that, in the isolated product−template duplex, G47 can sample either conformation.

Previous structures and modeling studies indicated a strong tendency for sugars in 2′−5′ linked oligoribonucleotides to adopt the C2′ *endo* conformation, rather than the C3′ *endo* ribose pucker found in canonical A-form RNA helices.[26,27] These 2′−5′

(22) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
(23) Rushizky, G. W.; Sober, H. A. *J. Biol. Chem.* **1963**, *238*, 371–376.
(24) Bochner, B. R.; Ames, B. N. *J. Biol. Chem.* **1982**, *257*, 9759–9769.
(25) Coleman, T. M.; Wang, G.; Huang, F. *Nucleic Acids Res.* **2004**, *32*, e14.
(26) Premraj, B. J.; Yathindra, N. *J. Biomol. Struct. Dyn.* **1998**, *16*, 313–328.
(27) Premraj, B. J.; Patel, P. K.; Kandimalla, E. R.; Agrawal, S.; Hosur, R. V.; Yathindra, N. *Biochem. Biophys. Res. Commun.* **2001**, *283*, 537–543.

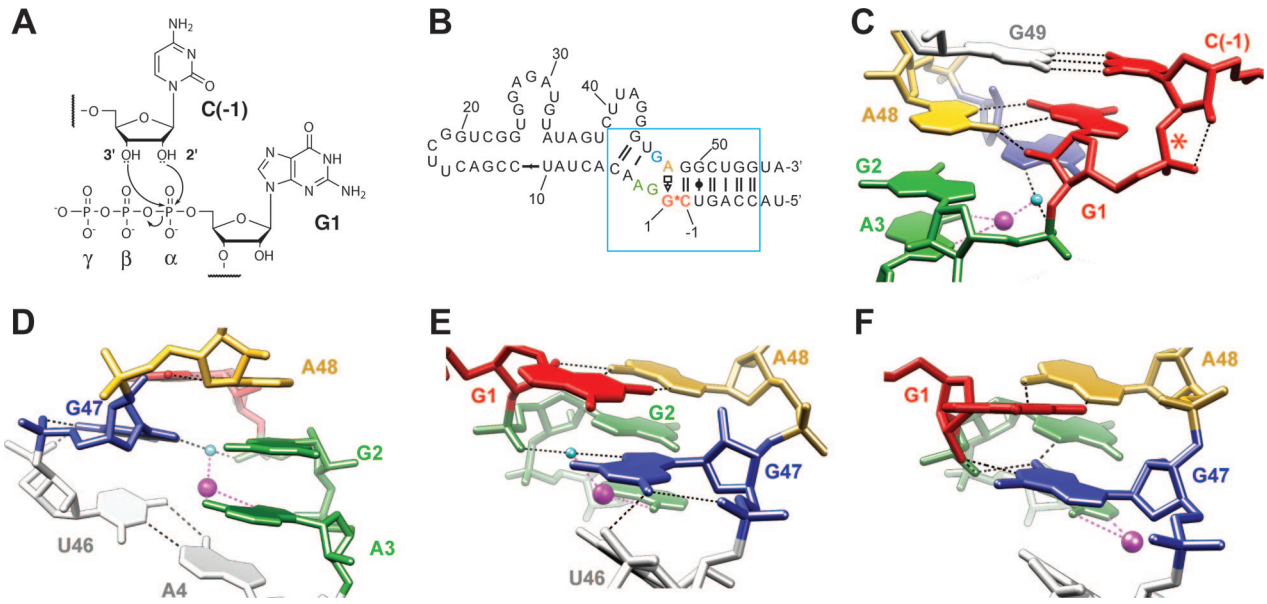

Figure 1. Structure of the product state of the class II ligase product−template duplex. (A) Chemical reaction catalyzed by RNA polymerase-like ligase ribozymes. The ligase template strand binds an RNA primer and catalyzes the attack of either of the terminal 2′-OH or 3′-OH on the α phosphate of the guanosine triphosphate residue at the 5′ end of the ribozyme, yielding a new phosphodiester bond and liberating pyrophosphate in a reaction that is analogous to a single step of RNA polymerization. (B) Secondary structure of the class II ligase A4−11.[7] Segment crystallized is shown in blue box. Residue numbering in all figures and text is based on that of the A4−11 ligase. Base-pairing symbols are as previously described.[45] Nucleotides 49−56 comprise the template strand, and nucleotides (−8) through (−1) are derived from the RNA substrate. The class II ligase catalyzes the attack of the 2′-OH of C(−1) on the 5′ triphosphate of G1. The 2′−5′ bond in the product is denoted by the asterisk. (C) The 2.3 Å crystal structure of the class II ligase product−template duplex from crystal 3, viewed from the minor groove. Inferred hydrogen bonds are shown as dashed lines. Well-ordered waters and Mg^{2+} ions are depicted as cyan and magenta spheres, respectively. Coloring of residues corresponds to panel B. (D) Cross-strand stacking of A48 (gold) on the unpaired steric wedge formed by G2 and A3 (green). (E) View from major groove side in crystal form II. G47 (blue) adopts the *syn* conformation, hydrogen-bonding to nonbridging phosphate oxygens of G47 and U46 (white). (F) Same view as in panel E in crystal form I. G47 adopts the canonical *anti* conformation, hydrogen-bonding to the 2′-OH of G1 and N7 of G2.

linked double helices are also stabilized by a hydrogen bond formed between the 3′-OH of the nucleotide 5′ to the 2′−5′ linkage to the pro-S_P nonbridging phosphate oxygen of the 2′−5′ linked phosphodiester bond.[27] In full agreement with these studies, in all three of our crystallographically independent views of the class II product−template duplex, the C(−1) ribose adopts the C2′ *endo* conformation, and our refined models are consistent with a hydrogen bond between the 3′-OH of C(−1) and the pro-S_P phosphate oxygen of G1. Modeling the C(−1) ribose in the C3′ *endo* pucker results in greater than 4σ peaks in the residual $|F_o| - |F_c|$ Fourier synthesis and a concomitant increase in the free-*R* factor (data not shown).

Structure−Function Analysis of the Class II Ligase. If the structure of the class II ligase substrate−template duplex does not vary substantially from our crystal structures of its product−template duplex, the base-pairing interactions we observe crystallographically should explain why this motif achieves a 326-fold increase in the yield of 2′−5′ ligation products compared to a canonical Watson−Crick A-form helix. We hypothesized that this regioselectivity is governed by three key structural features of the class II substrate−template duplex that control the orientation of the reactive functional groups. First, the G1·A48 sheared pair displaces G1 into the major groove, increasing the proximity of the target of nucleophilic attack, the α-phosphate of the G1 triphosphate, to the 2′-OH nucleophile of C(−1). Second, the cross-strand stacking of A48 on the non-base-paired wedge position G2 helps position the sheared G·A pair relative to the nucleophile. This wedge effect requires that G47 not base-pair with G2. Third, the position of the two potential nucleophiles (the 2′-OH and 3′-OH of C(−1)) is constrained by formation of a Watson−Crick pair between C(−1) and G49.

To determine the importance of the features observed in our structures of the class II ligase product−template duplex, we performed-site directed mutagenesis in the full-length class II ligase "A4−11" (Figure 1B). To test the importance of the G1·A48 hydrogen-bonding network, we sequentially removed functional groups stabilizing this sheared pair and measured the catalytic activity of the resulting mutants. Mutating G1 to A removes the hydrogen bond between the exocyclic amine of G1 and the N7 of A48. This mutation inhibits the ribozyme ∼1400-fold (Figure 2A and Table S2 in the Supporting Information). Removing the hydrogen bond between the exocyclic amine of A48 and the N3 of G1 by the mutation of A48 to G results in a ∼4000-fold decrease in activity. Combining this mutation with the G1A mutation cripples the ligase ∼55 000-fold (Figure 2A). These results corroborate the functional importance of the sheared G·A pair at the ligation junction of the class II product−template duplex.

In both of our crystal forms, G2 and A3 project into the minor groove. Because their Watson−Crick faces are not involved in any base-pairing interactions, we predicted that these bases function as a "steric wedge" and that their specific identity would be nonessential. The G2 position is highly tolerant of base substitution. Mutating G2 to either adenosine or uridine results in a <5-fold decrease in activity (Figure 2A). The G2C mutation, however, reduces the activity ∼100-fold, and kinetic analysis of this mutant reveals an initial lag phase (Figure 2A and data not shown), suggesting that the mutation may interfere with proper folding of the ribozyme. These data, along with the likely stability of a C2·G47 Watson−Crick pair, suggest that the G2 position must remain unpaired to maintain activity of the ligase. The A3 position is more sensitive to mutation, with all other nucleobases at this position decreasing the activity

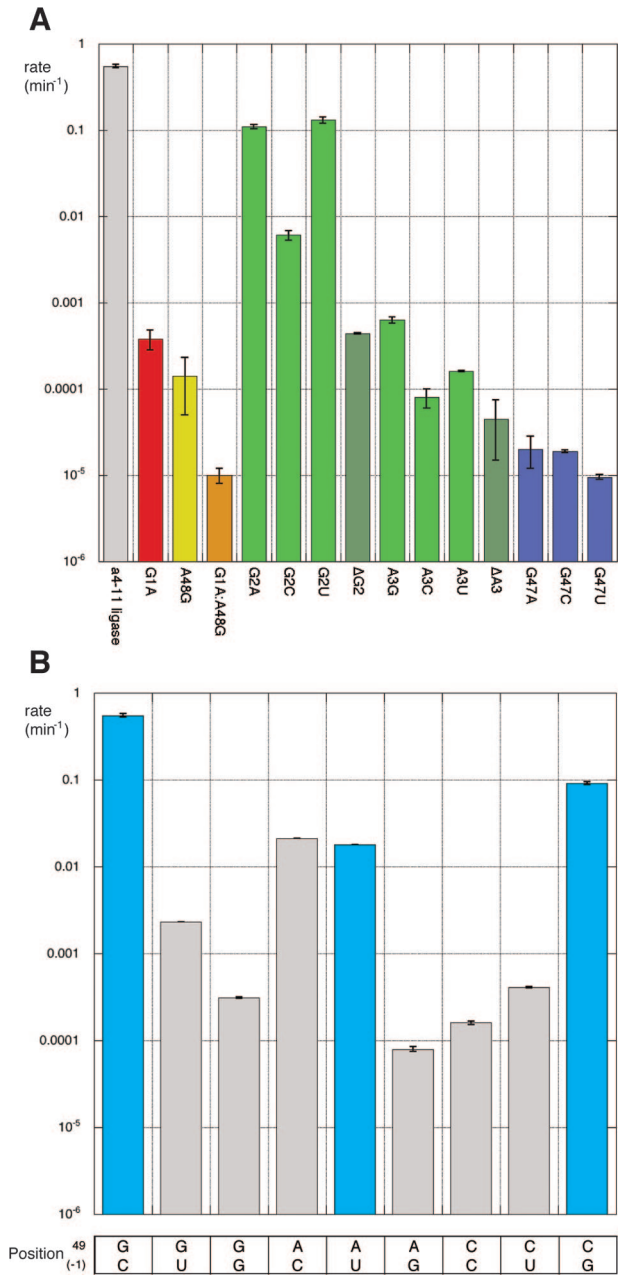

Figure 2. Structure−function analysis of class II ligase active site. (A) Observed ligation rates of class II ligase mutants. Color coding for mutant residues corresponds to that in Figure 1. (B) Rates of Watson−Crick (cyan) and non-Watson−Crick (gray) pairs at the C-1:G49 position. For rates and standard deviations, see Table S2 in the Supporting Information.

of the ligase at least 3 orders of magnitude (Figure 2A). We observe a strong peak in residual $|F_o| - |F_c|$ electron density maps in the major groove adjacent to A3 in all three of our crystallographically independent views of the product−template duplex. We have modeled this feature as a hydrated Mg^{2+} ion due to its electron density, coordination distances, and refined *B*-factor. This ion appears to make several water-mediated hydrogen bonds that stabilize the product−template duplex (Figure 1D). Loss of the ligands provided by A3 for this Mg^{2+} may be the cause for the observed decrease in catalytic rate. Collapsing the wedge by deletion of either G2 or A3 reduces the activity of the ligase by 3 and 4 orders of magnitude, respectively, supporting our hypothesis that the wedge is required to position A48 and G1. G47 is highly sensitive to mutation (Figure 2A). Our mutational analysis suggests that, of the two crystallographically observed conformations, the *syn* conformation is more relevant to catalysis, as removing the putative N7 hydrogen bond acceptor at the G2 position by mutation to U has relatively little effect on catalysis (Figure 2A). These results support our hypothesis that the A3 and G2 positions function as a steric wedge to properly position the G1•A48 sheared pair. The identity of the G2 base is nonessential, but it must not pair with G47, while an adenosine is required at position 3, probably to coordinate a Mg^{2+} ion.

The position of the two available nucleophiles (the 2′-OH and 3′-OH of C(−1)) relative to the G1 triphosphate in the precursor state is dictated by the base-pairing interaction immediately 5′ of the ligation junction (C(−1)•G49). Because class II ligases evolved to bind the oligonucleotide substrate through Watson−Crick pairing at this position and all Watson−Crick base pairs are isosteric with regard to helical geometry, we tested the ability of the ribozyme to tolerate other base pairs at this position. In all cases tested but one, canonical Watson−Crick pairs were more active than non-Watson−Crick pairs (Figure 2B). The one exception was the A49•C(−1) pair, which showed activity roughly equal to that of a canonical A49•U(−1) pair. If A49 were protonated, an A•C Watson−Crick pair may be tolerated by the ribozyme. However, because the catalytic efficiency of the ribozyme is highly pH-dependent and nearly undetectable at low pH, we were unable to discern any increase in the activity of the A49•C(−1) mutant at low pH (data not shown). These results indicate that, while the class II ligase has a slight preference for the base identity at the 49•(−1) positions, any Watson−Crick pair is tolerated. Thus, a Watson−Crick pair at this position is important for achieving the correct orientation of the nucleophile in the ligation reaction.

Comparison with the L1 Ligase. Recently, the structure of an independently isolated[3] 3′−5′ RNA ligase ribozyme, the L1 ligase (Figure 3A), was reported.[15] That structure of the ribozyme−product complex revealed a flexible, three-stemmed RNA that is held in an active conformation by surprisingly few tertiary interactions. The product−template stem A and the A-form geometry stem B flank a three-way junction that forms a hinge. This hinge allows the catalytic stem C to dock at the ligation site through a single base triple interaction (U38:G1:A51) and a Mg^{2+}-mediated interaction between nonbridging phosphate oxygens in stems A and C (G40:A39:G1).

Comparison of the ligation junctions of the L1 and the class II product−template duplexes shows that the formerly triphosphate-activated guanosine residue in the active sites of both ribozymes participates in a sheared G•A pair. In the L1 ligase, the Watson−Crick face of the adenosine making this interaction (A51) projects into the minor groove, forming a reverse Watson−Crick pair that is part of the base triple that brings stem C to the active site. Superposition of the L1 and class II product−template duplex structures (Figure 3B) suggests that the regioselectivity of both the L1 and class II ligases may be controlled by positioning the two available nucleophiles [the 2′-OH and 3′-OH of the nucleotide at position (−1)] by the geometry of the ligation site. As discussed above, in the class II ligase, the potential nucleophilic OH groups of C(−1) are constrained by Watson−Crick pairing of C(−1) with G49. If the nucleotide carrying the target of nucleophilic attack, G1, were also to form a Watson−Crick pair, helical geometry would favor attack by the 3′-OH, as observed in ligase ribozymes with continuous Watson−Crick pairing across the ligation junc-

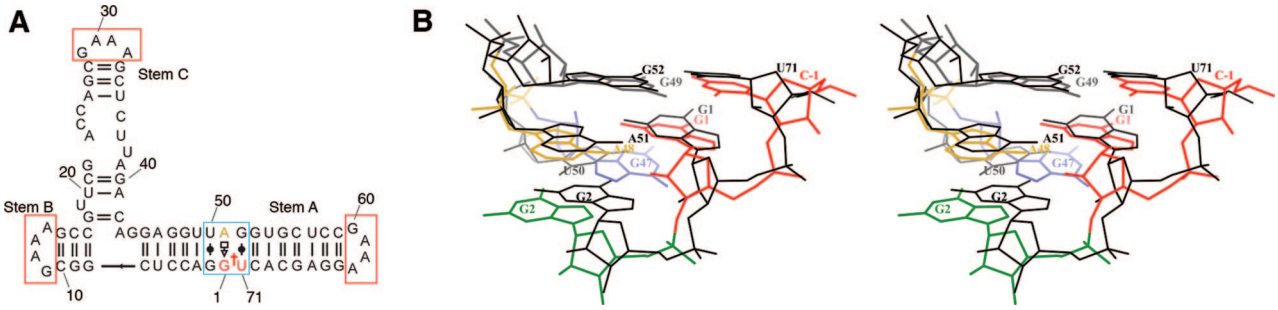

Figure 3. Comparison of the L1 and class II ligase ligation junctions. (A) Secondary structure of the L1 ligase L1x6c.[15] Red dagger denotes site of ligation. Red boxes are engineered GNRA tetraloops not present in the original L1 ligase.[3] Blue box denotes region shown in panel B. (B) Stereoview of a portion of the least-squares superposition (using phosphorus and O4' atoms) of product−template helices of the class II ligase (colored model, from crystal form II) and the docked conformation of the L1 ligase−product complex (in black, chain Q, PDB ID 2OIU).

tion.[4,28,29] However, when constrained by a G•A sheared pair, G1 is displaced into the major groove, which would favor attack by the 2′-OH due to its proximity.

The L1 ligase, which is highly regioselective for 3′−5′ bond formation,[3] contains a G•A pair at its ligation junction. However, in this ligase, the sheared pair is flanked by U•G and G•U wobble pairs (Figure 3A). When introduced to an A-form helix, wobble pairs impart a characteristic undertwist−overtwist to the backbone, altering helical geometry.[30] The two wobble pairs in the L1 ligase appear to favor attack by the 3′-OH by a combination of two factors. First, wobble pairing displaces U71 (and its 2′-OH) into the major groove, compensating for the displacement of G1. Second, the G2 wobble alters the backbone geometry, affecting the orientation of the G1 sugar and, therefore, the position of the α-phosphate in the precursor. The geometry of G•U wobble pairs also exposes a hydrophilic binding pocket in the minor groove[31] that allows formation of a water-mediated hydrogen bond between the 2′-OH of U and the exocyclic amine of G. Robertson and Scott[15] hypothesized that this water adjacent to the G52-U71 wobble pair, which in the L1 ligase is also coordinated by the 2′-OH of U38, quenches the alkoxide produced by activation of the 2′-OH of U71, leading to the strict regioselectivity for the 3′-OH of the ribozyme.

Structure-Based Engineering of Ribozyme Regioselectivity. Because the sheared G•A pairs at the ligation junctions of two ribozymes with opposite regioselectivity superimpose (Figure 3B), and the product−template duplex of the class II ligase is sufficient to specify much of the regioselectivity of the class II ligase,[7] we hypothesized that the residues flanking this G•A pair control the opposed regioselectivity of the class II and L1 ligases. In the class II ligase, where the nucleophilic residue participates in a G•C Watson−Crick pair, the 2′ oxygen is 1.6 Å from the G1 phosphate, compared with 3.6 Å for the 3′ oxygen. In the L1 ligase, where the nucleophilic residue participates in a U•G wobble, this relationship is reversed: the potential 2′ and 3′ nucleophiles are 4.0 Å and 1.6 Å from the G1 phosphorus, respectively. Since the position of these OH groups is governed by base-pair geometry, changing the base-pairing of the residue carrying the two potential nucleophiles should predictably alter the regioselectivity of the ligation reaction. The orientation of the G•A sheared pair relative to the two potential nucleophiles should also be critical for determining the position of the target of nucleophilic attack, the α-phosphate of G1, and therefore the regioselectivity of the reaction. In the class II ligase, G2 does not form any base-pairing interactions across the helix but instead functions as a steric wedge, stacking underneath A48, widening the minor groove, and affecting the orientation of G1 (Figure 1C). The resulting orientation of G1 allows for a Hoogsteen•sugar edge pair with A48 in which the G1 ribose adopts the C3′ *endo* conformation and three hydrogen bonds are formed (Figure 1C). In the L1 ligase, the G1 ribose adopts the C2′ *endo* conformation, most likely to avoid a steric clash between its 2′-OH and the O2 of U38. The critical tertiary interaction between A51 and U38 likely limits the acceptable base geometry surrounding the G1 ligation junction in the L1 ligase, as base-pairing interactions which cause distortions in backbone geometry will prevent the L1 from folding into an active conformation.[15] Therefore, the U38:G1:A51 tertiary interaction, along with the G•U and U•G wobble pairs, are likely critical for specifying the regioselectivity of the L1 ligase.

With the exception of the pucker of the G1 sugar in the L1 ligase product−ribozyme complex, all of the interactions we describe appear to be dictated by local geometry inherent to the product−template duplexes. Therefore, transplanting residues flanking the conserved G•A pair into their noncognate ligases should produce ribozymes with altered regioselectivity. To test this prediction, we engineered versions of the full-length class II ligase where we first removed those elements we predicted to favor formation of the 2′−5′ linked regioisomer and then inserted those elements from the L1 ligase predicted to favor formation of the 3′−5′ linked regioisomer. Collapsing the steric wedge formed by G2 and A3 in the class II motif, by the deletion of G2 (Figure 4C), significantly reduced the 2′−5′ regioselectivity of the resulting mutant class II ligase, resulting in a product ratio of 2:1 (2′−5′:3′−5′), compared to a product ratio of 77:1 for the wild-type A4−11 ligase[7] (Figure 4C,D). Consistent with our hypothesis that G•U wobble pairing at the position carrying the nucleophiles favors production of the 3′−5′ linked regioisomer, the regioselectivity of a class II mutant where C(−1) is changed to U (Figure 4E) leads to a further reduction in the yield of 2′−5′ linked product, with product ratio of 1:3 (Figure 4E,F). Interestingly this effect requires the deletion of G2. In the presence of the G2 wedge, the C-1 to U mutation does not show a significant change in regioselectivity (Table S2). In our structures of the class II ligase product−template duplex, the A3 position also functions as a wedge, forming no base pairs across the helix. When this is altered by

(28) Ekland, E. H.; Bartel, D. P. *Nucleic Acids Res.* **1995**, *23*, 3231–3238.
(29) Coppins, R. L.; Silverman, S. K. *J. Am. Chem. Soc.* **2004**, *126*, 16426–16432.
(30) Varani, G.; McClain, W. H. *EMBO Rep.* **2000**, *1*, 18–23.
(31) Masquida, B.; Westhof, E. *RNA* **2000**, *6*, 9–15.

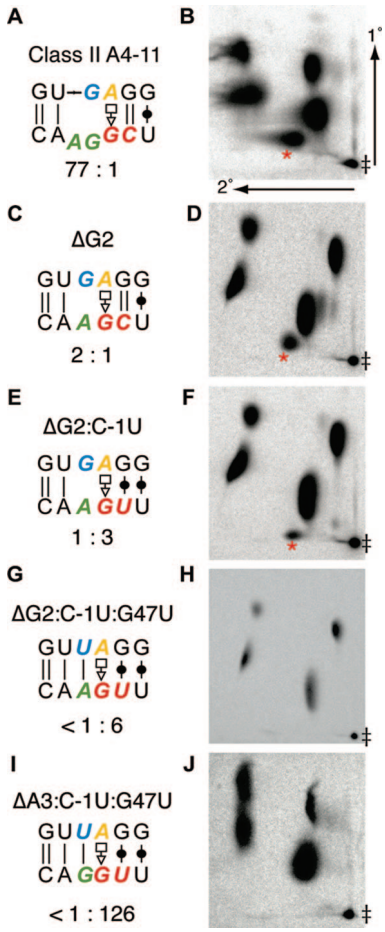

Figure 4. Engineered class II ligases with altered regioselectivity. (A) Predicted secondary structure of the wild-type substrate−template duplex in the class II ligase A4−11. Engineered residues are colored in bold. Mean product ratios of the two possible regioisomers (2′−5′:3′−5′) were determined by quantitation of triplicate 2D TLC (see Experimental Section). (B) RNase T2 digestion of purified, body-labeled ligation reaction products separated by 2D TLC. Arrows denote first and second dimensions. Origin is marked by a double dagger. Red asterisk denotes location of 2′−5′ linked dinucleotide. Panels C,D−I,J are as in panels A,B for each engineered ligase.

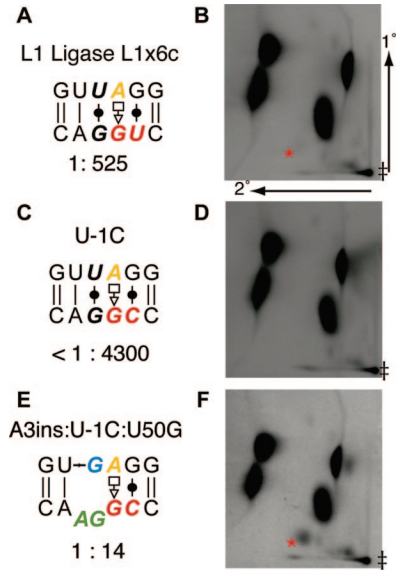

Figure 5. Engineered L1 ligase with altered regioselectivity. (A) Predicted secondary structures of the wild-type substrate−template duplexes in the L1 ligase L1x6c.[15] Mean product ratios of the two possible regioisomers (2′−5′:3′−5′) were determined by quantitation of triplicate 2D TLC (see Experimental Section). (B) RNase T2 digestion of purified body-labeled ligation reaction products separated by 2D TLC. Arrows denote first and second dimensions. Origin is marked by a double dagger. Red asterisk denotes location of 2′−5′ linked dinucleotide. Panels C,D and E,F are as in panels A,B for each engineered ligase.

Table 1. Ligation Rates of Engineered Ligase Ribozymes and Isolated Substrate−Template Duplexes[a]

mutation	rate (min^{-1})	σ (min^{-1})	R
Class II Ligase			
A4−11 class II ligase (wildtype)	5.5×10^{-1}		0.991
ΔG2	4.4×10^{-4}	1×10^{-5}	
ΔG2:C-1U	6.0×10^{-6}	1×10^{-7}	
ΔG2:C-1U:G47U	3.3×10^{-7}	2×10^{-8}	
ΔA3:C-1U:G47U	3.8×10^{-6}		0.999
L1 Ligase			
L1x6c L1 ligase (wild-type)	7.3×10^{-2}		0.994
U-1C	3.2×10^{-4}	1×10^{-5}	
A3 insertion:U-1C:U50G	7.4×10^{-5}		0.993
Isolated Substrate−Template Duplexes			
L1 ligase stem A	1.2×10^{-6}	2×10^{-7}	
L1 stem A:U-1C:U50G:A3 insertion	7.9×10^{-7}	8×10^{-8}	
class II ligase motif	7.5×10^{-7}	2×10^{-8}	
class II ligase motif:C-1U:G47U:ΔA3	4.8×10^{-7}	7×10^{-8}	

[a] Measured rates (min^{-1}) with standard deviations for single time-point experiments or R values for curve-fitting of multiple time point experiments. Isolated substrate−template duplexes are the ribozymes with their catalytic domains replaced by a tetraloop. Rates for substrates bound in a fully Watson−Crick duplex are 2.9×10^{-8} and 4×10^{-10} min^{-1} for 3′−5′ and 2′−5′ ligation, respectively.[7,46]

mutating G47 to U, which would allow A3 to form a canonical A-U Watson−Crick pair in the G2 deletion mutant (Figure 4G), we fail to detect any 2′−5′ linked product (Figure 4G,H), indicating that the regioselectivity of this triple-mutant ribozyme has been completely switched. When the substrate−template duplex of the class II ligase is fully converted to that present in the L1 by two base changes (G47U, C(−1)U) and the deletion of A3 (Figure 4I), we fail to detect any 2′−5′ linked product (Figure 4I,J), demonstrating that the resulting chimeric class II ligase has also completely switched its regioselectivity.

To test the generality of these results, we performed reciprocal engineering experiments on the L1 ligase, attempting to increase the yield of 2′−5′ linked products, by incorporating elements observed in our structures of the class II product−template duplex (Figure 5). A chimeric L1 ligase that contains the substrate−template duplex of the class II ligase (Figure 5E,F) yields a significant increase in 2′−5′ products, with a product ratio of 1:14 (2′−5′:3′−5′) compared with 1:525 for the wild-type L1 ligase (Figure 5A,B). Interestingly, replacing the G·U wobble pair that carries the two potential nucleophiles (G52·U71) with the G·C Watson−Crick pair found in the class II ligase (Figure 5C) is not sufficient to increase the yield of the 2′−5′ linked regioisomer (Figure 5C,D). This indicates that geometric effects, influencing the positions of the two potential nucleophiles and the electrophile, control the regioselectivity of the L1 and argues against the model proposed by Robertson and Scott, where the water coordinated by the potential 2-oxygen nucleophile of U71 inhibits 2′−5′ bond formation.[15]

Both fully engineered ligases (Figures 4I and 5E) are significantly less active than the original ribozymes (Table 1). Interestingly, we find that, like the class II minimal motif, the L1 ligation junction motif (5′-UGG-3′ 5′-GAG-3′) alone, absent

the catalytic core of the ribozyme, is able to accelerate the rate of the ligation reaction (Table 1). In fact, this rate enhancement of the minimal L1 motif is responsible for a substantial fraction of the catalytic activity seen in our fully engineered class II ribozyme (Table 1). This differs from our fully engineered L1 chimera, which shows at least 100-fold rate enhancement over that displayed by the class II substrate−template duplex alone. This discrepancy suggests that the overall three-dimensional architecture of the full-length L1 ribozyme is more flexible than that of the class II ribozyme, allowing it to reposition its catalytic stem C in a productive orientation with respect to the engineered substrate−template duplex. The flexibility of the overall structure of the L1 ligase is readily apparent in the structures reported by Robertson and Scott.[15] Of their two crystallographically independent ligase product structures, one shows the catalytic stem C making tertiary contacts, bringing it to the site of ligation, while the other has the stem C swung away from the active site.

Convergent in Vitro Evolution? Our analysis identifies a G·A sheared base-pairing interaction conserved between the ligation junctions of two independently evolved ribozymes with contrasting regioselectivity. The presence of the G is the result of limitations of the *in vitro* transcription system used for these experiments.[32] However, the presence of a sheared Hoogsteen· sugar-edge pair with an adenosine in both ribozymes points to chemical determinism or convergent evolution. During the *in vitro* selection experiments that gave rise to both the class II ligase and the L1 ligase ribozymes, the substrate binding site designed by the investigators was not used by the ribozymes that were most successful in the selection. Ribozymes that used substrate binding sequences inside the variable region of the selection library out-competed those that maintained the designed Watson−Crick substrate binding sites. This is a frequent observation in these types of selection experiments,[3,7,33] and while the L1 and class II are the only representatives of this group for which we have detailed structural information, we see two likely reasons that this sheared G·A pairing or other noncanonical pairs found at the ligation junctions of similar ribozymes may be selected. First, because the reactive groups are prealigned by engineered substrate binding stems, the key difficulty RNAs have at becoming selected may be proper placement of catalytic metal ions or other groups required to activate the nucleophile and achieve substantial rate enhancement. If catalytic metal ion binding is sterically incompatible with the designed substrate binding sites, other substrate binding sites in the ribozyme that afford access to the metal ions may be selected. Since these binding sites will have arisen *de novo*, and the probability of forming a perfectly complementary template is low, this alone could explain the repeated selection of non-Watson−Crick paired substrate−template duplexes like the class II motif. Second, these substrate binding modes may afford the ribozyme a catalytic advantage relative to a Watson−Crick duplex. Our data and previous mutational analyses of these ribozymes appear to support that conclusion (Figure 2, Table 1, and ref 34). Indeed, the observation that the isolated substrate−template duplexes of the class II and L1 ligases both afford some rate enhancement (ref 7 and Table 1) suggests that helical geometries other than A-form may be better substrates for ligation.

While there is some evidence that polymerases can use base pairing other than Watson−Crick for accurate replication,[35,36] it is difficult to imagine how a polymerase ribozyme that utilizes the base-pairing schemes found in the substrate−template duplexes of the class II and L1 ligases could function as an RNA replicase capable of accurate informational transfer, a necessary requirement of a replicase ribozyme.[37] Although it is possible to interpret the results of random selection as implying that these motifs are plausible inhabitants of the RNA world and therefore likely to resemble the first information-copying molecules, we favor the interpretation that these base-pairing motifs were selected due to their flexibility, i.e., the unpaired wedge in the class II ligase and the "conformationally soft" G·U wobbles[30] in the L1 ligase. We suggest that the highly stringent conditions of the *in vitro* selection that produced these ribozymes favored substrate−template duplexes that project favorable functional groups into the minor groove, allowing the formation of specific tertiary interactions at the site of ligation, at the cost of preserving informational transfer, simply because the selection protocols strongly weighed catalysis over accurate "replication".

Substrate−Template Duplex Control of Regioselectivity. Control of the regioselectivity of these ligase ribozymes by their substrate−template duplexes is not entirely unexpected. Work on the template-directed abiotic polymerization of activated ribonucleotides showed strong effects on regioselectivity of the substrate−template sequence, the identity of metal ions present in the reaction, and the nature of the leaving group (reviewed in refs 37and 38). However, when the reactive groups were constrained into A-form geometry by the use of a triphosphate-activated oligonucleotide on a complementary template,[39] the ligation reaction strongly favored the 3′−5′ linked regioisomer. Efforts to constrain *in vitro* selection experiments have exploited this observation and have been successful in isolating catalytic nucleic acids that maintain the continuous Watson−Crick pairing across the ligation junction and the desired 3′−5′ regioselectivity.[4,29] Our results inform these experiments by highlighting how ribozymes that deviate from canonical A-form geometry and Watson−Crick pairing control the regioselectivity of the polymerase-like ligation reaction.

Our results demonstrate how local geometric constraints, specified by both the base-pairing interaction of the nucleophilic residue and the backbone geometry resulting from base-pairing downstream of the electrophile, influence the regioselectivity of the ligation reaction. The effect of these geometric factors is, to a considerable extent, independent of the peripheral catalytic "domains" of the ribozymes, and in both the class II and L1 ligases rational modification of the substrate−template duplexes predictably alters the product ratios of the engineered ribozymes. While in terms of absolute rate enhancement these engineered ribozymes are inefficient catalysts, the ability rationally to engineer these ribozymes demonstrates our grasp

(32) Kennedy, W. P.; Momand, J. R.; Yin, Y. W. *J. Mol. Biol.* **2007**, *370*, 256–268.
(33) Hager, A. J.; Szostak, J. W. *Chem. Biol.* **1997**, *4*, 607–617.
(34) Robertson, M. P.; Hesselberth, J. R.; Ellington, A. D. *RNA* **2001**, *7*, 513–523.
(35) Nair, D. T.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K. *Structure* **2005**, *13*, 1569–1577.
(36) Nair, D. T.; Johnson, R. E.; Prakash, S.; Prakash, L.; Aggarwal, A. K. *Nature* **2004**, *430*, 377–380.
(37) Orgel, L. E. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 9–16.
(38) Joyce, G. F. *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 41–51.
(39) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 3340–3344.

of the underlying structural basis for the regioselectivity that arises from the substrate−template duplexes.

Both the class II and L1 substrate−template duplexes are themselves catalysts. Nonetheless, for both the class II and L1 ligases, the rates of the engineered full-length ribozymes exceed those of ribozymes comprised exclusively of the substrate−template duplexes capped by a tetraloop (Table 1). The peripheral catalytic elements or domains confer an additional 8- and 94-fold increase in rate to the engineered full-length class II and L1 ligases, respectively. In the class II ligase, these peripheral elements do not affect the regioselectivity of the reaction, as we are unable to detect any of the 2′−5′ regioisomer in the products of the engineered full-length ribozyme (Figure 4I,J). In the engineered L1 ligase, 6.6% of the reaction products are the desired 2′−5′ linked regioisomer (Figure 5E,F). If the rate of 2′−5′ ligation is extrapolated from the observed rate of formation of both regioisomers (6.6% of 7.4×10^{-5} min^{-1}), the resulting rate is 6 times faster than our observed rate for the engineered L1 ligase substrate−template duplex (closed with a tetraloop; Table 1). This indicates that, while most of the ligase product remains 3′−5′ linked, stems B and C do increase the rate of 2′−5′ bond formation compared to the engineered substrate−template duplex alone.

In order to achieve substantial rate enhancement, a ligase must activate the nucleophilic hydroxyl to form the alkoxide. Protein-based polymerases use divalent metal ions for this purpose,[40] but our crystal structures do not suggest if or how the class II ligase substrate−template duplex activates the 2′-OH (see discussion in Supporting Information). We favor the interpretation that the effect of the substrate−template duplex is almost exclusively due to geometric factors, aligning the 2′-OH and the α-phosphate for attack. The peripheral catalytic sequences of the class II ligase have been shown to contribute to regioselectivity on its cognate substrate−template duplex.[7] Because these sequences provide most of the rate enhancement of the ribozyme but contribute less of the regioselectivity than the substrate−template duplex,[7] we speculate that these peripheral sequences activate the 2′-OH nucleophile. The increased fraction of 2′−5′ product resulting from the full-length class II ribozyme compared to what is produced by its substrate−template duplex alone would then result from preferential activation of the 2′-OH. The ability of the peripheral catalytic domain of the class II to activate the 3′-OH on its noncognate substrate−template duplex, while limited, demonstrates that it may be possible to separate regioselectivity from overall rate enhancement. As such, we suggest that the peripheral sequences of our engineered ligases are not optimized for their new substrate−template duplexes. The inefficiency of our engineered ligases could be due to improper positioning of the nucleophiles themseleves, but this seems less likely because the engineered class II ligase can overcome the 6- to 9-fold higher intrinsic reactivity of the 2′-OH relative to the 3′-OH[41] and afford a further 8-fold increase in the rate of 3′−5′ bond formation over the rate of the engineered class II substrate−template duplex alone.

Conclusion

Structure-guided engineering of protein enzymes is often complicated by their plasticity, such that local perturbations propagate with unexpected consequences.[42] The comparative stability of isolated RNA structural elements may facilitate the structure-guided engineering of the catalytic properties of ribozymes. Unlike protein folding, in which hydrophobic collapse proceeds through a molten globule to arrive at a fully folded protein domain, RNA folding is governed by a hierarchy of strong local interactions, which sequentially achieve a final tertiary structure.[43,44] This modularity of RNA may facilitate the design or evolution of new ribozymes. We were, to varying degrees, able to engineer the regioselectivity of the class II and L1 ligases on the basis of structural analysis of the product−template duplexes and transplantation of key elements we predicted to control regioselectivity. While these experiments support the functional interpretation of our structures, our engineered ligase ribozymes would require more substantial rate enhancement to be of real use. Because our results establish that the local geometry of the substrate−template duplexes controls ligase regioselectivity, it is likely that further *in vitro* evolution, in which the sequences of the substrate−template duplexes are kept fixed while the peripheral catalytic sequences are mutated, would allow for rapid isolation of more active ligases with the desired regioselectivity.

Acknowledgment. We thank the staff at ALS beamlines 5.0.1 and 5.0.2 and J. Bolduc for assistance with synchrotron and home laboratory X-ray data collection, respectively, and T. Edwards, K. Fisher, T. Hamma, R. Kapur, B. Kurland, D. Klein, A. Luptak, J. Posakony, A. Roll-Mecak, M. Roth, J. Simon, B. Stoddard, A. Weiner, and H. Zaher for discussions. This work was supported in part by a grant from the W.M. Keck Foundation. J.N.P. was supported by the NASA Graduate Student Researchers Program (NGT5-50448) and by the Chromosome Metabolism and Cancer training grant (NIH T32 CA09657). A.R.F. is a Distinguished Young Scholar in Medical Research of the W.M. Keck Foundation and an Investigator of the Howard Hughes Medical Institute.

Supporting Information Available: Supporting discussion; crystallographic data collection, phasing and refinement statistics (Table S1); rates and regioselectivities of engineered ribozymes (Table S2); experimental electron density map (Figure S1); and anomalous difference Fourier map (Figure S2). This material is available free of charge via the Internet at http://pubs.acs.org.

JA8067325

(40) Steitz, T. A. *Nature* **1998**, *391*, 231–232.
(41) Lohrmann, R.; Orgel, L. E. *Tetrahedron* **1978**, *34*, 853–855.
(42) Brannigan, J. A.; Wilkinson, A. J. *Nat. Rev. Mol. Cell. Biol.* **2002**, *3*, 964–970.
(43) Sosnick, T. R. *Protein Sci.* **2008**, *17*, 1308–1318.
(44) Tinoco, I., Jr.; Bustamante, C. *J. Mol. Biol.* **1999**, *293*, 271–281.
(45) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497–3531.
(46) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 3332–3339.