The Structural Basis of RNA-Catalyzed RNA Polymerization

by

David M. Shechner

B.S., M.S. Molecular Biophysics and Biochemistry
Yale University, 2000

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2010

Signature of Author: _____
Department of Biology
February 11, 2010

Certified by: _____
David P. Bartel
Professor of Biology
Thesis Supervisor

Accepted by: _____
Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Committee

The Structural Basis of RNA-Catalyzed RNA Polymerization

by

David M. Shechner

Submitted to the Department of Biology
on February 11, 2010 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy
in Biology

ABSTRACT

The Class I ligase is an artificial ribozyme that catalyzes a reaction chemically identical to a single turnover of RNA-dependent RNA polymerization. Such an activity would have been requisite for the emergence of a self-replicase ribozyme, an enzyme that, according to the RNA World hypothesis, would be fundamental for the emergence of life. Demonstrating the plausibility of RNA-catalyzed self-replication, the Class I ligase catalytic machinery was previously harnessed to produce general RNA polymerase ribozymes. Hence, this ligase represents a robust model system for studying both the potential role RNA may have played in the origins of life and RNA catalysis in general. Through a combination of crystallographic and biochemical experiments, we have sought to elucidate the structure and mechanism of this ribozyme.

As a starting point for our experiments, the crystal structure of the self-ligated product was solved to 3.0 Angstrom resolution, revealing a tripodal architecture in which three helical domains converge in the vicinity of the ligation junction. A handful of tertiary interactions decorate this tripod scaffold; among them were two instances of a novel motif, the A-minor triad. The structure elucidated interactions that recognize and bind the primer-template duplex and those that position the reaction electrophile. It furthermore revealed functional groups that compose the active site. Biochemical evidence and the position of these groups lead us to propose a reaction mechanism similar to that used by proteinaceous polymerases.

Using a slowly reacting mutant, 3.05–3.15 Angstrom crystal structures were solved of unreacted, kinetically trapped ligase-substrate complexes bound to different metal ions. Comparison of the Ca^{2+}- and Mg^{2+}-bound structures explains the preference of the ligase for Mg^{2+}. Moreover, these structures revealed features missing in the product structure: interactions to the 5′-triphosphate and an active site catalytic metal ion. While this metal is positioned in a manner similar to the canonical “Metal A” of proteinaceous polymerases, the role of “Metal B” might have been supplanted by functional groups on the RNA. Kinetic isotope experiments and atomic mutagenesis of two active site functional groups imply that they may act in concert to electrostatically aid transition-state stabilization.

Thesis Supervisor: David P. Bartel
Title: Professor of Biology

Dedication

This work is dedicated to Mildred Winston, and in the loving memory of Fred Winston (1916–2007) and Isabel Shechner (1906–2005), all of whom taught me to appreciate and explore the world's diverse wonders. These are people of such unassuming warmth that, at first glance, one might almost overlook how thoroughly extraordinary they are. *Almost*.

Acknowledgements

My years at MIT were marked by profound challenges and equally profound joys, throughout all of which I was sustained by a loving family and a truly extraordinary group of friends and colleagues. I am eternally grateful for all that these people have given to me, and know that I would hardly be the scientist–nor the person–that I am today without their influence.

I would like to thank Dave Bartel for allowing me to be a part of his laboratory, and for the support and guidance he's provided during my time here. Given his notorious scientific prowess and expert mentorship, Dave in many ways serves as the model I strive to emulate in my own career.

I am also deeply grateful to the members of my thesis committee–Tom RajBhandary, Phil Sharp, Cathy Drennan, Thomas Schwartz and Jack Szostak–for their advice, support and enthusiasm, and for being a wonderful group of people to know. I am particularly indebted to Cathy and Jack, who for years have welcomed me into their labs and into their group meetings. The experiences I've had working with the outstanding people in their labs were vital to my development as a scientist, and were quite a lot of fun.

I am extremely lucky to have benefited from the outstanding mentorship of Bob Grant. Bob has been a wonderful teacher and a great friend, who, throughout the most difficult times in my research, was a constant source of invaluable advice and humor. Working alongside him on our numerous synchotron trips was a truly unique experience, one I shall treasure always.

My sincerest thanks to my collaborators, Sarah Bagby, Yelena Koldobskaya and Joe Piccirilli, and to Adrian Ferré-D'Amaré, Kaihong Zhou, Jennifer Doudna, Rob Batey and Fritz Eckstein for providing reagents and protocols. I am furthermore deeply grateful to Nick Yoder and Eric Spooner, who assisted with HPLC and mass spectometry experiments, respectively. Thanks to Vladimir Malashkevich, Frank Solomon, Bob Sauer, K. Rajashankar, Mike Robertson, Bill Scott, Mary Stahley, Jesse Cochrane and Scott Strobel for their invaluable advice and support.

It has been a rare honor to work in the Bartel lab and in the greater Whitehead community. While one might expect that institutions of this caliber would attract truly World-Class scientists, one could hardly imagine that they would also be such warm, welcoming and fun people. I am especially indebted to Erik Schultes, Margy Glasner and Nick Bergman, who got me started in the lab and who taught me the most profound truth I have learned in my adult life: avoid RNase at all costs. I'd also like to thank Chuck Merryman, Scott Baskerville, David Lee, Catherine Yen, Ed Curtis, Mike Lawrence, Uli Müller, and Wendy Johnston, with whom I shared many wonderful years studying ribozymes. I would particularly like to acknowledge Sarah Bagby, who endured working with me in both a scientific collaboration and a teaching assistantship. Sarah is a truly extraordinary person whom I'm honored to call my friend.

Moreover, the non-Ribozyme members of the Bartel lab have always proven a welcoming and open-minded group, and to all of them I am quite thankful. It has been my honor to know Graham Ruby, a treasured friend and a scientist of peerless quality. I'd like to thank Olivia Rissland for helping edit my thesis and for being a cherished friend and co-conspirator. I'd also

like to thank (in no particular order) Laura Resteghini, Calvin Jan, Soraya Yekta, Andrew Grimson, David Garcia, Anna Drinnenberg, Noah Spies, Robin Friedman, David Weinberg, Alena Shkumatava and Rosaria Chiang for making the Bartel lab a great place in which to work. And, during those (rare) times when it wasn't, I'd like to thank all the members of the Sabatini lab for always welcoming me as an adopted member.

I'd thank the Max Fünk Institut–Chris "C-Funk" Petersen, Graham "Jimmy-G" Ruby, Phil "High IQ" Iaquinta, "Jacketless" John Albeck, Bob "Dokta Koktagon" Kokta and Lee "Nekkid" Neckritz–for being the funkiest group of academics the world has ever known. Thank you for making me feel like a rock star, even off the stage.

I'd like to thank all of those involved in preparing the now infamous Whitehead retreat skit. It was a rare pleasure to write with you, perform with you, and in certain cases, smash you in the face with a cream pie.

Also, I'd thank all of my friends from outside the lab, without whose support and levity I would have surely lost my will to continue: Jeremy and Naomi Taylor, Matt Belinkie, Josh McNeil, Dave Caputo, Eric Frieden, Mike Stafford, Matt Wrather, Peter Fenzel, Emily Berry, Rebecca Merz, Jordan Stokes, Ben Drapkin, Ryan Sheely, Mark Lee, Nathan and Beth Willard, Adam Meninga, Patrick Hughes and Jonah Wallerstein. I thank you for your friendship, and for letting me endlessly pontificate about science.

Nothing I have experienced in this world has been more fascinating to me, or more beautiful, than Yasemin Sancak. I'll probably never understand why such a remarkable, wonderful person would wish to be my wife, but I intend to enjoy my good fortune for the rest of my days. It is with great love I also thank Naciye Ak, Yusuf Sancak and the rest of Yasemin's family for producing her, and for their love and patience they have shown in welcoming me into their fold. I'll add Lena Khibnik to that group, too, since she is more to us like a family member than merely a friend.

Finally, it is with the deepest love and eternal devotion that I thank my family–particularly my parents, Janet and Norman Shechner, my grandparents, Mildred and Fred Winston, my sister, Jessica Carter, my great aunt Isabel Shechner, my aunt Caron and uncle Jeff Winston, and my cousins Rachel McMinn and Robert Winston–for their endless support, love and dedication. I am truly lucky to be a member of this incredible family, and strive to make them as proud of me as I am of them. Without their influence, I would hardly be the person I am today.

…Which is to say, it's *their* fault.

Table of Contents

Chapter II has been published previously as:
D. M. Shechner, R. A. Grant, S. C. Bagby, Y. Koldobskaya, J.A. Piccirilli and D.P. Bartel, "Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme." Science 326(5957), 1271-1275. © 2009, AAAS.

The Appendix has been published previously as:
S. C. Bagby, N. H. Bergman, D. M. Shechner, C. Yen and D. P. Bartel, "A Class I Ligase Ribozyme with Reduced Mg^{2+} Dependence: Selection, Sequence Analysis and Identification of Functional Tertiary Interactions." RNA 15(12), 2129-2146. © 2009, RNA Society

Introduction

General Implications of RNA Catalysis

The discovery of catalytic RNA (1-4) represents a fundamental turning point in molecular biology. The ability of a single biopolymer to function as both a storehouse for genetic information and a catalyst evokes a concise solution to an intrinsic paradox underlying the origin of life (5-10). Since, in all extant organisms, DNA and proteins are almost exclusively assigned the roles of information storage and catalysis, respectively, it is hard to conceive of how either molecule might function in the absence of the other. Indeed, even in its simplest forms, the process of creating new copies of a given polypeptide is remarkably complex, requiring RNA polymerase, tRNAs, aminoacyl-tRNA synthetases, the ribosome and so forth. However, a system in which catalysts serve directly as the templates for their own replication would require only a single activity for both the transmission and the use of genetic information (11). Such a system would hence be easier to develop from prebiotic materials (9, 10).

It has been proposed (5-7) that in primordial organisms RNA might have served as the sole biopolymer used for both information storage and catalysis, a theory later termed the "RNA World hypothesis (8)." This theory is supported by the observation that RNA, its precursors or its derivatives are vital components in many of the most highly conserved–and hence ancient–processes in biology (12, 13). Remnants of ancient RNA-based chemistries can be found in so-called molecular fossils (14): the universal energy source ATP is an activated nucleotide, as are the essential enzyme cofactors NAD, FAD and Coenzyme A. Deoxyribonucleotides are not synthesized *de novo*, but rather from ribonucleotide precursors, as are the essential cofactor folate and the amino acid histidine. Furthermore, in addition to their notoriously important roles in DNA replication and protein synthesis, longer RNA species are of paramount importance

myriad biological processes, including but not limited to mRNA splicing, the maturation and modification of tRNA and rRNA, and the translocation of transmembrane proteins (12). Finally, the Ribosome, the enzyme responsible for all genetically-encoded polypeptide synthesis and perhaps the oldest surviving enzyme in all modern cells, is unequivocally a ribozyme (15, 16). Taken together, these data imply that biochemical reactions built from RNA likely predate those involving DNA or protein.

Although the existence of catalytic RNA thus provides an elegant solution to a longstanding evolutionary conundrum, it concomitantly raises basic biochemical questions. Compared to proteins, the RNA biopolymer has at its command a relative dearth of functional groups, accessible pK_as and unique steric profiles for use in catalysis (17). Given this, one might question both the range of chemical transformations that RNAs would be able to accelerate and the scope of catalytic strategies they could use in accelerating these transformations. This first issue is sparsely addressed by the known natural ribozymes, the chemical repertoire of which seems essentially limited to the transfer or hydrolysis of phosphodiester bonds and the formation of peptide bonds. Hence, the full potential catalytic repertoire of RNA has largely been revealed through *in vitro* selection, which has yielded a multitude of artificial RNAs with novel functions (18-20). Conversely, the myriad catalytic strategies now known to be the purview of RNA have largely been elucidated through intense biochemical scrutiny of the natural ribozymes (21). Importantly, much of this work has been aided by developments in RNA crystallography, a field that has dramatically flourished in recent years.

The Structural Enzymology of Natural Ribozymes

Currently, there are nine known classes of naturally occurring catalytic RNAs. Five classes are small, self-cleaving ribozymes: the Hammerhead (22, 23), Hairpin (24), Hepatitis Delta Virus (HDV) (25), Varkud Satellite (VS) (26) and GlmS (27). Two classes are self-splicing introns: the Group I (1, 4) and Group II (28). The remaining two classes are large ribozymes that process their substrates *in trans* and naturally function in obligate complexes with proteins: Ribonuclease P (2, 3), and the Ribosome (15). In the last ten years, modest (3.5–5 Å) to high (<3.5 Å) resolution crystal structures have been solved for all of these enzymes (**Figure 1**), except the VS ribozyme. Moreover, excluding RNAse P, the structure for each of the remaining ribozymes has been solved in multiple states along its reaction trajectory, providing a yet more detailed view into the catalytic strategies each employs. Progress is still being made in the two outlying cases: molecular modeling has allowed researchers to model RNase P in complex with its pre-tRNA substrate (29, 30), and a preliminary atomic model of the VS ribozyme has been derived from an elegant small-angle x-ray scattering experiment (31). However, thorough structure-function analyses of these enzymes await high-resolution crystal structures.

For years it was assumed that ribozymes would employ divalent metal ion cofactors as their sole catalytic functional groups, in large part because divalent metals are frequently required for RNA folding, and in part by analogy to the mechanisms of proteinaceous phosphoryl-transfer enzymes (32). Hence ribozymes were subordinated as being nothing more than a special class of metalloenzymes (33), mere scaffolds for positioning metal ion catalysts into proper alignment (34). This assumption, however, is now known to be untrue.

A more expansive theoretical framework for understanding the catalytic strategies available to RNA can be drawn by analogy to those used by its protein counterparts. Though protein enzymes accelerate a bewildering array of chemical transformations, decades of research have revealed that the catalytic strategies they use can be approximated by a small handful of general themes (35-37). In recent years, crystallographic examination of the natural ribozymes has shown that, with the exceptions of free radical or quantum tunneling catalysis, these enzymes employ many–if not all–of the same catalytic strategies used by their proteinaceous brethren (**Figure 2**) (21, 38, 39). For example, crystal structures have illustrated substrate approximation in the Ribosome (15, 40-42), and substrate reorientation in the Group I introns (43-46) and Hammerhead ribozyme (47-50). Pauling's conjecture–that an enzyme binds its reaction's transition state structure more tightly than its binds those of the ground states (51)–is illustrated by crystal structures of the Hairpin ribozyme trapped at various states along its reaction trajectory (52-57). Additional contributions from electrostatic and specific acid-base catalysis have also been ascribed to the Hairpin based on these results. Crystal structures have confirmed that functional groups at the active sites of three ribozymes–the HDV (58, 59), Hammerhead (49, 50) and GlmS (60-62)–are unequivocal general acid-base catalysts, a particularly controversial finding in that it implicitly requires dramatic perturbation of these functional groups' pK_as (63). Finally, catalysis by four ribozymes is known to require the assistance of at least one coenzyme. In the case of the GlmS ribozyme, the catalytic cofactor is the small molecule Glucosamine-6-phosphate, which functions as a general acid and protonates the oxyanion leaving group (61, 62, 64). The HDV ribozyme is thought to use a hydrated divalent metal ion as a general base, deprotonating the 2′-hydroxyl nucleophile (59, 65, 66).

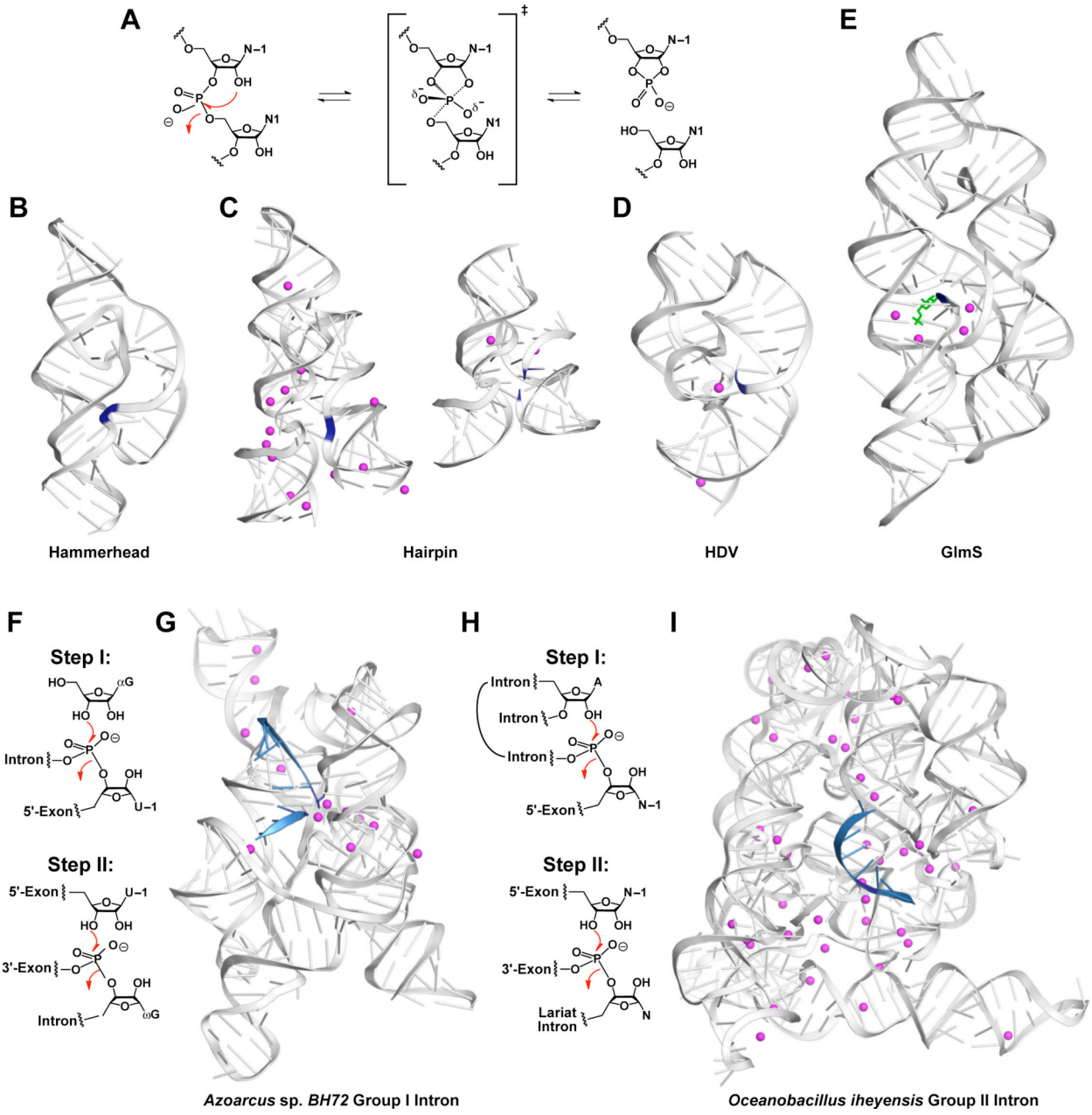


Figure 1. Crystal structures of the known natural ribozymes. In reaction schemes, protonation and deprotonation events are not specified. In structure panels, the reaction center (when present) is dark blue; divalent metal ions are shown as magenta spheres. (**A**) Reaction catalyzed by the self-cleaving ribozymes. Phosphorus centers in the reactions depicted in subsequent panels pass through the same transition state trigonal bipyramidal geometry, though the attacking nucleophiles and leaving groups differ. (**B**)–(**E**) Representative crystal structures of the self-cleaving ribozymes. Except for the "junctionless" Hairpin, each ribozyme has been chemically trapped prior to self-cleavage: (**B**) Hammerhead (49), (**C**) *left*, four-way junction (52) and *right*, "junctionless" Hairpin (56), (**D**) HDV (59), (**E**) GlmS (61). The GlmS coenzyme, glucoseamine-6-phosphate, is depicted as green sticks. (**F**) Two-step reaction catalyzed by the Group I self-splicing intron. αG refers to an exogenous guanosine, 5′-GMP, GDP or GDP. (*Continued*)

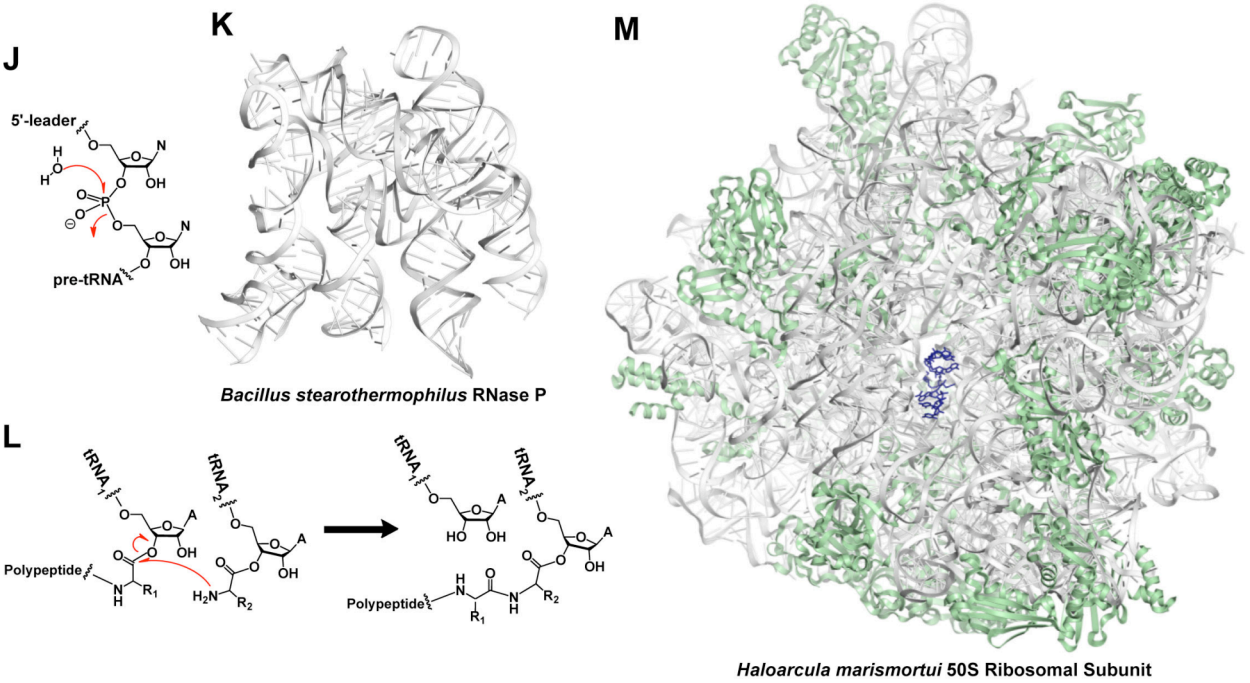


Figure 1. (*Continued*) ωG is the last nucleotide of the intron core. (**G**) Crystal structure of the *Azoarcus* Group I intron prior to the second splicing step (46). The two exons are colored light blue. (**H**) Two-step reaction catalyzed by the Group II self-splicing intron. The attacking residue in the first step is an internal adenosine on the intron itself. (**I**) Crystal structure of the *O. iheyensis* Group II intron following the second splicing step, prior to product release. Ligated exons are colored light blue. (**J**) Reaction catalyzed by Ribonuclease P (RNase P). (**K**) Crystal structure of an RNase P ribozyme; the view is peering in toward the active site, as presumed based on biochemical data. The structure was solved in absence of its protein cofactor and substrate (29). (**L**) Reaction catalyzed by the ribosome. (**M**) Crystal structure of the *H. marismortui* large ribosomal subunit, with chemically blocked acyl-tRNA substrate analogs (dark blue sticks) bound in the A- and P-sites (67). Ribosomal proteins are colored light green. Structure panels in this and all following images were made in PyMol (68).

Last, the Group I (46, 69) and Group II (70, 71) self-splicing introns are known to use a pair of divalent metal ions cofactors, and are hence the sole examples of ribozymes that conform to the two-metal-ion mechanism once proposed for them all (32). Though a full discussion of the data underlying the above assertions is beyond the scope of this chapter, biochemical evidence has largely confirmed the crystallographic results and *vice versa*.

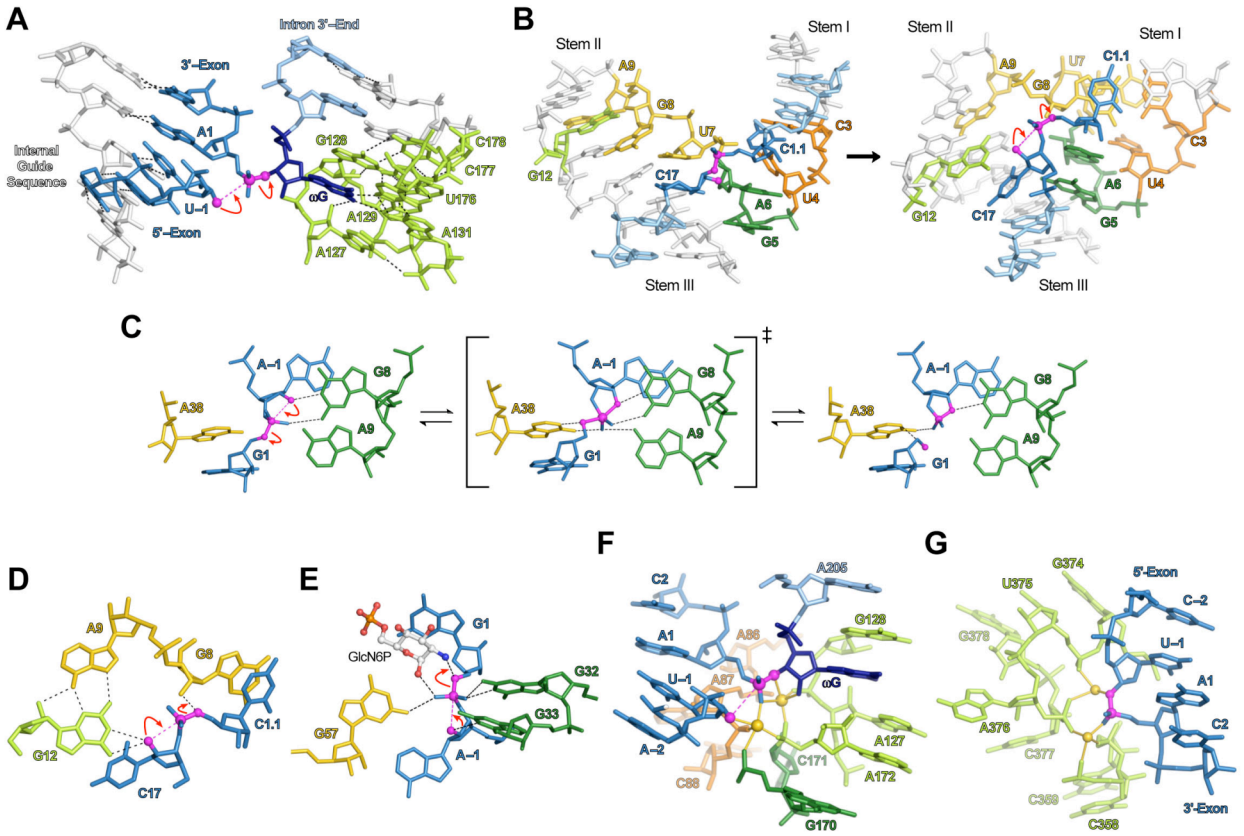


Figure 2. Crystallographic evidence for some of the catalytic strategies used by natural phosphoryl-transfer ribozymes. Magenta spheres mark nucelophiles, electrophiles and departing oxygens; magenta dotted lines indicate the direction of nucleophilic attack. Black dotted lines denote hydrogen bonds; gold bars indicate inner-sphere coordination. (**A**) Substrate approximation and orientation by the Group I intron (43, 46, 69). Docking ωG into its binding pocket (light green, left) forces the backbone into a ~180° turn, positioning the ligation junction phosphate in ideal geometry for in-line attack. (**B**) Substrate orientation in the Hammerhead ribozyme. Peripheral tertiary contacts reorganize the active site (49), contorting the scissile phosphate from the catalytically untenable geometry observed in minimal Hammerhead constructs (*left*)(48) into the ideal geometry for in-line attack (*right*). In the right-hand structure, the G8•C3 pair forces Stems I and II into closer proximity, swinging G5 and A6 outward. These present a wedge that contorts C17 and C1.1 into the catalytically viable orientation. (**C**) The Hairpin ribozyme binds more tightly to the geometry of its transition state than to those of the ground states. Active site residues G8, A9 and A38 make only two hydrogen bonds to the attacking nucleophile and scissile phosphate in the ground-state structure (*left*), and three in the cleaved product structure (*right*), but are observed to make five in the structures of vanadate transition-state mimics (*middle*) (53, 56). Additional work has implicated contributions from water-mediated specific base catalysis (55) or electrostatic catalysis by A38 and G8 (57), which are known to have neutral pK_as (39). (**D**) General acid-base catalysis in the Hammerhead ribozyme. G12 acts as a general base, activating the 2′-hydroxyl nucleophile for attack. The G8 2′-hydroxyl serves as a general acid, protonating the 5′-hydroxyl leaving group (49, 50). Active site water molecules (*not shown*) are thought to aid both functional groups (49, 72), though their roles remain unclear. (*Continued*)

Figure 2. (*Continued*) (**E**) Coenzyme-assisted general acid-base catalysis in the GlmS ribozyme. G33 provides the general base, while the 3-amino group of the gluocosamine-6-phosphate (GlcN6P) coenzyme serves as general acid (61). Functional groups on G32, G57 and the GlcN6P 2-hydroxyl group also serve as electrostatic catalysts, stabilizing the developing negative charge on the scissile phosphate. (**F**)–(**G**) Two-metal-ion catalysis by the (F) Group I (43, 69) and (G) Group II (70, 73) self-splicing introns. Magnesium ions are depicted as yellow spheres. Note that the phosphate-dense binding site in the Group I intron involves residues from five discrete secondary structural features: the ligation junction residues of helix P1, A87–C88 of J5/4, A127 of P7, C171 of J8/7 and ωG. In contrast, all Group II active site metal ion ligands are provided by an ~20-nt distorted helix in domain V.

The combined powers of crystallography and biochemistry have thus facilitated an incredibly detailed view into the mechanisms of the natural ribozymes. However, the catalytic RNAs born from *in vitro* selection experiments have not enjoyed the same depth of mechanistic and structural scrutiny.

Enzymology of The Artificial Ribozymes

The Scope of RNA Catalysis

Despite the limited scope of chemical transformations accelerated by natural ribozymes, *in vitro* selection has revealed the catalytic potential of nucleic acids to be quite broad (18, 74). *In vitro* selection follows a scheme evocative of Darwinian evolution. A genetically diverse population is subjected to a selective constraint; the propagation of any individual member is determined by its genetic fitness with regard to this constraint, and so, over successive generations the most fit individuals will come to dominate the population. Here, the genetically diverse starting population is a pool of 10^{14}–10^{16} random sequences, and an individual's propagation refers to its survival through a round of (RT)PCR. The selective constraint is applied by challenging this pool to perform a desired task, such as binding a target molecule or chemical self-modification with a novel functional group. The experiment is designed such that, by performing the desired task, successful individuals can be physically separated from their

unsuccessful cohort, amplified, and used as the starting point for subsequent rounds of the same procedure. Through iterative cycles of selection and amplification, successive pools become enriched for individuals possessing the desired activity. Variations of this method have been adapted for the isolation of novel RNA (75-77), DNA (78) and polypeptide (79-82) species.

The true power of *in vitro* selection lies in its flexibility, since in theory one could select for any activity that imbues an active species with biophysical properties facilitating its isolation from inactive species. For example, *in vitro* selection experiments have been used to explore the mechanism of RNA recognition by natural RNA-binding proteins (76, 83) or to evolve RNA species that bind proteins lacking this activity (84). Likewise, long before the discovery of riboswitches, artificial RNA aptamers had been isolated that bound small molecules (85); in some cases these exhibited extremely tight binding affinities (86) and acute selectivity between similar ligands (87). *In vitro* selection has also been used to alter the folding and activity of naturally occurring ribozymes (77, 88-92), but its use has truly flourished in the generation of artificial ribozymes with novel activities (18, 74, 93).

Akin to their natural ribozyme counterparts, many of the novel *in vitro* selected ribozymes catalyze reactions involving a phosphate center. As has been previously noted (18), the justification for this is largely technical, in that selection is often based on self-modification, and RNA molecules bristle with phosphate groups. Myriad catalysts for all imaginable permutations of phosphate transfer have thus been selected. Their reactions include: RNA cleavage (94-97), cyclic-phosphate hydrolysis (94), pyrophosphate hydrolysis (98), hydroxyl phosphorylation (99), aminoacyl-adenylation (100), adenylation of diverse cofactor precursors (101), synthesis of phosphamide bonds (81), synthesis or exchange of polyphosphates (102, 103) and the formation of 5′-5′ (102-104), 3′-5′ (105-110), and 2′-5′ phosphodiester bonds, both

linear (105) and branched (111, 112). This last reaction is of particular biological interest, since it resembles the first step of pre-mRNA splicing as catalyzed by both the spliceosome and group-II introns.

RNA has also proven to be a versatile catalyst for reactions not involving phosphorus centers. Two such ribozymes, catalyzing the stereoisomerization of a bridged biphenol compound (113) and porphyrin metalation (114), were isolated as aptamers for transition state analog compounds, a technique similar to that which had previously been used to generate catalytic antibodies (115). However, for the isolation of catalytic RNA, this has generally proven a less robust method than direct selection for catalytic activity. Direct selections have produced ribozymes that catalyze: imino alkylation (116), hydroxyl alkylation (117-119), acyl transfer (118), the formation of amide (120), peptide (121-123), or thioester (124) bonds, the Michael addition (125), Diels-Alder cycloaddition (126, 127), the aldol reaction (128), the Claisen condensation (129), the Amadori rearrangement (130), formation of the glycosidic bond in a pyrimidine (131) or a purine (132) nucleotide, metal-metal bond formation (133) and alcohol dehydrogenation (134).

Thus, with the exception of free-radical reactions, nearly every type of common chemical transformation accelerated by proteinaceous enzymes is now known to be within the catalytic scope of RNA. However, compared to the natural ribozoymes, the structural and mechanistic understanding of how artificial ribozymes catalyze these diverse reactions is far less developed. In most cases, studies have focused more on the catalytic capacity of a given ribozyme, and less on the manner in which it performs catalysis. Structural characterization of a novel ribozyme has often been limited to the determination of its secondary structure, the minimal catalytic motif and invariant nucleotides therein and–if the RNA was selected owing to its ability to self-modify–the

nucleotide thus modified. Mechanistic characterization has often been limited to an exploration of the substrate modifications and buffer conditions tolerated by the ribozyme. Some examples are of this are presented below.

Substrate modification studies can sometimes yield broad insights into the ways a ribozyme binds and turns over its substrates. For example, since an artificial peptidyl transferase ribozyme can catalyze the formation of over 30 different dipeptides from diverse AMP-activated substrates (122), this enzyme likely makes few contacts with the acyl groups on these substrates. A similar conclusion can be drawn from the promiscuity of a cofactor synthetase ribozyme (101). Conversely, since an aminoacyl-tRNA synthetase "flexizyme" (*see below*) is highly specific for an activated phenylalanine substrate, but will tolerate activation by various leaving groups (135), this ribozyme likely makes few specific contacts to those groups. Characterization of the diverse products formed by a purine synthase ribozyme (130) suggested that catalysis proceeds through the formation of a Schiff's base with its incoming 6-thioguanosine substrate. In none of these cases, however, have attempts been made to elucidate the specific functional groups within the ribozyme that are responsible for these phenomena.

The metal ion dependencies of many artificial ribozymes have been characterized, though, since most RNA species require divalent metal ions for folding, the mechanistic implications of these results can be difficult to interpret. Of potential interest are the obligate Ca^{2+}- and Zn^{2+}-depdencies of a phosphatase (98) and an alcohol dehydrogenase (128) ribozyme, respectively. Despite the ribozymes' having been selected from a complex mixture of available metals, these dependencies mirror the metal ion requirements of their analogous proteinaceous enzymes. Though little detailed mechanistic insight can be gleaned from these results (a similar phosphatase ribozyme prefers Mg^{2+} or Mn^{2+} over Ca^{2+}, for example (103)), they may be

considered as examples of convergent evolution of catalytic strategies. Very thorough investigations have explored metal ion binding by a diels-alderase ribozyme–which will be discussed in greater detail below–and self-acylating ribozymes (136-138). In this latter case, studies of metal ion tolerance have been combined with lanthanide footprinting and nucleotide analog interference mapping (NAIM) and revealed that this ribozyme probably does not employ a metal ion cofactor directly in catalysis.

Remarkably, although a study of pH-dependence would seem a requisite early experiment in the study of a ribozyme's catalytic mechanism, relatively few artificial ribozymes have been characterized even in this manner (136, 139-142). An outlying example is the artificial "X-motif" ribozyme, for which self-cleavage was shown to be log-linear with increasing pH (143). This result implies a single rate-determining deprotonization event during catalysis, in stark contrast with the natural self-cleaving ribozymes.

Beyond these experiments, there are a handful of examples in which the mechanism of an artificial ribozyme has been elucidated in greater detail. Ligase ribozymes in particular have proven popular subjects for detailed examination and will be discussed at greater length below. Another example is the pyrimidine synthetase ribozyme, which has been studied through a series of primary and secondary kinetic isotope effect (KIE) experiments, using a double-labeled substrate partitioning strategy (144). The primary KIE demonstrated little to no bond order between the anomeric ribose carbon and the pyrophosphate leaving group, while the secondary KIE was consistent with sp^2-hybridization at the anomeric carbon. Together, these results imply a dissociative reaction scheme, more similar to that employed by natural degylcosylating enzymes than it is to natural nucleotide synthetases. In a second example, elaborate kinetic schemes have been elucidated for two classes of self-capping ribozymes (141, 145), each of

which also catalyzes the competing reactions of cap exchange and hydrolysis. Owing to a rate-limiting pyrophosphate release step that precedes cap-formation, both enzymes had been predicted to form on-pathway covalent intermediates. For one of these ribozymes, an inspection of the reaction stereochemistry supports this model (141). In a third example, a peptidyl-transferase ribozyme has recently been shown to be faster in monovalent ions alone than it is in divalents, implying that catalysis does not require the direct participation of metal ions (142). Furthermore, given its log-linear pH dependence and a deuterium kinetic solvent isotope effect (KSIE) of ~10, it is possible that, unlike the ribosome, peptide bond formation by this small riboyme involves rate-limiting proton transfer. In a fourth example, a photolyase deoxyribozyme (DNAzyme) that repairs *cis*, *syn*-cyclobutane pyrimidine dimers has been characterized by mutagenesis, substrate cross linking and nucleotide analog studies (146). These and previous spectroscopic analyses imply that this enzyme uses a G-quartet structure as a photoactive "antenna," which is thought to harnesses the energy in a photon in order to promote electron transfer from an active site guanosine to its substrate.

In all, a deep mechanistic understanding has been developed for relatively few artificial catalytic RNAs. Although a wide gamut of powerful biochemical tools is available for probing RNA structure and function (147-150), some of which can simultaneously interrogate all of the residues on an RNA species with single-atomic precision, their use with *in vitro* selected ribozymes has been rather limited. Moreover, perhaps the most powerful structural analytic technique in molecular biology, x-ray crystallography, has previously been applied to only four *in vitro* selected catalytic RNAs, in most cases at only one point along each ribozyme's reaction trajectory (**Figure 3**). Each of these structures is presented in detail below.

The Diels-Alderase Ribozyme

In vitro selection has yielded two small catalytic RNAs that catalyze Diels-Alder [4+2] cycloaddition between an RNA-tethered diene and an exogenous maleimide dienophile (126, 127). The reaction these enzymes catalyze is invaluable for organic chemists, though its use by enzymes is still debatable (151). Nonetheless, it had previously been a popular target for the selection of catalytic antibodies (152, 153), and its acceleration by multiple RNA species was the first demonstration of RNA-catalyzed carbon-carbon bond formation. Since one of these ribozymes (126) is built using modified uridines, the other (127) has emerged as the more popular model system by far: in addition to crystallographic studies, the structure, folding and catalysis of this ribozyme have been examined by EPR (154), single-molecule FRET (155), AFM (156), active site photo-crosslinking (157) and NMR (158). Satisfyingly, the results of all of these have largely corroborated the crystallographic work described below.

Crystal structures have been solved of the free Diels-Alderase ribozyme at 3.5 Å, and of the ribozyme bound to its cycloaddition product at 3.0 Å resolution (159). Its structure is remarkably compact for a 49-nt species, being comprised of three helices pulled into close proximity by a central pseudoknot, converging into the shape of the Greek letter λ (**Figure 3A**). This architecture is clasped together through an extensive network of noncanonical base pairs and triples involving residues initially thought to be unpaired. Further stabilizing the overall structure are six Mg^{2+} ions in various states of hydration. While these metals make biochemically confirmed contacts that buttress interactions at the point of helical convergence (154, 155, 158), none are observed near the active site.

In fact, the active site of the Diels-Alderase is nearly devoid of polarizable or ionizable groups of any sort, being almost entirely lined by the aromatic rings of nucleobases in the central

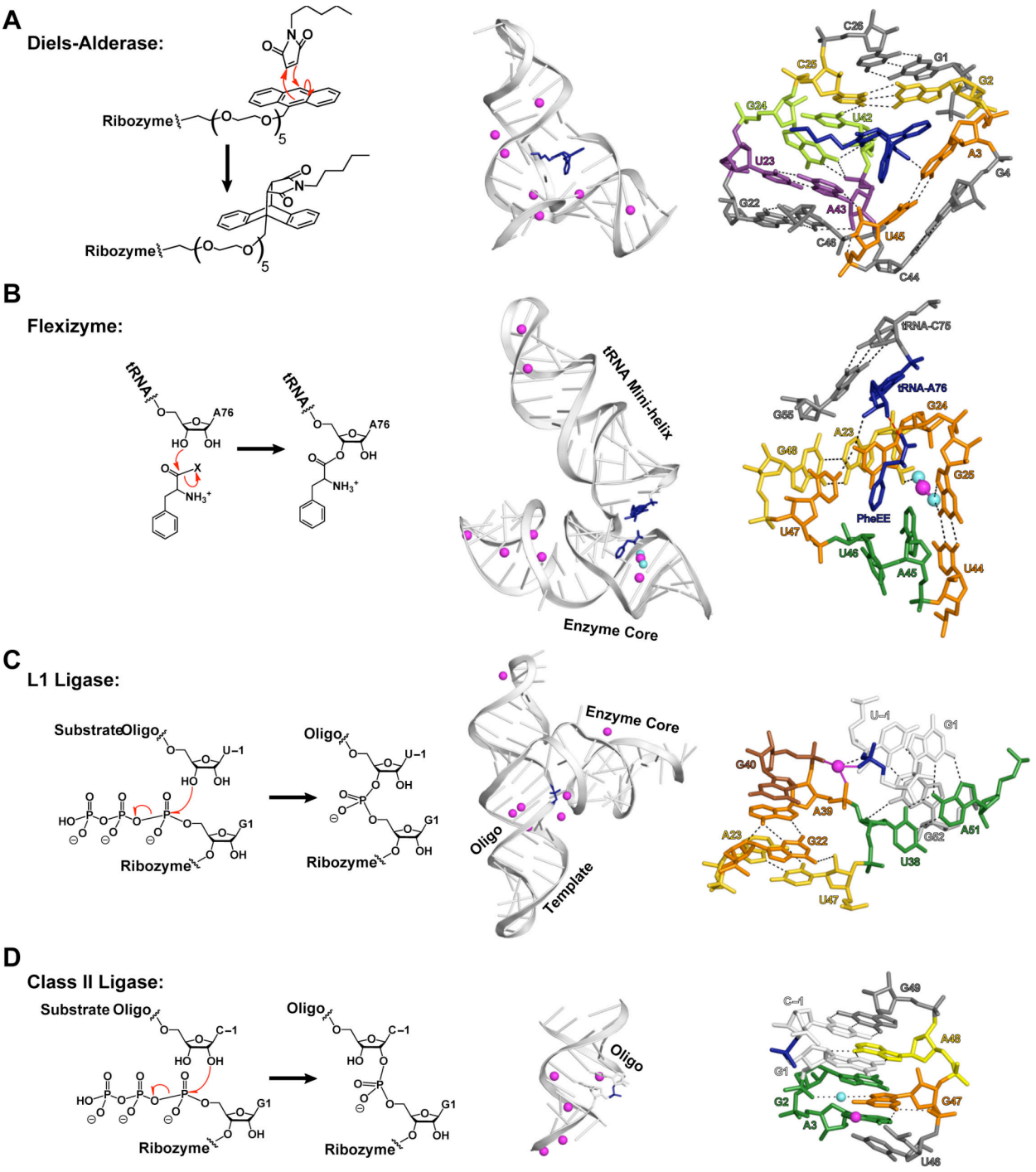


Figure 3. Crystal structures of artificial ribozymes. Each panel shows (*left*) the reaction catalyzed by, (*middle*) the global architecture of, and (*right*) the active site of its respective enzyme. In all structure panels, substrates, products or ligation junctions are colored dark blue. Where present, magnesium ions and waters are depicted as magenta and aqua spheres, respectively. In right-hand panels, bases involved in substrate binding or catalysis are colored; those abutting the active site are shown in gray. (*Continued*)

Figure 3. (*Continued*) Black dotted lines indicate hydrogen bonds; magenta bars indicate metal ion inner-sphere coordination (**A**) The Diels-Alderase ribozyme. Structure is of the enzyme bound to its cycloaddition product (159). (**B**) Flexizyme. "X" can be a variety of leaving groups. Structure is of the enzyme covalently linked to a tRNA mini-helix, bound to a substrate analog, PheEE. Model based on (160), with the docked PheEE coordinates kindly provided by Adrian Ferré-D'Amaré. The red dashed line in the right panel indicates the proposed line of nucleophilic attack. (**C**) The L1 ligase ribozyme. Structure is of a circularized self-ligation product (161). Note the U38•A51•G1 base triple, the sole tertiary contact. (**D**) The Class II ligase ribozyme. Structure is of the minimal motif product (162).

pseudoknot. In the apoenzyme structure, the pseudoknot houses a wedge-shaped central cleft spanning a ~31º angle, encased by canonical A•U pair, a noncanonical A•U pair and an A•G•C triple on each of its three sides. These are abutted by an unpaired guanosine that induces an asymmetric widening of the central cavity, creating a "hydrophobic canyon." In the product structure, the A•U pairs and A•G•C triple stack perfectly against each aromatic face of the cycloaddition product, while the canyon provides a binding pocket complementary to its aliphatic linker. The asymmetry of this canyon limits the orientation in which the two substrates could approach one another, explaining why the Diels-Alderase exclusively catalyzes the enantiospecific formation of (S,S) products (163).

Of particular note, the free and product-bound Diels-Alderase structures are almost perfectly superposable, which confirms the previous prediction of a rigid, pre-formed active site (164) and has been further corroborated by NMR (158). Active sites consisting of little more than sterically complementary aromatic residues were also observed in the Diels-Alderase catalytic antibodies (152, 153). These results can easily be justified mechanistically, since active sites of this sort would provide ideal binding spots for aromatic substrates and would present complementary binding sites for stabilizing the transition state geometry. Catalysis by these

enzymes, then, is thought to be achieved entirely through substrate approximation and orientation, though contributions from quantum effects are hard to gauge.

The Aminoacyl-tRNA Synthetase "Flexizyme"

Though several aminoacylating ribozymes have been selected *in vitro* (117, 135, 139, 165), the so-called "flexizyme" is unique in multiple regards. First, although its selection scheme was based on self-modification, it was designed to favor the isolation of species capable of acylating the 3′-terminus of a tRNA *in trans* (135). Since the resulting ribozyme appeared to recognize its tRNA substrate exclusively by base pairing with its 3′-CCA sequence, flexizyme variants could be programmed to specifically acylate an individual tRNA species by engineering additional contacts between the ribozyme and the tRNA acceptor stem (166). Second, while flexizyme achieves ~10^4–fold discrimination between its proper phenylalanyl substrate and noncognate amino acids, it tolerates a variety of modifications in the chemical groups used to activate this substrate, ranging from 5′-AMP to cyanomethyl ester (167). Together, these properties have made flexizyme an invaluable resource for engineering *in vitro* translation systems that incorporate nonnatural amino acids, thus yielding programmable ribosomal synthesis of nonbiological polymers (168, 169).

The 2.8 Å crystal structure has been solved of the flexizyme covalently bound to a tRNA-like minihelix substrate, a species evocative of a noncovalent flexizyme•tRNA complex (**Figure 3B**) (160). Compared to the compact, tightly knitted architecture of the Diels-Alderase ribozyme, the flexizme structure is relatively austere. Its architecture consists of an irregular helix nestled between two A-form helices, assembled into a single pseudocontinuous duplex without the participation of any long-range tertiary interactions. The sole inter-domain contact is

a three-nucleotide hairpin turn that joins the enzyme to the substrate tRNA, presenting the acceptor stem into the active site at an angle roughly perpendicular to the enzyme core. As previously predicted (170), the tRNA is recognized exclusively by base pairing between its 3′-CCA terminus and the 3′-end of the enzyme.

The flexizyme active site is entirely comprised of a short irregular helix, consisting of two unpaired residues that are enclosed by noncanonical base pairs: on one end a G•U wobble, and on the other a G•A pair and a single hydrogen bond "stretched" G•U pair. A partially hydrated Mg^{2+} may tether the structure together through major groove contacts to each of the two G•U pairs (138), but this is too far removed from the active site to be directly involved in catalysis. However, the stretched G•U pair seems to be functionally quite important: the unusual geometry widens the active site duplex, allowing its guanosine to stack against the substrate phenylalanine ring. This same guanosine forms a hydrogen bond with the tRNA 2′-hydroxyl through its N2 amine, an occlusive contact that may explain why flexizyme acylates the tRNA 3′-hydroxyl regioselectively. The importance of this stretched G•U pair was furthermore confirmed through a targeted *in vitro* selection in which only the catalytic core residues were mutagenized. Only three residues were revealed to be invariantly required for activity: the two nucleotides in this pair and the uridine that caps the hairpin loop.

Since neither the 3′-aminoacylated product nor any of the activated phenylalanyl substrates was sufficiently stable in water to survive crystallization, crystals were grown in the presence of a longer-lived substrate analog, PheEE. Only one of the two crystallographically independent monomers was observed to bind this compound. It was therefore argued that each of the monomer structures was analogous to a discreet step along the reaction trajectory, corresponding to the states immediately prior to and after substrate binding. If this hypothesis is

correct, then considering the gross structural differences between these two monomers, binding activated phenylalanine reorients several active site residues by ~0.5–1.2 Å, and concomitantly induces a global reorientation of the tRNA substrate helix with respect to the enzyme core. Coupling of phenylalanine and tRNA binding had been previously predicted for this ribozyme (170) and is reminiscent of the induced-fit mechanism reported for some natural aminoacyl-tRNA synthetases (171). However, this enzyme exhibits a paucity of candidates for functional groups directly involved in catalysis. Considering the sequence malleability of all other active site residues, and since the crystal structure revealed no specific contacts made to functional groups on either substrate, it is possible that flexizyme catalyzes its reaction purely through substrate approximation.

The L1 Ligase Ribozyme

The L1 is one of several ligase ribozymes to have been isolated from similar *in vitro* selection schemes (*see below*). This particular ribozyme catalyzes the formation of a 3′-5′phosphodiester bond between an oligonucleotide substrate and the α-phosphate of its own 5′-triphosphate (106), a reaction chemically similar to that of proteinaceous nucleic acid polymerases. Unlike these polymerases, however, catalysis by the L1 ligase is highly sequence dependent: mutagenesis studies have revealed that the ribozyme requires its nucleophile to be positioned by a noncanonical G•A pair, adjoined on either side by G•U wobble pairs (172). Initial characterization also revealed that the L1 is highly flexible; certain sequence variants are even able to stably populate two discrete secondary structures, of which only one is catalytically viable. This trait has been exploited in the engineering of allosteric L1 variants that function as molecular sensors, wherein ligation is responsive to the presence of oligonucleotides (106), small

molecules (106, 173) and polypeptides (174, 175), in cases achieving activation ratios as high as 50,000.

The crystal structure of a self-circularized L1 ligation product has been solved to 2.6 Å resolution (**Figure 3C**)(161), a technical achievement that required the development of a novel *ab initio* phasing method (176) that has proven useful in the solution of other RNA crystal structures (50, 160). Two domains comprise the roughly γ-shaped L1 ligase architecture, one bearing the template•oligonucleotide substrate duplex, the other bearing the "enzyme core;" a short hinge region positions these domains nearly perpendicular to one another. However, much like the flexizyme structure described above, the asymmetric unit of this L1 ligase crystal contained two crystallographically independent monomers in dramatically different conformations. One molecule is thought to represent a catalytically valid "docked" conformation, in which the ligation junction is positioned within the enzyme's active site in large part through the action of an inter-domain G•A•U base triple. In the other molecule, an ~79° unwinding of the hinge region precludes the formation of this contact, and as a result the enzyme domain is pulled away from the ligation junction and rotated ~180° away from the viable conformation. Proper folding of the hinge region thus appears vital for catalysis, a property further corroborated by its high degree of conservation among active L1 isolates (172).

As observed in the docked conformer, the L1 ligase active site appears to be built from two adjoining structural motifs. The first is a duo of noncanonical U•A and G•A base pairs within the enzyme core; these enclose an unpaired uridine that is consequently splayed out of the helical stack. Extrusion from the enzyme core enables this uridine to participate in the second active site structural motif: the aforementioned G•A•U base triple with the triphosphate-bearing G1 nucleotide, the lone contact between the enzyme and template•substrate domains. The

combination of these features induces a profound kink in the backbone near the extruded uridine, contorting the phosphates of two nearby sequential residues into a binding site for a high-occupancy Mg^{2+} ion. This metal is placed ~2.2 Å from the ligation junction phosphate, positioned with nearly ideal coordination geometry to play a direct role during catalysis. Moreover, coordination of this metal pulls one of the active site phosphate oxygens ~2.9 Å away from the ligation junction 3′-bridging oxygen, the remnant of the reaction nucleophile. It was therefore proposed that metal coordination might perturb the pK_a of this phosphate, potentially allowing it to function as a general base during catalysis by abstracting a proton from the hydroxyl nucleophile. While this contact is apparently the only direct interaction to the nucleophile, a high-occupancy water molecule was observed participating in a network of hydrogen bonds with the adjacent 2′-hydroxyl. Based on this, the authors originally speculated that this water might underlie the L1 ligase's observed 3′-hydroxyl regioselectivity, since it would immediately quench any specious deprotonation of the 2′-hydroxyl. However, subsequent mutagenesis studies (*see below*) have implied that local sequence context, more than any potential solvent interactions, may have a greater role in determining regiospecificity.

Overall, then, the L1 ligase ribozyme might catalyze RNA assembly by positioning the reactive hydroxyl and triphosphate in close proximity to a metal ion cofactor (32). The proposition that this metal cofactor, in addition to transition state charge stabilization, might also activate a phosphate general base is compelling, though untested.

The Class II Ligase Ribozyme

The first *in vitro* selection experiment to isolate novel catalytic RNAs from random sequences ultimately produced three distinct structural classes of self-ligating ribozymes (93,

105). Each of these enzymes promoted the nucleophilic attack by the terminal hydroxyl of an oligonucleotide substrate on their own 5′-triphosphate, generating a new phosphodiester bond and yielding a pyrophosphate. While the so-called Class I ligase selectively employed the 3′-hydroxyl on its substrate as the reaction nucleophile, generating a common 3′-5′ phosphodiester linkage as a result, the Class II and Class III ligases regioselectively accelerated the formation of 2′-5′ linkages. Members of the Class II family had the simplest secondary structures of all three classes, consisting of a single irregular duplex that situated the triphosphate-bearing nucleotide within an asymmetric bulge. While other structural elements unquestionably aided catalysis–the best class II isolate accelerating its reaction some 10^{10}-fold over the uncatalyzed rate–the central asymmetric bulge could be thought of as a seven-nucleotide minimal catalytic motif. An otherwise unmodified helix positioning its oligonucleotide substrate and 5′-triphosphate within this motif accelerated phosphodiester bond formation ~10^4-fold over the background rate, and maintained its preference for using the 2′-hydroxyl nucleophile by a factor of 4.5:1 (105).

Efforts to crystallize the intact Class II ligase at any point along its reaction pathway have proven unsuccessful, as have attempts to crystallize the minimal motif prior to ligation. However, two crystal forms of oligonucleotide duplexes corresponding to the minimal motif ligation product have been solved to 2.7 and 2.3 Å resolution (**Figure 3D**) (162). Each of these Class II minimal motifs is essentially a short, irregular helix, from which the authors distilled three distinct structural features that contribute to catalytic activity and regioselectivity. First is the triphosphate-bearing G1 nucleotide itself, which participates in a *trans* sugar-edge•Hoogsteen G•A base pair (177) and is consequently pulled into the major groove. The adenosine of this pair forms a cross-stand stack atop the second structural feature, a “steric wedge” formed by two consecutive unpaired purines. This wedge is pushed into and concomitantly widens the minor

groove just below the ligation junction. Opposite the wedge lies the third feature, an unpaired guanosine that alternately adopts *syn* and *anti* glycosidic bond conformations in the two crystal forms. The changes induced by this variation are slight, inducing a minor alteration in the G1•A propeller twist and creating different binding sites for partially hydrated Mg^{2+} ions. Critically, in neither of these structures is a metal ion observed within 8 Å of the ligation junction.

Systematic mutagenesis studies on the full-length Class II ligase confirmed that all three of these features contribute to the catalytic rate; alterations that reposition the G1 nucleotide or which alter the unpaired guanosine were most damaging (162). Moreover, observing that the Class II and L1 ligases position their triphosphate-bearing nucleotides by the same noncanonical base pair, but selectively form 2′-5′ and 3′-5′ linkages, respectively, an attempt was made to alter the regioselectivity of each enzyme through structure-guided design. Comparison of the two crystal structures implied that the geometries of substructures bracketing the triphosphate-bearing G1 might have the most profound effect on regioselectivity: since the G1•A base pair is shifted into the major groove, compensatory alterations to the nucleophile-bearing nucleotide and downstream steric wedge might reorient the triphosphate with respect to the potential 2′- or 3′-hydroxyl nucleophiles. A hybrid Class II ligase in which each of these active site structures had been converted to the analogous L1 ligase sequences did indeed exhibit a complete conversion in regioselectivity, though this came at the expense of a 10^8-fold loss in rate enhancement. The converse experiment–in which Class II ligase structures were transplanted into the L1–resulted in a more modest conversion of stereoselectivity. At the expense of a ~100-fold drop in overall activity, this chimeric construct showed an ~35-fold increase in 2′-5′ bond formation relative to the wild type L1, though still favoring 3′-5′ linkages by a factor of 14:1. Collectively, these experiments provide a structural glimpse into a potential mechanism by which ligase ribozymes

select the hydroxyl nucleophile they activate, but are relatively uninformative in explaining how this activation is achieved.

Trends in Artificial Ribozyme Structure

Progress in understanding artificial ribozymes now allows one to speculate in a more informed manner how the products of billions of years of natural selection might differ from those isolated after a few rounds of *in vitro* selection. Though one should take caution in extrapolating from such a small sample set, the artificial ribozyme crystal structures may reveal some general themes regarding the structural and mechanistic differences between natural and artificial RNA catalysts.

First, *in vitro* evolved ribozymes appear less compact than their natural counterparts. This can be attributed to their relative paucity of highly defined long-range tertiary contacts, which are often observed to buttress the architecture of natural structured RNAs (178), including ribozymes (179). The Hammerhead ribozyme illustrates a beautiful example of how such buttressing might play a paramount role in stabilizing a catalytically viable active site conformation (49, 50), though equally impressive examples abound in the other natural catalytic RNAs (60, 61, 73, 180). While, among the artificial ribozymes, the tightly bundled architecture of the Diels-Alderase ribozyme is an outlying example, other members of its cohort suffer from a near or total lack of long-range tertiary contacts.

A related theme concerns the complexity of artificial ribozyme active sites. Complex architectures allow the natural ribozymes to place a multitude of functional groups in close proximity to one another without many constraints on their relative orientation *a priori*. Comparatively simple architectures would force ribozymes to build active sites from residues

that are close in primary structure, thus imposing tighter spatial restraints on the directions from which the enzyme can engage the reaction center. As an illustration, the GlmS ribozyme forms discrete contacts with each oxygen on the scissile phosphate, makes base- and ribose-specific contacts to the departing nucleotide, and encapsulates its catalytic cofactor with an array of interactions that position it at the active site. Altogether, it employs functional groups on nucleotides presented by four different secondary structural elements to achieve these aims (60, 61). A more dramatic example is found in the Group I intron, which positions its Mg^{2+} catalytic cofactors through the concerted action of six phosphates that converge at the active site from three distant strands (43-45, 69). Binding sites for each of its substrates are built from relatively local structures–grasping each exon in a simple Watson-Crick duplex and guanosine through an irregular helix. The converse strategy is adopted by the Group II intron: while the "catalytic triad" responsible for positioning Mg^{2+} cofactors at its active site resides entirely within the single bulged-helix of domain V (70), binding its exon substrates requires the direct participation of three disparate domains (71) and is buttressed by a staggering number of complex tertiary structures (73). Even the substantially simpler Hairpin ribozyme, using a handful of tertiary contacts to dock its two domains, contorts its substrate into a catalytically competent conformation and positions functional groups from each domain on either side of the scissile phosphate (52, 54). By comparison, the active sites of Flexizyme or the L1 ligase, for example, seem unadorned in features that might specifically modulate substrate binding and catalysis.

A final theme, predicated on the previous two, concerns the inferred catalytic strategies used by artificial ribozymes. Given the comparatively simpler global architectures and active sites of these enzymes, it is perhaps not surprising that their assumed catalytic strategies are also less subtle than those of the natural ribozymes. Considering the four artificial catalytic RNAs

whose crystal structures have been solved, the most convincing evidence indicates that they achieve catalysis purely through substrate approximation, aided in one case by a metal ion cofactor. Of particular note are the diminutive contributions from nucleobases. Although in the natural ribozymes, these can function as electrostatic or acid-base catalysts, in the artificial ribozymes they are thus far known only to function as stacking surfaces. That the subtler aspects of biological catalysis may be beyond the purview of artificial ribozymes might explain their diminished catalytic capacity. A cursory glance at the pantheon of artificial ribozymes reveals average rate enhancements of ~10^2–10^6–fold over their corresponding uncatalyzed reactions (74), as compared to ~10^9–fold for natural self-cleaving RNAs (181) and the famously staggering rate enhancements of 10^{10}–10^{17} (182) achieved by some natural protein enzymes.

Therefore, despite their potential to accelerate a multitude of different chemistries, the artificial ribozymes seem in many regards far simpler catalysts than even the "simple" natural self-cleaving ribozymes. Concerning the RNA World hypothesis, a pessimistic interpretation of these data might be that RNA catalysts are simply best suited for the two chemistries they are currently known to accelerate in nature. Alternatively, a rosier interpretation might posit that these data are merely evidence that the selective pressure exerted by experimenters *in vitro* simply has not matched that imposed during genuine evolution *in vivo*.

RNA Replication in the Absence of Protein

The Hunt for a Replicase Ribozyme

As noted earlier, the RNA World hypothesis proposes a simple solution to the "chicken and egg" paradox underlying the mechanics of the Central Dogma. Yet, an unresolved issue fundamental to this theory is its implicit requirement that RNA be capable of self-replication

(11). After all, in the absence of robust, uncatalyzed polymerization, the putative "ribo-organisms" of the RNA World would reap little benefit from having enzymes that serve as templates for their own replication unless one of these enzymes could catalyze this replication (13, 183-185). However, if such a replicase ribozyme exists in nature, it has not yet been observed.

Multiple lines of investigation have therefore sought to demonstrate that RNA is capable of catalyzing its own replication. Initial progress was made using derivatives of the Group I intron, which could be coaxed into disproportionation reactions that extended a primer with multiple templated pyrimidines (186) or both untemplated (187) and templated mononucleotides (188) with fewer sequence constraints. However, the polymerization fidelity in these reactions is untenably low. Selection of the cognate Watson-Crick base over competing mismatches was at best ~65% in single-nucleotide primer extension reactions (188) and showed a variable degree of improvement when activated trinucleotides were used as monomers (189). Recognizing that primordial polymerization may have exploited longer activated RNAs as monomers, parallel efforts have employed Group I derivatives to assemble multiple oligonucleotides in a template-directed fashion (190), even generating long RNAs complementary to part (191) or all (88) of the enzyme's sequence. In a similar vein, an *Azoarcus* Group I intron that had been split into four smaller, inert fragments was shown catalyze its own covalent assembly *in trans* (192), representing the first step in a kind of self-assembled self-replication. Importantly, all of these Group I-derived experiments mimicked the second step of the intron's natural splicing reaction and employed a mononucleotide (typically guanosine) rather than pyrophosphate as the leaving group. Hence, given the energetics of the transformations they catalyzed, many of these reactions suffered not only from severe sequence limitations, but also from the competing back-

reactions in which extended primers are lysed by exogenous guanosine. An intriguing recent development has been the observation that a Group I isolate from *Anabaena* naturally catalyzes the formation of a 3′-5′ phosphodiester bond using a pyrophosphate leaving group (193), a reaction similar to that of natural proteinaceous polymerases. The extent to which this reaction can be generalized with other natural isolates, and to which this sequence-dependent single-turnover reaction can be broadened to general polymerization, remains to be seen.

Given the limitations in building a replicase from existing natural ribozymes, other efforts have employed *in vitro* selection to generate artificial enzymes that may serve as more productive starting points. For example, an artificial ligase ribozyme termed "R3C" base pairs directly to two short oligonucleotides and catalyzes the formation of a 3′-5′ phosphodiester linkage between them, yielding a pyrophosphate leaving group (108). Using a design philosophy similar to that of the Group I-catalyzed oligonucleotide-linking experiments mentioned above, this enzyme was engineered to ligate two substrates corresponding to truncations of the enzyme itself, hence catalyzing a kind of self-regeneration (194). However, since the enzyme remains base-paired to the new copy, this reaction is strongly product-inhibited. This hurdle was partially alleviated by moving to a cross-catalytic ping-pong format, in which one R3C variant ligated truncated fragments of another, which in turn ligated truncations of the original enzyme (195). Further *in vitro* selection was used to optimize this system for multiple-turnover, yielding a pair of R3C variants that would replicate one another with exponential growth kinetics. Diluting aliquots of the reaction mixture intro fresh buffer and substrates, this reaction could be continued indefinitely (196). While achieving this kind of self-sustained replication by RNA is unquestionably a great milestone, it is unlikely that a system of this sort represents an intermediate on the pathway to general replication. A true general replicase ribozyme would be

a single macromolecular species capable of copying any RNA in a template-directed fashion, using activated mononucleotides. As such, the R3C system falls short of this goal in several regards. First, it is an obligate bipartite system, requiring the presence of both ribozyme variants for replication. Second, replication of this sort is far from general: since each R3C variant positions its substrates by directly base pairing to them, the available sequence space this system could copy is extremely limited. Third, catalysis by the R3C ligase requires that the triphosphate-bearing nucleotide be a guanosine, and that it be unpaired. Finally, the substrates in the R3C reaction are oligonucleotides, not mononucleotides.

The Class I ligase Ribozyme, its Cohort and its Derivatives

To date, the most successful attempts at isolating a general replicase ribozyme have started with an artificial catalytic RNA termed the Class I ligase (**Figure 4A**) which was one of the first novel ribozymes to be isolated by *in vitro* selection (93). Selection in this case was designed to isolate species capable of catalyzing a reaction analogous to RNA polymerization in the manner performed by proteinaceous enzymes. This ultimately yielded three distinct structural classes of self-ligase ribozymes (105), all of which could also ligate exogenous substrate complexes *in trans* (**Figure 4B**). Two structural classes were also able to catalyze multiple-turnovers in this format, making them true enzymes. However, the Class I ligase was unique in multiple regards. Subsequent *in vitro* selection and optimization (197) revealed the Class I secondary structure to be markedly more complex than that of its brethren, consisting of seven paired regions converging in a central four-way junction bracketed by a double pseudoknot (**Figure 4A**). Catalysis by the Class I ligase is by far the fastest of the three classes, its *trans*-ligation rate approaching 10^9–fold enhancement above the uncatalyzed reaction, its self-

ligation rate being too fast to measure by manual pipetting at pH ≥7.5 (105, 198). This was also the only structural class for which the nucleophile and electrophile were positioned by Watson-Crick base pairing. Finally, while the Class II and Class III ligases preferentially accelerated the formation of a 2′-5′ linkage, the Class I ligase was exclusively regioselective for a 3′-hydroxyl nucleophile (105). Overall, Class I ligase reaction–the Watson-Crick context dependent, regiospecific attack by a 3′-hydroxyl on a 5′-triphosphate, yielding a pyrophosphate–is chemically equivalent to a turnover of templated RNA polymerization as it is currently performed by proteinaceous enzymes.

This confluence of traits–structural complexity, catalytic efficiency in both *cis* and *trans* contexts, selection for Watson-Crick base pairing and 3′-regioselectivity–was unique for the Class I ligase at the time of its isolation, and remains nearly unmatched by the numerous ligase ribozymes that have been isolated since. For example, five other ligases are now known to selectively promote the formation of 3′-5′ phosphodiester bonds (106-110). Of these, only the Class hc (109) and the DSL (110) ligases can tolerate a nucleophile and electrophile that are positioned by Watson-Crick pairing; for the Class hc this orientation is inhibitory. Moreover, both of these ribozymes are less efficient than the Class I at *trans* ligation by a factor of approximately 10^5. While the Class hc ligase has been optimized through further *in vitro* selection for ligation in this format (199), the resulting enzyme showed strong sequence dependence at substrate sites distal to the ligation junction, making it an unlikely candidate for a general polymerase.

In addition to its catalytic prowess, the Class I ligase has also served as a versatile and flexible system for studying the population dynamics that occur during the *in vitro* selection process and for altering the format in which selection is performed. An early example is so-

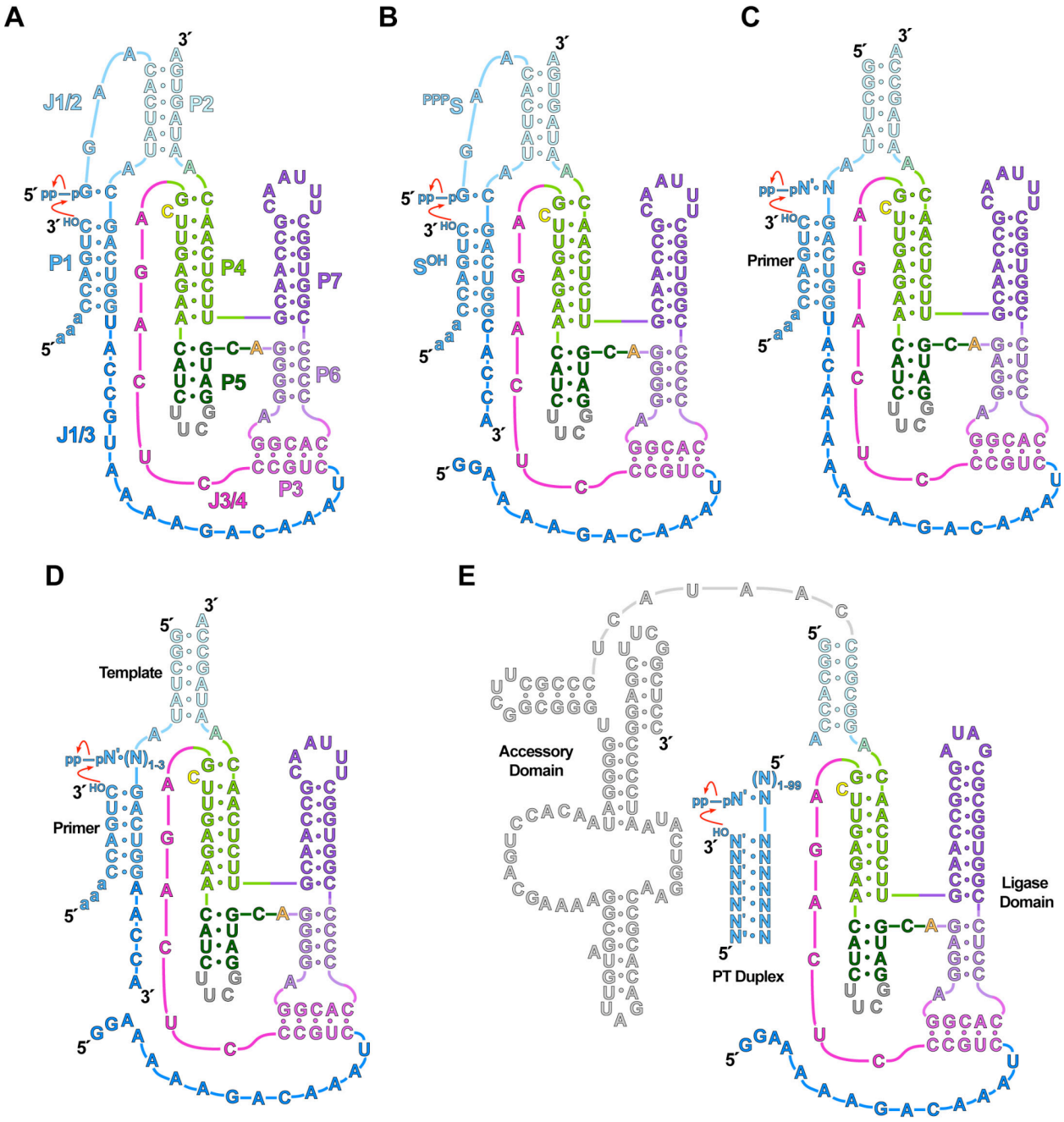


Figure 4. Reactions catalyzed by the Class I ligase and its derivatives. Red arrows indicate electron movement during catalysis. N is any nucleotide; N′ is its Watson-Crick complement. Lowercase letters indicate 2′-deoxy nucleotides. (**A**) *Cis*-ligation with an oligonucleotide substrate. Shown is the secondary structure of the "clone 207" ligase (105, 197), modified as in (200). (**B**) *Trans*-ligation of two oligonucleotide substrates, S^{OH} and ^{ppp}S, as in (198). (**C**) Extension of "primer" oligonucleotide using exogenous NTPs and a template that is covalently linked to the ligase core, as in (201). (**D**) Primer-extension as in (C), whereas the template is bound to the enzyme core by base pairing, as in (202). The enzyme tolerates templates of 1–3 nucleotides in length. (**E**) General RNA-templated RNA polymerization as in (D), using an exogenous Primer•Template (PT) duplex, catalyzed by the "Round-18" polymerase (203).

called “continuous *in vitro* evolution,” in which all steps of the *in vitro* selection process are simultaneously performed in a single reaction vessel (204). By juxtaposing transcription, ribozyme-catalyzed ligation (which in this format appends the ribozyme with a T7 RNA polymerase promoter), and reverse-transcription, the survival of any given RNA sequence from one generation to the next requires that it catalyze its reaction before having its structure disrupted by reverse transcription. This scheme allows hundreds of rounds of selection to be performed in relatively little time, providing a facile route to visualizing population dynamics *in vitro*. Moreover, although kinetic competition with reverse polymerase serves as the sole selective pressure in its simplest context, continuous evolution has been used to subject the ligase to the additional constraints of limiting magnesium (205), altered divalent metal ion compositions (206), extremes of pH (207), and competition with a ribonucleolytic DNAzyme “predator” (208) or with a different ligase species also under continuous selection (209). Furthermore, the Class I ligase has been co-opted for direct selection of multiple-turnover in an *in vitro* compartmentalization experiment (210), for computer-controlled selection on a microfluidic chip (211) and for zeptomole detection of Hepatitis C genomic RNA *in vivo* (212).

The Class I Ligase: An RNA Polymerase Ribozyme

Given the impressive catalytic efficiency of the Class I ligase and the similarity between its reaction and that of proteinaceous polymerases, this ribozyme represented an ideal starting point from which to build a general RNA polymerase ribozyme. To test whether the ligase is itself capable of templated primer extension, a construct was designed in which the features analogous to the template and NTP in the self-ligase were removed (**Figure 4D**) (202). By

providing a template-primer pair and NTPs *in trans*, the sequence dependence with which the ligase recognized its template, primer and incoming NTP could be systematically interrogated. Paired with a template coding for a single addition and provided with the complementary NTP, the ligase was able to extend its primer by one nucleotide. This could be generalized for templates of any sequence, although reaction rates differed according to the identity of the template nucleotide. However, in each case primer extension was selective for Watson-Crick pairing between the template and incoming NTP, even when the enzyme was challenged with a mixture of all four NTPs. Based on k_{cat}/K_M measurements for every possible template•NTP pair, the average fidelity of single nucleotide primer-extension was ~85%, a substantial improvement over the best selectivity gleaned from Group I intron-derived polymerase constructs (188). Moreover, by using a longer template, the Class I ribozyme was able to extend a primer by three additions during a 24-hour incubation, achieving an average Watson-Crick fidelity of 88% among the tested templates. The total number of additions increases to six if the template-primer pair is designed to bind the ribozyme core in multiple sequence registers. As these reactions showed little bias for the sequence of the primer used, the ligase hence embodied all of the requisite features of a general RNA polymerase ribozyme (202).

In-Depth Characterization of the Ligase

Several lines of experimentation sought to further characterize the ligase catalytic mechanism and compare it to that of natural polymerases. Proteinaceous polymerases are thought to use a universally conserved catalytic mechanism employing a pair of divalent metal ions liganded by active site aspartic acids (213). One of these metals, "Metal A," activates the 3′-hydroxyl nucleophile for attack and, in cohort with the second metal, "Metal B," stabilizes the

geometry and charge distribution of the transition state. Recent work has implicated that an active site general acid and general base may further contribute to this mechanism (214); the base receives a proton from the activated nucleophile, while the acid protonates the pyrophosphate leaving group. Though sequence identity of residues responsible for these functions is poorly conserved, the acid has been identified and confirmed biochemically (215).

A screen of various divalent metal ions revealed that Class I self-ligation was only supported in the presence of Mg^{2+} or, to a much lesser extent, Mn^{2+}; all other metals were inhibitory (140). These preferences parallel those of proteinaceous polymerases (216), though Mn^{2+} is typically more highly tolerated. Based on the Hill coefficients of Mg^{2+} binding, two to five Mg^{2+} ions were estimated as being critically required for folding and catalysis. In addition, hydroxyl-radical probing, which serves as readout of tertiary structure by assaying the solvent accessibilities of each residue, revealed that the ligase adopts a nearly identical structure if reconstituted in Mg^{2+} or Ca^{2+}. This observation provided the foundation of a method for measuring self-ligation rates under conditions where folding is rate limiting. By pre-folding the ligase in Ca^{2+} and initiating the time course by simultaneous addition of EGTA (a Ca^{2+}-specific chelator) and Mg^{2+}, the catalytic rate could be measured irrespective of the rate of folding. A key result from these studies is the observation that self-ligation rates follow a pH dependence that is log-linear (slope of 1.0) between 5.8 and 8.5 (140). This is different from the bell-shaped pH dependencies exhibited by proteinaceous polymerases (215) but resembles that of the natural ribozyme RNase P (217), for which the nucleophile is a metal-activated hydroxide ion.

An attempt was also made at modeling the three-dimensional architecture of the ligase. Using the hydroxyl-radical probing data mentioned above to identify residues occluded from solvent, and a chemical crosslink between helices P5 and P7 (**Figure 4A**) as a topological

constraint, the orientation and topology of the seven helical stems was modeled computationally (200). In the resulting model, the ligase architecture is comprised of three domains built from helices P1-P2, P3-P6-P7 and P4-P5. These domains are compactly bundled parallel to one another, with P2 abutting P3 and P4, and P1 abutting P5 and P6. While a predicted Watson-Crick base pair between G46 and C113 was experimentally confirmed, no other explicit geometric constraints could be applied to the long joining regions J1/3 and J3/4.

Meanwhile, biochemical experiments continued to reveal similarities and differences between polymerases built from RNA and protein. Using variants (**Figure 4C**) of the primer-extending construct developed in (202), contacts made to the incoming NTP were probed using a series of chemically modified nucleotide derivatives (201). Phosphorothiolate substitutions on the incoming NTP α-phosphate resulted in a loss of activity that is dependant on the stereochemistry of the sulfur substitution, suggesting that the NTP α-phosphate pro-*Rp* phosphate oxygen, but not the pro-*Sp,* is potentially a direct metal ion ligand. Modifications to other positions on the incoming NTP revealed interactions responsible for its binding. While Watson-Crick pairing provided the strongest energetic contribution to nucleotide recognition, minor contributions from contacts made to the 2′-hydroxyl, γ-phosphate and purine N7 groups were also observed. That the 2′-hydroxyl provided such a modest contribution to substrate binding is in stark contrast with proteinaceous polymerases, but could be justified as a residual effect from the original *in vitro* selection process. Unlike natural polymerases, the ligase had never been challenged with 2′-deoxy substrates during its selection, and hence there is no selective advantage in avoiding them.

Such residual effects of the *in vitro* selection process were also observed in a study using longer substrates. During the process of converting the self-ligase to the primer-extending

enzyme, four nucleotides had been trimmed from the 5′-end of the ligase core. Restoring truncations of this sequence in the form of 5′-triphosphorylated oligonucleotide substrates improves the k_{cat}/K_M of primer extension reactions by up to ~1300-fold, as compared to reactions with GTP. However, this added efficiency was strongly dependent on the new sequence added to the substrate, those deviating from the wild type being worse than GTP alone. This implies that the wild type substrates are more efficient because they restore electrophile-positioning interactions selected for in the original self-ligation context, but which had been ablated in the primer-extending enzyme. In contrast, proteinaceous polymerases efficiently bind and position an incoming NTP using only the interactions made to that NTP.

Ligase-Derived Polymerase Ribozymes

Though the Class I ligase and its primer-extending derivative resemble natural RNA polymerases in multiple regards, properties remnant from their original selection preclude them from functioning as general RNA replicases. Most notable of these is the manner in which they bind their templates. In its original self-ligation context, the ligase is covalently linked to its template through J1/3 (**Figure 4A**). But, even when this linkage is severed, as in the primer-extending enzyme shown in (**Figure 4C**), template binding is accomplished through Watson-Crick pairing to helix P2, a structural feature required for folding. The consequences of this are twofold. First, since the ligase can only bind templates with a programmed sequence at their 3′-ends, the available sequence space of viable templates is substantially limited. Second, binding the template by pairing with P2 limits the length of extension products the ligase can polymerize. Since the template also provides one strand of the P2 duplex, a transcript could never be extended past the point where it intrudes into P2. Indeed, the longest non-repetitive sequence the

ligase core could extend is three nucleotides (202). Having reached the limit of what could be accomplished through design alone, further *in vitro* selection would be necessary to overcome these restrictions.

Such efforts have proven extremely fruitful. Seminal work by Johnston *et al*. (203) harnessed the Class I ligase catalytic machinery in an *in vitro* selection experiment aimed at isolating general polymerase ribozymes. Since binding an exogenous primer-template (PT) duplex through means other than base pairing was assumed to be an activity far from the ligase's grasp, this experiment sought to augment the ligase catalytic machinery by evolving a second PT-binding domain. A starting pool of $>10^{15}$ sequence complexity was built by appending the 3′-end of a partially randomized ligase core (severed from its PT duplex as in (202)) with a randomized tail of 76 nucleotides. In addition, the 5′-end of the pool was appended by a long leader sequence terminating in a 5′-5′ linkage to a short primer. Selection was based on the ability to extend this primer with several 4-thio-uridine or biotinylated adenosine residues in a template directed fashion, hence facilitating the enrichment of active sequences without requiring that they modify themselves directly. While this method does not explicitly require that enzymes bind their PT duplexes exogenously *per se*, it was hoped that tethering these groups by a flexible linker would approximate bimolecular binding. Further efforts were also taken to avoid selecting ribozymes that bound their PT duplexes in a sequence dependent manner.

Ten rounds of selection produced a single isolate with observable polymerase activity, the so-called "round-10" ribozyme. This isolate was used as the basis for a new pool of $>10^{15}$ partially randomized sequence variants, which was subjected to eight additional rounds of more stringent selection. This gave rise to 22 sequences with measurable polymerase activity, of

which the most active was coined the "round-18" ribozyme (**Figure 4E**) and chosen for further analysis.

While its predecessors could only be said to approximate such an activity, the round-18 ribozyme is a *bona fide* general RNA-dependent RNA polymerase. Provided with NTPs, the ribozyme catalyzed primer extension by at least a single nucleotide addition with any of the PT duplexes tested. Much like its natural counterparts, the round-18 polymerase did prefer some templates to others, and with these most efficient templates it extended its primer by up to fourteen nucleotides in a 24-hour incubation. Activity was only modestly affected when either the coding region or the primer-template helix was extended five- to ten-fold, implying that there are likely few steric constraints limiting the PT duplexes. Moreover, cloning and sequencing of its products revealed that the average fidelity achieved by the round-18 ribozyme was substantially improved over its parental ligase. Full primer extension to the end of a coding region proceeded with an average error frequency of 1.1% per residue. A more systematic study of misincorporation rates measured the average fidelity at 96.7%, which could be improved to 98.5% by slightly lowering the GTP concentration. Taken together, these data demonstrated that the round-18 ribozyme could bind a primer-template duplex of any sequence in *trans* and extend the primer in a Watson-Crick dependent manner. Its only apparent limitation is in the length of its products.

Characterization and Improvement of the Polymerase Ribozyme

The ability of the round-18 polymerase to bind and utilize PT-duplexes was further characterized structurally (218) and mechanistically (219). The former study surveyed individual binding contacts by observing polymerization kinetics using a systematic series of modified

substrates bearing a single 2′-deoxyribose on either the primer or the template. Each substrate actually reported the effect incurred by modification in multiple contexts, since the position of a given modification changed relative to the end of the primer each time the primer was extended. Considering this, and repeating the experiments with PT pairs of different sequences, errors due to sequence-specific incorporation differences could be partially alleviated. Overall, the round-18 ribozyme makes only a few contacts to 2′-hydroxyls on its PT substrates, binding at just two positions on either strand of the primer-template duplex and at three downstream positions in the template coding region. Moreover, the magnitude of interference due to each substitution could reasonably be attributed to the loss of a hydrogen bond. While these experiments are by no means exhaustive, having probed only a single type of functional group, they imply that the round-18 ribozyme binds its substrate through a small cadre of relatively weak interactions.

A more mechanistic approach sought to characterize interactions between the polymerase and its substrate by directly measuring binding affinities and processivity (219). However, each of these measurements was hampered by the observation that the ribozyme is inhibited in the presence of excess RNA, limiting both the range over which reagents could be titrated and the available techniques used in measurement. In the case of binding affinity, the PT-duplex K_D could be estimated at ~3mM, based on the magnitude of a burst observed during a prebinding-dilution experiment. In the case of processivity, a clever kinetic modeling technique was developed that used a single template and a series of primers corresponding to polymerization intermediates along that template. Comparing the rates of extension at a given template position when that position is the first or second to be read, processivity could be modeled by optimizing a dissociation probability for each template residue. Although these probabilities were developed by modeling the extension of a primer by only two nucleotides, the resulting model

accurately described polymerization kinetics with longer templates, implying that it is a sufficient descriptor of the system overall.

However, in considering the prospects for building an RNA replicase from the round-18 ribozyme, the results of this study are mixed. On the one hand, polymerization by the round-18 ribozyme is certainly not distributive: the average probability that the enzyme would remain bound to its substrate following a round of addition was ~23%. In the best sequence contexts, this probability approached 90%, conferring a "running start" advantage to subsequent additions over the first. On the other hand, processivity was highly sequence dependent; in certain contexts it was below measurement. Despite the polymerase's many encouraging traits, its transformation into a true replicase ribozyme would require an approximate 100-fold improvement in substrate affinity.

One line of experimentation tried to overcome the limiting K_D by colocalizing the ribozyme and its substrate, thus reducing the number of dimensions in which they diffused (220). To accomplish this, the ribozyme and PT-duplex were tagged with hydrophobic (cholesteryl or stearyl) anchors and assembled into micelles. Reconstitution into these assemblies improved neither the apparent rate of single-addition extension nor in the maximum product length. However, primer extension in this context was more processive, populating fewer intermediates and improving the fraction of fully extended primers tenfold. Still, even this modest improvement was only observed with certain PT-duplexes, and would likely prove insufficient to alleviate the shortcomings of the polymerase.

Two separate efforts turned to *in vitro* selection in hopes of isolating polymerases with properties superior to those of the round-18 ribozyme. The first of these started with the observation that the round-18 polymerase is the product of a population bottleneck, having been

evolved from a partially randomized pool derived from a single active species (203). By returning to a pool from an early stage of the original polymerase selection and applying more stringent selective pressures, the authors hoped to isolate more efficient catalysts that had been missed during the first pass (221). This yielded a total of eight new polymerase ribozymes with accessory domains dramatically different from the round-18 ribozyme. Six of these species were active with all of the PT-duplexes tested, but none of them were as efficient as the round-18 polymerase. In theory, each of these new ribozymes could provide a potentially productive starting point for further design and selection experiments. However, it is also possible that the known isolates embody the catalytic optimum that can be obtained from this pool using the available selection technology.

Inspired by this possibility, further work has thus sought to improve the round-18 ribozyme through selection using an *in vitro* compartmentalization (IVC) methodology. As noted above, the original polymerase selection did not explicitly require that enzymes bind their PT-duplexes *in trans*; in the resulting polymerase ribozymes a PT-binding deficiency limits activity. To address this issue, Zaher and Unrau (222) developed an *in vitro* selection scheme that requires active isolates to bind a PT-duplex in *trans*, similar to methods previously used to select ligase (210) and diels-alderase (223) variants. Briefly, a DNA pool corresponding to $>10^{14}$ sequence variants of the round-18 polymerase was ligated to an RNA PT-duplex. This pool was emulsified in oil droplets along with all the components necessary for T7 transcription and ribozyme activity. Transcripts arising from this pool cannot diffuse between oil droplets, and so active polymerase ribozymes–while not physically coupled to their substrates *per se*–are kept in close proximity to them. Selection is based on the isolation of DNA sequences that are covalently linked to extended RNA primers. This selection yielded a polymerase variant, termed

B6.61, that was superior to its predecessor in every assay performed. It was approximately fourfold faster, exhibited higher fidelity and incorporated fewer G•U wobble mismatches than the round-18 ribozyme. Most impressive, however, is that the B6.61 polymerase could extend its primer by up to 22 additions during a 24-hour incubation, nearly twice as far as the round-18 polymerase. This boost in activity was due to relatively minor sequence changes: a five-nucleotide A-rich expansion near the 5′-terminus, and a point mutation in the accessory domain. Such minor changes might be predictable, as the starting pool was a relatively conservative departure from the parental sequence, incorporating only a 5-nt random domain and mutagenizing at 3-10% elsewhere. It is therefore possible that a more aggressively redesigned pool, coupled with the IVC selection technique, may yield even more highly active polymerase ribozymes or even a genuine replicase.

However, it is likely that designing the most productive route from polymerase to replicase will require a more explicit understanding of the ligase ribozyme that lies at the polymerase core.

Summary of this Study

Since a higher-resolution understanding of catalysis by the Class I ligase ribozyme would likely be of great interest in several fields of study, we have sought to elucidate its structure and mechanism through a combination of x-ray crystallographic and biochemical experiments. Moreover, since artificial catalytic RNAs have not typically been dissected at the level of enzymological scrutiny enjoyed by the natural ribozymes, subtler structure-function analysis of the ligase mechanism may serve as a starting point for exploring the reactions it and other synthetic ribozymes catalyze in greater detail.

We opted to begin our analysis by solving the crystal structure of the self-ligated product species. Focusing on the product ensured that all molecules in our experiments had successfully traversed a transition state, and were hence not covalently defective in some manner. Previous attempts at crystallizing the ligase product had been unsuccessful (N. H. Bergman, unpublished data), but our approach differed in several key regards. First, our efforts were focused on an improved ligase variant that had emerged from an *in vitro* selection experiment aimed at optimizing catalytic efficiency. Compared to its predecessor, this species showed a dramatically reduced dependence on Mg^{2+} for folding and catalysis (224), implying that it might prove a more robust crystallization target. Focusing on this improved ligase was inspired by attempts to use *in vitro* selection to circumvent folding heterogeneities in other RNAs (58); these studies had yielded species with improved thermal stability (91, 225) and crystal quality (45, 226). Finally, we employed the U1A "crystallization module," a technology that had previously been successful in facilitating the crystallization of two recalcitrant RNA species (52, 58).

The crystal structure of the Class I ligase product revealed a wealth of information about its global fold and active site architecture. This initial structure was not only consistent with the existing body of biochemical probing data but has also withstood the scrutiny of further biochemical experiments (224). This structure was furthermore shown to be reproducible in different crystalline environments and has proven a predictive model for the experiments described in the present work. Overall, the ribozyme takes the form of a tripod, its legs built from three coaxially stacked domains, P1-P2, P3-P6-P7 and P4-P5, which converge in the vicinity of the ligation junction. While the topology and relative orientation of these domains were accurately predicted by previous molecular modeling experiments, (200), the novel tripodal form was unanticipated. Surprisingly, three longest joining regions, J1/2, J1/3 and J3/4, do not

mediate long-range tertiary contacts, as is often observed in large RNA structures, but rather decorate the exterior of the tripod scaffold. Although most of the interactions within these regions are new examples of previously characterized motifs, two incarnations of a novel four-nucleotide motif were observed in J1/3. This was termed the A-minor triad, but is likely to be quite rare in nature: in known RNA crystal structures, only one other example was observed.

Beyond the novelty of its overall architecture and the new motif discovered therein, this first class I ligase structure identified the functional groups comprising the active site. The ligation junction is placed at the point of convergence between the three helical domains, opposite the gap between the two A-minor triads. It is in close proximity to functional groups on three universally conserved residues: the C47 N4 and the backbone phosphates of A29 and C30. Each of these groups was probed directly: mutation of C47 to uridine decreased the self-ligation rate by a factor of greater than 10^4, while oxygen-to-sulfur modifications on A29 or C30 showed interference values among the highest measured for any position in the ribozyme. Given these and other biochemical data, it was proposed that during catalysis, the C47 N4 hydrogen bonds to the leaving group, thereby stabilizing its developing negative charge, while the A29 and C30 phosphates directly ligand an active site Mg^{2+} ion. This metal would activate the 3′ hydroxyl for nucleophilic attack, in a role analogous to "Metal A" in proteinaceous polymerases and "Metal 1" in the group I intron. However, this initial assignment was speculative. Since no electron density was observed for a metal ion in the active site, it was argued that a high-affinity metal site might have been lost during the course of catalysis or crystallization. Given this possibility, the model for transition state stabilization by the ligase implied by these data resembles that of natural proteinaceous polymerases. This work represents a substantial advancement in our understanding of the structure and catalytic mechanism of this ribozyme.

Because the Class I ligase is an obligate single turnover catalyst in its self-ligation format, the RNA species in the initial product structure was inert, lacking some the functionalities in a complete active site. To further explore the structural basis of ligase catalysis, crystal structures of chemically trapped, unreacted ligase•substrate complexes were solved. Though several routes to trapping this complex were theoretically available, high quality crystals were only obtained from samples in which the C47U mutant ligase was reconstituted and crystallized in Ca^{2+}. In isolation, neither the active site mutation nor the change in divalent metal composition ablates folding, though each alteration diminishes self-ligation activity by a factor of 10^4-10^5. Furthermore, since Ca^{2+}-grown crystals of the C47U mutant could be rapidly re-equilibrated in Mg^{2+} prior to freezing, this approach facilitated not only the visualization of several intact ligase active sites, but also a structural understanding of the Mg^{2+}-dependence of the ligase. Overall, the folds of Ca^{2+}-reconstituted and Mg^{2+}-soaked structures were extremely similar to the product structure, a result that corroborates previous chemical probing experiments performed under both conditions. However, the P1-P2 domain, which bears structures analogous to the template, primer and incoming NTP used by the polymerase, is structurally perturbed in Ca^{2+}-bound crystal. Since these structural pathologies were relieved in Mg^{2+}-soaked crystals, the metal ion preference of the ligase may be the explained by its inability to productively position the template and electrophile in the absence of magnesium ions.

Structures of the unreacted ligase•substrate complexes revealed features of the active site that were absent in the original product structure: the proposed divalent metal ion bound by A29 and C30, as well as the 5′-triphosphate. An inner-sphere contact between the active site metal ion and the primer 3′-hydroxyl pulls the reaction nucleophile closer to C30 than in the product structure, positioning it with nearly ideal geometry for in-line attack on the G1 α-phosphate. The

G1 β- and γ-phosphates deviate dramatically from the extended conformation seen in proteinaceous polymerase•substrate complexes, instead making an unusual hooked structure that docks into the minor groove of the P2 helix. While numerous metal ion contacts made to these phosphates seem responsible for stabilizing this conformation, none of them occupy a position similar to the canonical "Metal B" in proteinaceous polymerases. The only functional groups near that location are the C47 N4 and C30 2′-hydroxyl.

A series of biochemical experiments were performed to test the functional significance of these structural observations. C47 was altered to a series of natural and modified bases, facilitating assignment between functional groups in the cytidine (or, in the case of the mutant, uridine) ring and the presumed roles played during catalysis. Parallel chemical alteration studies of the C30 2′-hydroxyl revealed that this group functions by donating either a proton or a hydrogen bond and that its ability to do so is thermodynamically coupled to the chemical state at residue 47. Solvent kinetic isotope experiments revealed that the wild type and chemically altered ligase species most likely transfer a single proton during the transition state of its rate-determining step. Taken together, these experiments augment the previous mechanistic model predicted from the product structure, and demonstrate the first example of an artificial ribozyme that employs an active site nucleobase.

Overall, the experiments described here provide a detailed structural and mechanistic model of catalysis by the Class I ligase ribozyme, and begin to establish a concrete understanding of the strategies that underlie its remarkable efficiency. Results from these studies will be invaluable in future efforts to design or select RNA polymerase ribozymes. Moreover, they demonstrate that artificial ribozymes can employ subtle catalytic strategies similar to their

natural counterparts. Efforts to further explore the mechanistic basis of artificial ribozyme catalysis may facilitate the generation of ribozymes with more complex and robust activities.

References

1. K. Kruger *et al.* (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell.* **31**, 147.
2. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell.* **35**, 849.
3. C. Guerrier-Takada, S. Altman. (1984) Catalytic activity of an RNA molecule prepared by transcription in vitro. *Science.* **223**, 285.
4. A. J. Zaug, T. R. Cech. (1986) The intervening sequence RNA of Tetrahymena is an enzyme. *Science.* **231**, 470.
5. C. Woese. (1967) The molecular basis for genetic expression. In *The genetic code*. pp. 179-195. Harper and Row, New York.
6. L. E. Orgel. (1968) Evolution of the genetic apparatus. *J Mol Biol.* **38**, 381.
7. F. H. Crick. (1968) The origin of the genetic code. *J Mol Biol.* **38**, 367.
8. W. Gilbert. (1986) The RNA World. *Nature.* **319**, 618.
9. G. F. Joyce, L. E. Orgel. (1999) Prospects for Understanding the Origin of the RNA World. In *The RNA World, Second Edition,* R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds., pp. 49-77. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10. L. E. Orgel. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol.* **39**, 99.
11. D. P. Bartel. (1999) Re-creating an RNA Replicase. In *The RNA World, Second Edition,* R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds., pp. 143-162. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
12. G. F. Joyce. (1989) RNA evolution and the origins of life. *Nature.* **338**, 217.
13. P. A. Sharp. (1985) On the origin of RNA splicing and introns. *Cell.* **42**, 397.
14. H. B. White, 3rd. (1976) Coenzymes as fossils of an earlier metabolic state. *J Mol Evol.* **7**, 101.
15. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science.* **289**, 920.
16. T. A. Steitz, P. B. Moore. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci.* **28**, 411.

17. W. Saenger. (1984) In *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.
18. D. S. Wilson, J. W. Szostak. (1999) In vitro selection of functional nucleic acids. *Annu Rev Biochem.* **68**, 611.
19. D. P. Bartel, P. J. Unrau. (1999) Constructing an RNA world. *Trends Cell Biol.* **9**, M9.
20. U. F. Müller. (2006) Re-creating an RNA world. *Cell Mol Life Sci.* **63**, 1278.
21. M. J. Fedor, J. R. Williamson. (2005) The catalytic diversity of RNAs. *Nat Rev Mol Cell Biol.* **6**, 399.
22. A. C. Forster, R. H. Symons. (1987) Self-cleavage of virusoid RNA is performed by the proposed 55-nucleotide active site. *Cell.* **50**, 9.
23. O. C. Uhlenbeck. (1987) A small catalytic oligoribonucleotide. *Nature.* **328**, 596.
24. J. M. Buzayan, Gerlachm W.L., G. Bruening. (1986) Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA. *Nature.* **323**, 349.
25. L. Sharmeen, M. Y. Kuo, G. Dinter-Gottlieb, J. Taylor. (1988) Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol.* **62**, 2674.
26. B. J. Saville, R. A. Collins. (1990) A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell.* **61**, 685.
27. W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins, R. R. Breaker. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature.* **428**, 281.
28. C. L. Peebles *et al.* (1986) A self-splicing RNA excises an intron lariat. *Cell.* **44**, 213.
29. A. V. Kazantsev *et al.* (2005) Crystal structure of a bacterial ribonuclease P RNA. *Proc Natl Acad Sci U S A.* **102**, 13392.
30. S. Niranjanakumari *et al.* (2007) Probing the architecture of the B. subtilis RNase P holoenzyme active site by cross-linking and affinity cleavage. *RNA.* **13**, 521.
31. J. Lipfert, J. Ouellet, D. G. Norman, S. Doniach, D. M. Lilley. (2008) The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure.* **16**, 1357.
32. T. A. Steitz, J. A. Steitz. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A.* **90**, 6498.
33. A. M. Pyle. (1993) Ribozymes: a distinct class of metalloenzymes. *Science.* **261**, 709.
34. M. Yarus. (1993) How many catalytic RNAs? Ions and the Cheshire cat conjecture. *FASEB J.* **7**, 31.

35. J. R. Knowles. (1991) Enzyme catalysis: not different, just better. *Nature.* **350**, 121.
36. W. P. Jencks. (1969) In *Catalysis in Chemistry and Enzymology*. Dover Publications, Inc., New York.
37. R. B. Silverman. (2000) In *The Organic Chemistry of Enzyme-Catalyzed Reactions*. Academic Press, San Diego, CA.
38. J. A. Doudna, J. R. Lorsch. (2005) Ribozyme catalysis: not different, just worse. *Nat Struct Mol Biol.* **12**, 395.
39. M. J. Fedor. (2009) Comparative enzymology and structural biology of RNA self-cleavage. *Annu Rev Biophys.* **38**, 271.
40. J. L. Hansen, T. M. Schmeing, P. B. Moore, T. A. Steitz. (2002) Structural insights into peptide bond formation. *Proc Natl Acad Sci U S A.* **99**, 11670.
41. T. M. Schmeing *et al.* (2002) A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nat Struct Biol.* **9**, 225.
42. M. Selmer *et al.* (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science.* **313**, 1935.
43. P. L. Adams, M. R. Stahley, A. B. Kosek, J. Wang, S. A. Strobel. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature.* **430**, 45.
44. B. L. Golden, H. Kim, E. Chase. (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol.* **12**, 82.
45. F. Guo, A. R. Gooding, T. R. Cech. (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell.* **16**, 351.
46. S. V. Lipchock, S. A. Strobel. (2008) A relaxed active site after exon ligation by the group I intron. *Proc Natl Acad Sci U S A.* **105**, 5699.
47. H. W. Pley, K. M. Flaherty, D. B. McKay. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature.* **372**, 68.
48. W. G. Scott, J. T. Finch, A. Klug. (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell.* **81**, 991.
49. M. Martick, W. G. Scott. (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell.* **126**, 309.
50. Y. I. Chi *et al.* (2008) Capturing hammerhead ribozyme structures in action by modulating general base catalysis. *PLoS Biol.* **6**, e234.

51. L. Pauling. (1946) Molecular Architecture and Biological Reactions. *Chem Eng News.* **24**, 1375.

52. P. B. Rupert, A. R. Ferre-D'Amare. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature.* **410**, 780.

53. P. B. Rupert, A. P. Massey, S. T. Sigurdsson, A. R. Ferre-D'Amare. (2002) Transition state stabilization by a catalytic RNA. *Science.* **298**, 1421.

54. S. Alam, V. Grum-Tokars, J. Krucinska, M. L. Kundracik, J. E. Wedekind. (2005) Conformational heterogeneity at position U37 of an all-RNA hairpin ribozyme with implications for metal binding and the catalytic structure of the S-turn. *Biochemistry.* **44**, 14396.

55. J. Salter, J. Krucinska, S. Alam, V. Grum-Tokars, J. E. Wedekind. (2006) Water in the active site of an all-RNA hairpin ribozyme and effects of Gua8 base variants on the geometry of phosphoryl transfer. *Biochemistry.* **45**, 686.

56. A. T. Torelli, J. Krucinska, J. E. Wedekind. (2007) A comparison of vanadate to a 2'-5' linkage at the active site of a small ribozyme suggests a role for water in transition-state stabilization. *RNA.* **13**, 1052.

57. C. MacElrevey, J. D. Salter, J. Krucinska, J. E. Wedekind. (2008) Structural effects of nucleobase variations at key active site residue Ade38 in the hairpin ribozyme. *RNA.* **14**, 1600.

58. A. R. Ferre-D'Amare, K. Zhou, J. A. Doudna. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature.* **395**, 567.

59. A. Ke, K. Zhou, F. Ding, J. H. Cate, J. A. Doudna. (2004) A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature.* **429**, 201.

60. D. J. Klein, A. R. Ferre-D'Amare. (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science.* **313**, 1752.

61. J. C. Cochrane, S. V. Lipchock, S. A. Strobel. (2007) Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. *Chem Biol.* **14**, 97.

62. D. J. Klein, M. D. Been, A. R. Ferre-D'Amare. (2007) Essential role of an active-site guanine in glmS ribozyme catalysis. *J Am Chem Soc.* **129**, 14858.

63. P. C. Bevilacqua, T. S. Brown, S. Nakano, R. Yajima. (2004) Catalytic roles for proton transfer and protonation in ribozymes. *Biopolymers.* **73**, 90.

64. T. J. McCarthy *et al.* (2005) Ligand requirements for glmS ribozyme self-cleavage. *Chem Biol.* **12**, 1221.

65. S. Nakano, D. M. Chadalavada, P. C. Bevilacqua. (2000) General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science.* **287**, 1493.

66. J. H. Chen, B. Gong, P. C. Bevilacqua, P. R. Carey, B. L. Golden. (2009) A catalytic metal ion interacts with the cleavage Site G.U wobble in the HDV ribozyme. *Biochemistry.* **48**, 1498.

67. T. M. Schmeing, K. S. Huang, S. A. Strobel, T. A. Steitz. (2005) An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature.* **438**, 520.

68. W. L. Delano. (2002) The PyMOL Molecular Graphics System. *http://pymol.sourceforge.net*. DeLano Scientific.

69. M. R. Stahley, S. A. Strobel. (2005) Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science.* **309**, 1587.

70. N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle. (2008) Crystal structure of a self-spliced group II intron. *Science.* **320**, 77.

71. N. Toor, K. Rajashankar, K. S. Keating, A. M. Pyle. (2008) Structural basis for exon recognition by a group II intron. *Nat Struct Mol Biol.* **15**, 1221.

72. M. Martick, T. S. Lee, D. M. York, W. G. Scott. (2008) Solvent structure and hammerhead ribozyme catalysis. *Chem Biol.* **15**, 332.

73. N. Toor *et al.*) Tertiary architecture of the Oceanobacillus iheyensis group II intron. *RNA.* **16**, 57.

74. X. Chen, N. Li, A. D. Ellington. (2007) Ribozyme catalysis of metabolism in the RNA world. *Chem Biodivers.* **4**, 633.

75. A. D. Ellington, J. W. Szostak. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature.* **346**, 818.

76. C. Tuerk, L. Gold. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science.* **249**, 505.

77. A. A. Beaudry, G. F. Joyce. (1992) Directed evolution of an RNA enzyme. *Science.* **257**, 635.

78. R. R. Breaker, G. F. Joyce. (1994) A DNA enzyme that cleaves RNA. *Chem Biol.* **1**, 223.
79. L. C. Mattheakis, R. R. Bhatt, W. J. Dower. (1994) An in vitro polysome display system for identifying ligands from very large peptide libraries. *Proc Natl Acad Sci U S A.* **91**, 9022.
80. R. W. Roberts, J. W. Szostak. (1997) RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc Natl Acad Sci U S A.* **94**, 12297.
81. S. Baskerville, D. P. Bartel. (2002) A ribozyme that ligates RNA to protein. *Proc Natl Acad Sci U S A.* **99**, 9154.
82. C. Merryman, E. Weinstein, S. F. Wnuk, D. P. Bartel. (2002) A bifunctional tRNA for in vitro selection. *Chem Biol.* **9**, 741.
83. D. P. Bartel, M. L. Zapp, M. R. Green, J. W. Szostak. (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell.* **67**, 529.
84. J. C. Cox, A. D. Ellington. (2001) Automated selection of anti-protein aptamers. *Bioorg Med Chem.* **9**, 2525.
85. M. Sassanfar, J. W. Szostak. (1993) An RNA motif that binds ATP. *Nature.* **364**, 550.
86. J. M. Carothers, S. C. Oestreich, J. H. Davis, J. W. Szostak. (2004) Informational complexity and functional activity of RNA structures. *J Am Chem Soc.* **126**, 5130.
87. R. D. Jenison, S. C. Gill, A. Pardi, B. Polisky. (1994) High-resolution molecular discrimination by RNA. *Science.* **263**, 1425.
88. R. Green, J. W. Szostak. (1992) Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science.* **258**, 1910.
89. A. Berzal-Herranz, S. Joseph, J. M. Burke. (1992) In vitro selection of active hairpin ribozymes by sequential RNA-catalyzed cleavage and ligation reactions. *Genes Dev.* **6**, 129.
90. N. K. Vaish, P. A. Heaton, O. Fedorova, F. Eckstein. (1998) In vitro selection of a purine nucleotide-specific hammerheadlike ribozyme. *Proc Natl Acad Sci U S A.* **95**, 2158.
91. K. Juneau, T. R. Cech. (1999) In vitro selection of RNAs with increased tertiary structure stability. *RNA.* **5**, 1119.
92. K. H. Link, L. Guo, R. R. Breaker. (2006) Examination of the structural and functional versatility of glmS ribozymes by using in vitro selection. *Nucleic Acids Res.* **34**, 4968.

93. D. P. Bartel, J. W. Szostak. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science.* **261**, 1411.
94. T. Pan, O. C. Uhlenbeck. (1992) In vitro selection of RNAs that undergo autolytic cleavage with Pb2+. *Biochemistry.* **31**, 3887.
95. K. P. Williams, S. Ciafre, G. P. Tocchini-Valentini. (1995) Selection of novel Mg(2+)-dependent self-cleaving ribozymes. *EMBO J.* **14**, 4551.
96. V. K. Jayasena, L. Gold. (1997) In vitro selection of self-cleaving RNAs with a low pH optimum. *Proc Natl Acad Sci U S A.* **94**, 10612.
97. J. Tang, R. R. Breaker. (2000) Structural diversity of self-cleaving ribozymes. *Proc Natl Acad Sci U S A.* **97**, 5784.
98. F. Huang, M. Yarus. (1997) A calcium-metalloribozyme with autodecapping and pyrophosphatase activities. *Biochemistry.* **36**, 14107.
99. J. R. Lorsch, J. W. Szostak. (1994) In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature.* **371**, 31.
100. R. K. Kumar, M. Yarus. (2001) RNA-catalyzed amino acid activation. *Biochemistry.* **40**, 6998.
101. F. Huang, C. W. Bugg, M. Yarus. (2000) RNA-Catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry.* **39**, 15548.
102. F. Huang, M. Yarus. (1997) 5'-RNA self-capping from guanosine diphosphate. *Biochemistry.* **36**, 6557.
103. H. S. Zaher, R. A. Watkins, P. J. Unrau. (2006) Two independently selected capping ribozymes share similar substrate requirements. *RNA.* **12**, 1949.
104. K. B. Chapman, J. W. Szostak. (1995) Isolation of a ribozyme with 5'-5' ligase activity. *Chem Biol.* **2**, 325.
105. E. H. Ekland, J. W. Szostak, D. P. Bartel. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science.* **269**, 364.
106. M. P. Robertson, A. D. Ellington. (1999) In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol.* **17**, 62.

107. L. F. Landweber, I. D. Pokrovskaya. (1999) Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: the spandrels of RNA evolution. *Proc Natl Acad Sci U S A.* **96**, 173.

108. J. Rogers, G. F. Joyce. (2001) The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA.* **7**, 395.

109. L. Jaeger, M. C. Wright, G. F. Joyce. (1999) A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc Natl Acad Sci U S A.* **96**, 14712.

110. Y. Ikawa, K. Tsuda, S. Matsumura, T. Inoue. (2004) De novo synthesis and development of an RNA enzyme. *Proc Natl Acad Sci U S A.* **101**, 13750.

111. T. Tuschl, P. A. Sharp, D. P. Bartel. (2001) A ribozyme selected from variants of U6 snRNA promotes 2',5'-branch formation. *RNA.* **7**, 29.

112. Y. Wang, S. K. Silverman. (2003) Deoxyribozymes that synthesize branched and lariat RNA. *J Am Chem Soc.* **125**, 6880.

113. J. R. Prudent, T. Uno, P. G. Schultz. (1994) Expanding the scope of RNA catalysis. *Science.* **264**, 1924.

114. M. M. Conn, J. R. Prudent, P. G. Schultz. (1996) Porphyrin Metalation Catalyzed by a Small RNA Molecule. *J Am Chem Soc.* **118**, 7012.

115. S. J. Pollack, J. W. Jacobs, P. G. Schultz. (1986) Selective chemical catalysis by an antibody. *Science.* **234**, 1570.

116. C. Wilson, J. W. Szostak. (1995) In vitro evolution of a self-alkylating ribozyme. *Nature.* **374**, 777.

117. M. Illangasekare, G. Sanchez, T. Nickles, M. Yarus. (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. *Science.* **267**, 643.

118. P. A. Lohse, J. W. Szostak. (1996) Ribozyme-catalysed amino-acid transfer reactions. *Nature.* **381**, 442.

119. A. Jenne, M. Famulok. (1998) A novel ribozyme with ester transferase activity. *Chem Biol.* **5**, 23.

120. T. W. Wiegand, R. C. Janssen, B. E. Eaton. (1997) Selection of RNA amide synthases. *Chem Biol.* **4**, 675.

121. B. Zhang, T. R. Cech. (1997) Peptide bond formation by in vitro selected ribozymes. *Nature.* **390**, 96.

122. L. Sun, Z. Cui, R. L. Gottlieb, B. Zhang. (2002) A selected ribozyme catalyzing diverse dipeptide synthesis. *Chem Biol.* **9**, 619.

123. M. Illangasekare, M. Yarus. (1999) A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis. *RNA.* **5**, 1482.

124. T. M. Coleman, F. Huang. (2002) RNA-catalyzed thioester synthesis. *Chem Biol.* **9**, 1227.

125. G. Sengle, A. Eisenfuhr, P. S. Arora, J. S. Nowick, M. Famulok. (2001) Novel RNA catalysts for the Michael reaction. *Chem Biol.* **8**, 459.

126. T. M. Tarasow, S. L. Tarasow, B. E. Eaton. (1997) RNA-catalysed carbon-carbon bond formation. *Nature.* **389**, 54.

127. B. Seelig, A. Jaschke. (1999) A small catalytic RNA motif with Diels-Alderase activity. *Chem Biol.* **6**, 167.

128. S. Fusz, A. Eisenfuhr, S. G. Srivatsan, A. Heckel, M. Famulok. (2005) A ribozyme for the aldol reaction. *Chem Biol.* **12**, 941.

129. Y. Ryu, K. J. Kim, C. A. Roessner, A. I. Scott. (2006) Decarboxylative Claisen condensation catalyzed by in vitro selected ribozymes. *Chem Commun (Camb).* 1439.

130. M. W. Lau, P. J. Unrau. (2009) A promiscuous ribozyme promotes nucleotide synthesis in addition to ribose chemistry. *Chem Biol.* **16**, 815.

131. P. J. Unrau, D. P. Bartel. (1998) RNA-catalysed nucleotide synthesis. *Nature.* **395**, 260.

132. M. W. Lau, K. E. Cadieux, P. J. Unrau. (2004) Isolation of fast purine nucleotide synthase ribozymes. *J Am Chem Soc.* **126**, 15686.

133. L. A. Gugliotti, D. L. Feldheim, B. E. Eaton. (2004) RNA-mediated metal-metal bond formation in the synthesis of hexagonal palladium nanoparticles. *Science.* **304**, 850.

134. S. Tsukiji, S. B. Pattnaik, H. Suga. (2003) An alcohol dehydrogenase ribozyme. *Nat Struct Biol.* **10**, 713.

135. H. Saito, D. Kourouklis, H. Suga. (2001) An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO J.* **20**, 1797.

136. H. Suga, J. A. Cowan, J. W. Szostak. (1998) Unusual metal ion catalysis in an acyl-transferase ribozyme. *Biochemistry.* **37**, 10118.

137. A. Flynn-Charlebois, N. Lee, H. Suga. (2001) A single metal ion plays structural and chemical roles in an aminoacyl-transferase ribozyme. *Biochemistry.* **40**, 13623.

138. H. Saito, H. Suga. (2002) Outersphere and innersphere coordinated metal ions in an aminoacyl-tRNA synthetase ribozyme. *Nucleic Acids Res.* **30**, 5151.

139. M. Illangasekare, M. Yarus. (1999) Specific, rapid synthesis of Phe-RNA by RNA. *Proc Natl Acad Sci U S A.* **96**, 5470.

140. M. E. Glasner, N. H. Bergman, D. P. Bartel. (2002) Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry.* **41**, 8103.

141. H. S. Zaher, P. J. Unrau. (2006) A general RNA-capping ribozyme retains stereochemistry during cap exchange. *J Am Chem Soc.* **128**, 13894.

142. L. Sun, Z. Cui, C. Li, S. Huang, B. Zhang. (2007) Ribozyme-catalyzed dipeptide synthesis in monovalent metal ions alone. *Biochemistry.* **46**, 3714.

143. D. Lazarev, I. Puskarz, R. R. Breaker. (2003) Substrate specificity and reaction kinetics of an X-motif ribozyme. *RNA.* **9**, 688.

144. P. J. Unrau, D. P. Bartel. (2003) An oxocarbenium-ion intermediate of a ribozyme reaction indicated by kinetic isotope effects. *Proc Natl Acad Sci U S A.* **100**, 15393.

145. F. Huang, M. Yarus. (1998) Kinetics at a multifunctional RNA active site. *J Mol Biol.* **284**, 255.

146. D. J.-F. Chinnapen, D. Sen. (2006) Towards elucidation of the mechanism of UV1C, a deoxyribozyme with photolyase activity. *J Mol Biol.* **365**, 1326.

147. W. A. Ziehler, D. R. Engelke. (2001) Probing RNA structure with chemical reagents and enzymes. *Curr Protoc Nucleic Acid Chem.* **Chapter 6**, Unit 6 1.

148. D. W. Celander. (2001) Probing RNA structures with hydroxyl radicals. *Curr Protoc Nucleic Acid Chem.* **Chapter 6**, Unit 6 5.

149. J. C. Cochrane, S. A. Strobel. (2004) Probing RNA structure and function by nucleotide analog interference mapping. *Curr Protoc Nucleic Acid Chem.* **Chapter 6**, Unit 6 9.

150. K. A. Wilkinson, E. J. Merino, K. M. Weeks. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc.* **1**, 1610.

151. W. L. Kelly. (2008) Intramolecular cyclizations of polyketide biosynthesis: mining for a "Diels-Alderase"? *Org Biomol Chem.* **6**, 4483.

152. A. Heine *et al.* (1998) An antibody exo Diels-Alderase inhibitor complex at 1.95 angstrom resolution. *Science.* **279**, 1934.

153. F. E. Romesberg, B. Spiller, P. G. Schultz, R. C. Stevens. (1998) Immunological origins of binding and catalysis in a Diels-Alderase antibody. *Science.* **279**, 1929.

154. N. Kisseleva, S. Kraut, A. Jaschke, O. Schiemann. (2007) Characterizing multiple metal ion binding sites within a ribozyme by cadmium-induced EPR silencing. *HFSP J.* **1**, 127.

155. A. Y. Kobitski, A. Nierth, M. Helm, A. Jaschke, G. U. Nienhaus. (2007) Mg2+-dependent folding of a Diels-Alderase ribozyme probed by single-molecule FRET analysis. *Nucleic Acids Res.* **35**, 2047.

156. A. M. Chiorcea-Paquim, J. A. Piedade, R. Wombacher, A. Jaschke, A. M. Oliveira-Brett. (2006) Atomic force microscopy and anodic voltammetry characterization of a 49-mer diels-alderase ribozyme. *Anal Chem.* **78**, 8256.

157. R. Wombacher, A. Jaschke. (2008) Probing the active site of a diels-alderase ribozyme by photoaffinity cross-linking. *J Am Chem Soc.* **130**, 8594.

158. V. Manoharan, B. Furtig, A. Jaschke, H. Schwalbe. (2009) Metal-induced folding of Diels-Alderase ribozymes studied by static and time-resolved NMR spectroscopy. *J Am Chem Soc.* **131**, 6261.

159. A. Serganov *et al.* (2005) Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat Struct Mol Biol.* **12**, 218.

160. H. Xiao, H. Murakami, H. Suga, A. R. Ferre-D'Amare. (2008) Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature.* **454**, 358.

161. M. P. Robertson, W. G. Scott. (2007) The structural basis of ribozyme-catalyzed RNA assembly. *Science.* **315**, 1549.

162. J. N. Pitt, A. R. Ferre-D'Amare. (2009) Structure-guided engineering of the regioselectivity of RNA ligase ribozymes. *J Am Chem Soc.* **131**, 3532.

163. B. Seelig, S. Keiper, F. Stuhlmann, A. Jaschke. (2000) Enantioselective Ribozyme Catalysis of a Bimolecular Cycloaddition Reaction This work was supported by the Deutsche Forschungsgemeinschaft (Grant no.: Ja 794/3-1) and the Bundesministerium fur Bildung und Forschung (Grant no.: BEO 0311861). We thank Dr. S. Klussmann and Dr. S. Vonhoff (Noxxon Pharma AG, Berlin) for the synthesis of the L-ribozyme. *Angew Chem Int Ed Engl.* **39**, 4576.

164. S. Keiper, D. Bebenroth, B. Seelig, E. Westhof, A. Jaschke. (2004) Architecture of a Diels-Alderase ribozyme with a preformed catalytic pocket. *Chem Biol.* **11**, 1217.

165. N. Lee, Y. Bessho, K. Wei, J. W. Szostak, H. Suga. (2000) Ribozyme-catalyzed tRNA aminoacylation. *Nat Struct Biol.* **7**, 28.

166. K. Ramaswamy, H. Saito, H. Murakami, K. Shiba, H. Suga. (2004) Designer ribozymes: programming the tRNA specificity into flexizyme. *J Am Chem Soc.* **126**, 11454.

167. H. Saito, H. Suga. (2001) A ribozyme exclusively aminoacylates the 3'-hydroxyl group of the tRNA terminal adenosine. *J Am Chem Soc.* **123**, 7178.

168. A. Ohta, H. Murakami, E. Higashimura, H. Suga. (2007) Synthesis of polyester by means of genetic code reprogramming. *Chem Biol.* **14**, 1315.

169. Y. Goto, H. Suga. (2009) Translation initiation with initiator tRNA charged with exotic peptides. *J Am Chem Soc.* **131**, 5040.

170. H. Saito, K. Watanabe, H. Suga. (2001) Concurrent molecular recognition of the amino acid and tRNA by a ribozyme. *RNA.* **7**, 1867.

171. J. J. Perona, Y. M. Hou. (2007) Indirect readout of tRNA for aminoacylation. *Biochemistry.* **46**, 10419.

172. M. P. Robertson, J. R. Hesselberth, A. D. Ellington. (2001) Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings. *RNA.* **7**, 513.

173. M. P. Robertson, A. D. Ellington. (2000) Design and optimization of effector-activated ribozyme ligases. *Nucleic Acids Res.* **28**, 1751.

174. M. P. Robertson, A. D. Ellington. (2001) In vitro selection of nucleoprotein enzymes. *Nat Biotechnol.* **19**, 650.

175. M. P. Robertson, S. M. Knudsen, A. D. Ellington. (2004) In vitro selection of ribozymes dependent on peptides for activity. *RNA.* **10**, 114.

176. M. P. Robertson, W. G. Scott. (2008) A general method for phasing novel complex RNA crystal structures without heavy-atom derivatives. *Acta Crystallogr D Biol Crystallogr.* **D64**, 738.

177. N. B. Leontis, E. Westhof. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA.* **7**, 499.

178. R. K. Montange, R. T. Batey. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys.* **37**, 117.

179. A. Serganov, D. J. Patel. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet.* **8**, 776.

180. P. L. Adams *et al.* (2004) Crystal structure of a group I intron splicing intermediate. *RNA.* **10**, 1867.

181. R. R. Breaker *et al.* (2003) A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA.* **9**, 949.

182. S. Raugei, M. Cascella, P. Carloni. (2004) A proficient enzyme: insights on the mechanism of orotidine monophosphate decarboxylase from computer simulations. *J Am Chem Soc.* **126**, 15730.

183. N. R. Pace, T. L. Marsh. (1985) RNA catalysis and the origin of life. *Orig Life Evol Biosph.* **16**, 97.

184. T. R. Cech. (1986) A model for the RNA-catalyzed replication of RNA. *Proc Natl Acad Sci U S A.* **83**, 4360.

185. K. E. McGinness, G. F. Joyce. (2003) In search of an RNA replicase ribozyme. *Chem Biol.* **10**, 5.

186. M. D. Been, T. R. Cech. (1988) RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the Tetrahymena ribozyme. *Science.* **239**, 1412.

187. P. S. Kay, T. Inoue. (1987) Catalysis of splicing-related reactions between dinucleotides by a ribozyme. *Nature.* **327**, 343.

188. D. P. Bartel, J. A. Doudna, N. Usman, J. W. Szostak. (1991) Template-directed primer extension catalyzed by the Tetrahymena ribozyme. *Mol Cell Biol.* **11**, 3390.

189. J. A. Doudna, N. Usman, J. W. Szostak. (1993) Ribozyme-catalyzed primer extension by trinucleotides: a model for the RNA-catalyzed replication of RNA. *Biochemistry.* **32**, 2111.

190. J. A. Doudna, J. W. Szostak. (1989) RNA-catalysed synthesis of complementary-strand RNA. *Nature.* **339**, 519.

191. J. A. Doudna, S. Couture, J. W. Szostak. (1991) A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science.* **251**, 1605.

192. E. J. Hayden, N. Lehman. (2006) Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem Biol.* **13**, 909.

193. Q. Vicens, T. R. Cech. (2009) A natural ribozyme with 3',5' RNA ligase activity. *Nat Chem Biol.* **5**, 97.

194. N. Paul, G. F. Joyce. (2002) Inaugural Article: a self-replicating ligase ribozyme. *Proc Natl Acad Sci U S A.* **99**, 12733.

195. D. E. Kim, G. F. Joyce. (2004) Cross-catalytic replication of an RNA ligase ribozyme. *Chem Biol.* **11**, 1505.

196. T. A. Lincoln, G. F. Joyce. (2009) Self-sustained replication of an RNA enzyme. *Science.* **323**, 1229.

197. E. H. Ekland, D. P. Bartel. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231.

198. N. H. Bergman, W. K. Johnston, D. P. Bartel. (2000) Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry.* **39**, 3115.

199. K. E. McGinness, G. F. Joyce. (2002) RNA-catalyzed RNA ligation on an external RNA template. *Chem Biol.* **9**, 297.

200. N. H. Bergman, N. C. Lau, V. Lehnert, E. Westhof, D. P. Bartel. (2004) The three-dimensional architecture of the class I ligase ribozyme. *RNA.* **10**, 176.

201. M. E. Glasner, C. C. Yen, E. H. Ekland, D. P. Bartel. (2000) Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. *Biochemistry.* **39**, 15556.

202. E. H. Ekland, D. P. Bartel. (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature.* **382**, 373.

203. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science.* **292**, 1319.

204. M. C. Wright, G. F. Joyce. (1997) Continuous in vitro evolution of catalytic function. *Science.* **276**, 614.

205. T. Schmitt, N. Lehman. (1999) Non-unity molecular heritability demonstrated by continuous evolution in vitro. *Chem Biol.* **6**, 857.

206. C. A. Riley, N. Lehman. (2003) Expanded divalent metal-ion tolerance of evolved ligase ribozymes. *Biochimie.* **85**, 683.

207. H. Kühne, G. F. Joyce. (2003) Continuous in vitro evolution of ribozymes that operate under conditions of extreme pH. *J Mol Evol.* **57**, 292.

208. P. Ordoukhanian, G. F. Joyce. (1999) A molecular description of the evolution of resistance. *Chem Biol.* **6**, 881.

209. S. B. Voytek, G. F. Joyce. (2009) Niche partitioning in the coevolution of 2 distinct RNA enzymes. *Proc Natl Acad Sci U S A.* **106**, 7780.

210. M. Levy, K. E. Griswold, A. D. Ellington. (2005) Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *RNA.* **11**, 1555.

211. B. M. Paegel, G. F. Joyce. (2008) Darwinian evolution on a chip. *PLoS Biol.* **6**, e85.

212. N. K. Vaish *et al.* (2003) Zeptomole detection of a viral nucleic acid using a target-activated ribozyme. *RNA.* **9**, 1058.

213. N. Sträter, W. N. Lipscomb, T. Klabunde, B. Krebs. (1996) Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. *Angew. Chem. Int. Ed. Engl.* **35**, 2024.

214. C. Castro *et al.* (2007) Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc Natl Acad Sci U S A.* **104**, 4267.

215. C. Castro *et al.* (2009) Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol.* **16**, 212.

216. A. S. Mildvan, L. A. Loeb. (1979) The role of metal ions in the mechanisms of DNA and RNA polymerases. *CRC Crit Rev Biochem.* **6**, 219.

217. D. Smith, N. R. Pace. (1993) Multiple magnesium ions in the ribonuclease P reaction mechanism. *Biochemistry.* **32**, 5273.

218. U. F. Müller, D. P. Bartel. (2003) Substrate 2'-hydroxyl groups required for ribozyme-catalyzed polymerization. *Chem Biol.* **10**, 799.

219. M. S. Lawrence, D. P. Bartel. (2003) Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry.* **42**, 8748.

220. U. F. Müller, D. P. Bartel. (2008) Improved polymerase ribozyme efficiency on hydrophobic assemblies. *RNA.* **14**, 552.

221. M. S. Lawrence, D. P. Bartel. (2005) New ligase-derived RNA polymerase ribozymes. *RNA.* **11**, 1173.

222. H. S. Zaher, P. J. Unrau. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA.* **13**, 1017.

223. J. J. Agresti, B. T. Kelly, A. Jaschke, A. D. Griffiths. (2005) Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization. *Proc Natl Acad Sci U S A.* **102**, 16170.

224. S. C. Bagby, N. H. Bergman, D. M. Shechner, C. Yen, D. P. Bartel. (2009) A class I ligase ribozyme with reduced Mg2+ dependence: Selection, sequence analysis, and identification of functional tertiary interactions. *RNA.* **15**, 2129.

225. F. Guo, T. R. Cech. (2002) Evolution of Tetrahymena ribozyme mutants with increased structural stability. *Nat Struct Biol.* **9**, 855.

226. K. Juneau, E. Podell, D. J. Harrington, T. R. Cech. (2001) Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA--solvent interactions. *Structure.* **9**, 221.

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

The work presented here was done in collaboration with Robert A. Grant, Sarah C. Bagby of MIT, and with Yelena Koldobskaya of Joseph Piccirilli's laboratory in the Chemistry Department of the University of Chicago. Robert educated me in all aspects of the crystallographic process, and directly assisted in data collection, and phase determination. Sarah performed chemical interference experiments on the ligase. Using RNA I prepared, Yelena selected and characterized a synthetic antibody, and solved a crystal structure of this antibody in complex with the ligase. All other biochemical and crystallographic results are my own.

X-ray data were collected with the Northeastern Collaborative Access Team (NE-CAT) at the Advanced Photo Source (APS), Argonne National Laboratory.

Abstract

Primordial organisms of the putative RNA world would have required polymerase ribozymes able to replicate RNA. In known polymerase ribozymes with activity best approximating that needed for RNA replication, the catalytic machinery for templated phosphodiester bond formation is derived from an artificial ribozyme, the class I RNA ligase. Here we present the 3.0-angstrom crystal structure of this ligase. The architecture comprises three coaxially stacked domains that converge near the ligation junction. These domains are brought together by the constraints of the enzyme's secondary structure and a small cadre of additional interactions, such that the ribozyme resembles a tripod. Interacting with this tripod scaffold through a series of 10 minor-groove interactions (including two A-minor triads) is the unpaired segment that recognizes Watson-Crick base pairs in the primer-template duplex and contributes to the active site. Absolutely conserved residues form a pair of structural motifs that enclose the active site, positioning two phosphates and the exocyclic amine of a cytidine base within close proximity of the ligation junction. Biochemical and structural data imply a model for transition state stabilization similar to that used by natural protein enzymes, possibly augmented by the additional participation of a catalytic nucleobase. This structure provides insight into the potential mechanisms by which an RNA molecule might decode genetic information and catalyze its own replication in a simple primordial organism.

Introduction

The RNA world hypothesis proposes that early life forms lacked both protein enzymes and DNA genes, and depended instead on RNA for both chemical catalysis and information storage (1). Supporting this hypothesis is the observation that many fundamental, conserved biological processes depend upon RNA, or its metabolic precursors or derivatives (1-4). Perhaps its most compelling aspect, however, is that enzymes made of RNA would have been much easier to replicate than those made of protein. In contemporary organisms, replication of a single protein enzyme is a complex process requiring dozens of macromolecular components (including mRNA, tRNA, aminoacyl-tRNA sythetases, and the ribosome). As this schema of gene expression employs information storing and catalytic biopolymers that are each rendered useless without contributions from the other, it is difficult to conceive of a way in which such a system could have spontaneously emerged from prebiotic materials. However, replication of an RNA enzyme requires only a single macromolecular entity: an RNA-directed RNA polymerase (RdRP) that synthesizes first a reverse complement, and then a copy, of the ribozyme (5).

Because RNA enzymes (ribozymes) with RNA replicase activity have not been found in extant biology, efforts to explore the ability of RNA to catalyze this central activity of the RNA world have focused on engineered derivatives of natural (6-9) and on entirely artificial ribozymes (5). A simple system based on such artificial ribozymes has been engineered to perform a kind of sustainable self-replication (10). However, given both its use of oligonucleotides rather than mononucleoside triphosphates (NTPs) as substrates, and the substantial limitations in the sequence space of these substrates, this system doesn't achieve the kind of general RNA replicase activity that would have been vital in the RNA world.

To date the RNA species that best approximate this general self-replicase activity contain at their core an artificial ribozyme known as the class I RNA ligase (11-14), which was selected *in vitro* from a large pool of random sequences (15). It catalyzes regiospecific attack by the 3′-hydroxyl of an oligonucleotide substrate on the triphosphate at its own 5′ terminus, yielding a 3′-5′ phosphodiester linkage with concomitant release of pyrophosphate, while requiring both the nucleophile and the electrophile to be positioned by Watson-Crick pairing (Figure 1A). Because this reaction chemistry is identical to that catalyzed by proteinaceous enzymes that replicate RNA, the ligase has provided a productive starting point for developing more sophisticated RNA enzymes that use NTPs and the information from an external RNA template to synthesize short strands of RNA (12-14, 16). Although more efficient with some templates than with others, this primer-extension reaction is general in that all templates tested support detectable extension (12-14), in some cases with modest processivity (17). The ligase has also served as a flexible system for studying the evolutionary process itself *in vitro*, having been subsequently selected under numerous diverse constraints, including continuous selection (18, 19). To understand the structural basis behind RNA-catalyzed RNA polymerization, we have solved the crystal structure of the class I ligase ribozyme.

Results and Discussion

A ligase variant suitable for crystallization

The ligase variant we crystallized was the product of three successive *in vitro* selection experiments. After initial isolation of the class I ligase from random sequences (11, 15), more active variants were isolated from a large pool of mutant ligases (20). A composite ligase that combined features of the more active variants from this second selection produced only low-resolution crystals. Therefore, we performed a third selection experiment, targeting for mutagenesis those segments not participating in known base pairs (termed "joining regions" because they link together paired regions) and selecting variants that folded and reacted within milliseconds (21). A particularly active ligase from this third experiment crystallized in three forms, one of which diffracted to 3.0Å resolution. This variant is more tolerant of low Mg^{2+} concentrations; it reacts 15 times faster than the predecessor in 1 mM Mg^{2+} (21) but only slightly faster than the predecessor in high Mg^{2+} [predecessor reaction rate, 800/min in 60 mM Mg^{2+}, pH 9 (22)]. As with the predecessor, its reaction is pH-dependent, slowing to 2.2/min in our crystallization conditions (10 mM Mg^{2+}, pH 6.0). In addition to producing a ligase suitable for crystallization, the last two selection experiments generated diverse but active sequence variants suitable for analyses of nucleotide conservation and covariation, which revealed residues and pairing interactions presumed critical for efficient ligase activity. Seven paired regions, designated P1 through P7, comprise most of the ligase secondary structure (Figure 1A), with additional conserved residues falling in the joining regions.

Structure determination and overview

Crystallization focused on the self-ligation product, obtained by purifying those molecules that ligated the substrate oligonucleotide to their 5´ terminus. To facilitate lattice formation, a dispensable loop (20) was replaced with the U1A-binding module, and the ligase was co-crystallized with the U1A protein (Figure 1B) (23). Crystals of this ligase-U1A complex formed in space-group P1, which contained two nearly identical copies in the asymmetric unit (all-atom ‡RMSD = 0.92Å, excluding the U1A protein and loop). The structure was solved by single isomorphous replacement with anomalous scattering (SIRAS), using iridium- and cobalthexamine to obtain heavy-atom and pseudo-native datasets, respectively (24), and refined to R_{work} and R_{free} values of 20.2 and 23.4%, respectively (Table 1). The final model included all nucleotides in each monomer, although some of the first nine residues in J1/3 (the joining region linking P1 to P3) were poorly ordered (Figures 2, 3).

A parallel effort used a phage-display system to generate antibodies for cocrystallization (25), which yielded crystals with data to 3.1 Å. The ligase structure in this second crystal form, solved independently, confirmed the structure presented here (all-atom RMSD = 1.48 Å) (Figure 4). Details of this structure will be published elsewhere.

The class I ligase structure consists of three coaxially stacked domains: P1-P2, P3-P6-P7 and P4-P5 (Fig. 1B, C), the overall topology and relative orientation of w'hich are consistent with a model previously derived from chemical-probing data (26). That model bundles the three domains nearly parallel to one another, but the crystal structure revealed them to be placed at relative angles of 58–71°, converging near the ligation junction so as to resemble a tripod (Fig. 1D). The domains appear to be positioned by the constraints of the secondary structure, aided by contacts to extruded nucleotides and residues joining the paired regions. Residues abutting the

active site are occluded from solvent, in good correspondence with positions predicted by free-radical mapping to lie in the ribozyme interior (26) (Figure 5). Overall, however, because the legs of the tripod each protrude into solvent, the amount of surface area the ligase occludes from solvent is toward the low end of the distribution observed in other structured RNAs, more closely resembling that of tRNA and other smaller (60-100 nt) natural species, rather than that of RNAs in this size range (~115-175 nt) (Figure 6).

J3/4 bridges all three domains

Positioning the three helical domains are tertiary interactions at the top of the tripod (Figures 7A, 8A). G45, the final residue of J3/4, contacts each of the helical domains, perhaps helping to orient them. This residue abuts P4 and stacks with U76, the residue at the heart of the P3-P6 pseudoknot (Figures 7A, 8A). This interaction pulls the 5′ strands of P4 and P6 close to—and nearly parallel with—the J1/2 joining region, facilitating a contact between a C5 nonbridging oxygen and the 2´-hydroxyl of G45, a group with confirmed function (21). This appears to be the sole direct contact between the P3-P6-P7 and P1-P2 domains. U76, the residue upon which G45 stacks, spans the short gap between the 5´ strand of P6 and the 3´strand of P3, pulling the termini of these strands into close proximity, with the dual effects of enforcing coaxial stacking between P3 and P6 and inducing a profound kink in the U76 backbone. Residues 75 and 77 could further buttress this kink through the joint backbone coordination of a partially hydrated magnesium ion, aided by a hydrogen bond between the 2´-hydroxyls of U76 and G44 (21).

Preceding the G47:U76 interaction is a Watson-Crick base pair between C40 and G44–the first and fifth nucleotides of J3/4–both of which were previously considered unpaired (Figures 7A, 8A). This unanticipated pair continues the helical axis of P3 and encloses a three-

nucleotide loop (Fig 8A). Disruption and restoration of Watson-Crick pairing demonstrated the importance of the C40:G44 pair for ligase activity (Figure 8B,C). Moreover, the enclosed loop is superimposable with a GNRA tetraloop (Figure 9A), implying that the purpose of these residues is to provide a thermodynamically stable cap for the P3-P6-P7 domain. Indeed, replacement of this sequence with a GAAA or UUUU tetraloop, and expansion of the helix by an additional two base pairs, had negligible effect on the ligation rate (Figure 8C).

The GNRA-tetraloop-like triloop enclosed between C40 and G44 is one of two familiar structural motifs that cap opposite ends of the P3-P6-P7 domain. L7 (U93-A98) forms a canonical "uridine-turn (U-turn)" motif, observed first in the anticodon loop of yeast phenylalanyl tRNA (27) and subsequently in the active site of the hammerhead ribozyme (Figure 9B, (28, 29)). The ubiquity of this motif in diverse RNA species might imply that its role in the ligase is analogous to that of the J3/4 GAA triloop, namely, to provide a short, thermodynamically stable cap at the end of a helical region. This being said, this motif is not required for ligase activity, as the L7 sequence is variable in active ligase and polymerase isolates (11, 12, 20, 21).

Recognition of the primer-template helix

The A-rich J1/3 makes a series of tertiary contacts decorating the minor grooves of helices P1 (corresponding to the primer-template duplex used by the polymerase) and P6, crossing between them in the vicinity of the ligation junction (Figures 1B–C and 7B). Of the 10 minor-groove interactions, eight involve adenosines of J1/3. Adenosines are particularly well suited to make interactions to the minor grove (30), and despite relatively high noise in the electron density map at residues U20–G28, interactions with the P1 minor groove could be

assigned for three well-conserved adenosines in this region. A22 makes a type I A-minor interaction (21, 30) with the second base pair of P1, while A25, A26 and A27 dock into the fourth and fifth base pairs of P1, which corresponds to the primer-template duplex used by the polymerase (Figures 7C and 10). Each of these interactions is essentially blind to the identities of the base pairs in P1: modeling any of the four Watson-Crick pairs at these positions retains at least two hydrogen bonds with the corresponding adenosine in J1/3, and allowing modest flexibility in the J1/3 strand (Figure 3A, C), no Watson-Crick substitution would introduce a steric clash. Of particular note are the hydrogen bonds involving the 2´-hydroxyls of U16 and G–3. In the polymerase, 2´-deoxy substitution is more detrimental at the position analogous to U16 than at any other template residue, and 2´-deoxy substitution at the position analogous to G–3 is among the most detrimental primer substitutions (31). Of particular note is the interaction between the N1 of A25 and the 2´-hydroxyl of U16, which corresponds to the template residue four nucleotides upstream of the polymerization site. Hence A25–A27 make defined, yet sequence-independent, contacts that help explain the ability of the polymerase to utilize primer-template helices of any sequence (12-14).

Of potential interest is the structural homology between interactions made by A22–A25 (near the 5´ end of J1/3) as they dock into P1 and that of the beet western yellow virus (BWYV) ribosomal frameshifting pseudoknot (Figure 11, (32)). U23 is of particular interest, as the analogous residue in the viral pseudoknot is an adenosine, yet both are able to share a network of hydrogen bonds with the following cytidine residue and the 2´-hydroxyl of the docking nucleotide (Figure 11B). Supporting the isosteric nature of these interactions, an adenosine at position 23 is frequently observed among highly active isolates from our most recent ligase

selection (21), and improved polymerase variants (14) show a sequence bias for adenosines in the region analogous to the 5´ end of J1/3.

A novel structural motif in J1/3

At the other end of J1/3, A31, A32 and A33 form a kinked structure in which their backbone sweeps out a ~90° angle as they dock into the P6 minor groove, passing from one helical strand to the other through a succession of hydrogen bonds identical to that of A25–A27 (Figures 7B and C). Both sets of adenosines are followed by a base that bears a major-groove carbonyl: G28 at the 3´ end of the P1 triad, U34 at the 3´ end of the P6 triad. In both cases this fourth base is rotated away from the docking helix, directing its carbonyl downward toward the base triples which preceded it, such that the backbone strand of the motif assumes an ~90° angle. Hence we propose that this recurring four-nucleotide AAA(U/G) structure represents a novel motif, which we term the "A-minor triad" (Figure 12). The two incarnations of this motif found in the ligase are superposable (Figure 12B, all-atom RMSD = 1.29Å, without the 5´-phosphate and 3´-(U/G), RMSD = 0.14Å).

With the exception of a single hydrogen bond (between A26 N1 and G-3 N3, and between A32 N1 and G74 N3), the contacts made by an A-minor triad could form irrespective of the sequence into which it docks. Hence, it is possible that this motif represents a simple structural module in which single-stranded RNA docks into the minor groove of a regular A-form duplex. If so, however, the motif is exceedingly rare. A search through known RNA crystal structures revealed only a single additional instance–nucleotides 607-610 of the 16S rRNA (*T. thermophilus* numbering, Figure 12C). Perfect A-minor triads are observed at this position both in *T. thermophilus* (33) and in *E. coli* (34). Note that in the *T. thermophilus* A-

minor triad, the three adenosines are followed by a guanosine, whereas in *E. coli* this last residue is a uridine. These observations, in addition to the repeated selection of a uridine or guanosine at the end of each motif in active ligase or polymerase ribozyme variants (12, 14, 20, 21), further support the notion that the A-minor triad motif preferentially terminates in a residue bearing a major-groove carbonyl.

The P6 A-minor triad helps form a Mg^{2+}-binding site, which is modeled as making a combination of inner- and outer-sphere contacts with each of the last five nucleotides of J1/3 (Figures 7B, 13A and 13C). Placement of a metal ion at this position was supported by strong (~5.5σ) positive peaks in $|F_{obs}|-|F_{calc}|$ difference Fourier maps during the initial stages of refinement. Direct metal coordination by the A31 and A32 nonbridging (pro-R_P) phosphate oxygens brings these oxygens ~3.1 Å from one another, inducing a 90° kink that positions C30 out of the helical docking register of A31–A33. Outer-sphere contacts involving N7 of A32, N7 and N2 of A33, and O4 of U34 further stabilize this interaction—roles that, in concert with their packing into P6, explain both the absolute conservation of these nucleotides in active ribozyme isolates (12, 14, 20, 21, 35) and the deleterious effects of chemically modifying them (21). The *T. thermophilus* 16S rRNA A-minor triad potentially parallels this metal-binding ability: an A608 nonbridging oxygen binds to a solvent atom of unassigned identity (33), around which the rest of the motif hooks. Hence, the propensity to form a metal-binding site may be a general trait of A-minor triad motifs, though, given the ligase P1 triad, not a universal one.

Tertiary interactions bracketing the active site

Between the two A-minor triads lies the active site (Figures 7B and 13A–C). Forming the "floor" of the active site is A71, an absolutely conserved residue at the center of the four-way

junction linking the P4-P5 and P3-P6-P7 domains (Figure 1A–B). A71 forms an imperfect type I A-minor interaction (30) with C86:G105, the first base pair of P7 (Figures 10 and 13C). Chemical modification of A71, or the 2´-OH of C86, severely affects catalysis (21). The A71 A-minor interaction also influences the structure of J1/3. A29 stacks on top of A71, itself forming a type I A-minor motif with the G72:C85 pair at the base of P6 (Figures 10 and 13A). Tracing the chain backward from its 3´ end, until this point J1/3 packs into the minor groove of the P6, and would be capable of continuing in a similar fashion along P7 were it not for the barrier presented by A71. Stacking between A29 and A71 displaces the J1/3 strand, facilitating its transit from P6 to P1. In so doing, the A71 barrier, in concert with the metal-stabilized backbone kink, encloses and buttresses the presumed active site centered on residues A29, C30, and C47 (Figure 13C).

The consequences of bracketing A29 and C30 between these motifs are twofold. First, C30 is extruded from the minor groove of P6, the rotation of its base constrained by A73 so as to form a cross-strand stack with C47 (Figure 13A–C). C47 is likewise extruded from helix P4, the plane of its base roughly perpendicular to the adjoining base pairs, with its N4 exocyclic is positioned 3.1Å away from the ligation junction. Second, the A71 interaction prevents G28 from stacking below A29. The consequent rotation of G28 places the phosphates of A29 and C30 in close proximity to one another (~5 Å between phosphorus centers) and facing P1. The nonbridging oxygens at these residues are 4.5 to 5.3 Å from the 3′-hydroxyl and phosphate oxygens of the ligation junction.

We therefore propose that C47, as positioned by C30, and the backbone phosphates of A29 and C30 compose the ligase active site. Both cytidines are conserved among active isolates (12, 14, 20, 21, 35), although their contributions to activity differ (Figure 14). The C30U

substitution decreases activity by a factor of five, perhaps from disrupting the hydrogen bond between the C30 N4 and the C47 O4′ (Figure 13B), whereas the C47U substitution diminishes activity by a factor of $>10^4$, consistent with a more direct role in catalysis.

With only minor perturbation, the A29 and C30 backbone phosphates could provide a binding site for at least one catalytic magnesium ion as observed at active sites for some natural ribozymes (36, 37) and the L1 ligase, an artificial ribozyme that promotes a reaction resembling that of our ligase (38). Although we observe no electron density corresponding to such a metal in the crystal structure of the ligated product, a high-affinity metal-binding site in the active ribozyme might nonetheless be lost during catalysis. To test for a functional role for these and other backbone phosphate oxygens, we performed a nucleotide-analog interference mapping experiment (NAIM, (21, 39)), randomly incorporating *Rp*-phosphorothioate (oxygen-to-sulfur) substitutions and mapping those that interfered with activity (Figure 13D). As expected, the sites of interference generally corresponded to the observed inner-sphere magnesium-binding sites in the crystal structure (21). For example, substitution of sulfur for non-bridging oxygens at A31 or A32 had among the most dramatic interference effects observed, supporting the idea that inner-sphere contacts at these phosphates create a metal-stabilized kink needed to form the cross-strand C30:C47 stack at the active site. That substitution at A29 or C30 similarly abrogated function supports the hypothesis that other phosphates coordinate at least one magnesium ion that is catalytically critical in the transition state but not bound tightly in the crystallized product, although steric effects due to the larger sulfur atom cannot be excluded.

Interactions with the triphosphate-bearing nucleotide

The electrophile for the self-ligation reaction is the triphosphoguanosine (G1) at the 5´ terminus of the ribozyme. In polymerase ribozymes that perform primer extension reactions, G1 is missing, its role performed instead by free nucleoside triphosphates (NTPs). Also missing are the J1/2 residues tethering G1 to the ribozyme core. The crystal structure revealed highly specific contacts to J1/2, which could help orient G1 in the active site and explain not only the extremely fast rate of self-ligation but also why primer extension with 5´-triphosphorylated oligonucleotides corresponding to truncations of G1–A3 are up to 10^3 fold more efficient than with GTP alone (Figure 15) (21, 40). G2 forms a *trans* sugar edge/Hoogsteen base pair (41) with A11 (Figure 15B, *left*), a nearly universally conserved residue in all active isolates (12-14, 18-21, 35). A3 in turn makes a perfect A-minor motif (30), with G48:C113 (Figure 15B, *middle*). Finally, A4 stacks above another nearly universally conserved residue, A114, which in turn stacks on the G48:C113 pair (Figure 15B, *right*).

Each of the interactions involving residues in J1/2 is supported by biochemical evidence. For the G2 interaction, methylation of A11 N1, which would induce a tautomeric shift in A11 and ablate one of the hydrogen bonds in the G2:A11 pair, results in a loss of self-ligation activity (21). Conversely, in primer extension reactions, restoration of the G2:A11 pair by using a pppGG substrate results in a >30-fold increase in k_{cat}/K_M relative to GTP. Moreover, ablation of this pair with a pppGAAA substrate results in a two-fold loss of k_{cat}/K_M relative to GTP, and a >80-fold loss relative to pppGGAA (40). For the A3:G48:C113 triple, chemical alteration of C113, ether by methylation of N3 or 2´-deoxy substitution, results in a substantial loss of self-ligation activity (21). Conversely, primer extension with pppGGA, which restores the potential for forming the A3:G48:C113 triple in *trans*, is 10^3 more efficient than with GTP alone (40). For

the A4/A114 stack, methylation of A114 N1, which increases the van der Waals profile and hence stacking potential of that base, increases self-ligation activity (21). Primer extension reaction with pppGGAA, which would restore this interaction in *trans*, is >40-fold more efficient than with GTP alone, but is likewise only 3% as efficient as with the pppGGA substrate (40). This might be explained by the fact that A4 does not dock into its position via formation of a base pair or triple, as do G2 and A3, but instead must thread itself through the intersection of the P1-P2 and P4-P5 domains (Figure 15A and 15B, *right*), which may be entropically unfavorable.

RNA polymerases built from protein and RNA

The crystal structure of another ligase ribozyme, the L1, has been solved (38), allowing us to compare multiple active sites–built from both RNA and protein–that accelerate similar reactions. Proteinaceous nucleic acid polymerases employ a universally conserved catalytic mechanism requiring a pair of aspartic acid-liganded divalent metal ions (42, 43), and have been proposed to employ general acid to stabilize the pyrophosphate leaving group (44) (Figure 16A). We aligned the L1 and class I ligase structures to that of a Pol-α family DNA polymerase (45), anchoring this alignment on the last base pair in the primer-template duplex. This juxtaposed the conserved active-site aspartic acids in the protein enzyme with proposed active site phosphates in both ligases (Figure 13E), further supporting the notion that in the ribozymes these moieties may serve to position one or two catalytically critical metal ions. However, while catalytically critical residues in the class I ligase (C47) and DNA polymerase (Lys560, (44)) occupy similar positions in their respective active sites, no analogous functionality is seen in the L1 ligase. This may explain, in part, why the class I is several orders of magnitude more efficient than the L1.

We therefore propose a preliminary model for catalysis by the class I ligase and its polymerase derivatives that resembles the mechanism of proteinaceous enzymes. In our model, the substrate α-phosphate and backbone phosphates of A29 and C30 jointly ligand a catalytic magnesium ion (Figure 16B). This metal activates the primer 3´-hydroxyl for nucleophilic attack and stabilizes the trigonal bipyrimidal transition-state geometry, directly analogous to the canonical "Metal A" in known polymerases (43). In addition, because free nucleotide triphosphates bind divalent cations, we suggest that the G1 triphosphate (or the incoming NTP) enters the active site complexed with a second metal, which, after binding, would remain coordinated by oxygens on the β- and γ-phosphates. At the transition state, this second metal helps stabilize the developing negative charge on the pyrophosphate leaving group. Although we cannot discretely assign moieties on the ribozyme that would aid in its binding, this second metal would stabilize the developing negative charge on the pyrophosphate leaving group, aided by the exocyclic amine of C47, which hydrogen bonds to the (α,β) bridging oxygen.

Our model, which postulates a hydrogen bond to the leaving oxygen in the transition state, differs from that of proteinaceous polymerases, which involves proton transfer to this oxygen in the transition state (44). Although nucleobases can act as general acids at ribozyme active sites (46), we disfavor ascribing such a function to C47. Formally, the C47 N4 exocyclic amine could serve as a general acid catalyst according to two models: (**1**) C47 acts from the neutral form, generating the nucleobase anion following proton transfer from N4, or (**2**) the catalytically active form of C47 is a cation that is protonated at N3. N3 would hence serve as an electron sink, favoring a tautomeric shift to the N4-imino form as a proton is transferred from N4 during catalysis. We consider both of these models unlikely. For Model **1**, the neutral form of C (pK_a >12) is not sufficiently acidic to serve as an effective general acid. Moreover, with

increasing pH, the ribozyme ligation rate increases log-linearly with a slope of 1.0 (pH 5.7 to 8.5), consistent with the net loss of one proton, presumably that of the nucleophile, when proceeding from the ground state to the transition state (22). If general-acid catalysis by C47 were dominant at the transition state, the pH–rate profile would likely deviate from linearity over this pH range. Otherwise, for Model **2**, in order to provide general acid catalysis without influencing the pH-rate profile over the range 5.7-8.5, the ribozyme would have to shift the pKa of the C47 cation by at least four pKa units (from 4 to > 8.5). Finally, if C47 were a general acid, the functional group donating the proton would differ from that of the active-site cytidine of the hepatitis delta virus (HDV) ribozyme, wherein the N3 imine is thought to act as the proton donor (46). For the ligase, methylating N3–which would perturb the pK_a for both N3 and N4, has a negligable effect on catalysis, which rules out Model **2**, and direct participation of N3 but not N4 (21).

Regioselectivity by the Class I ligase ribozyme

Although many RNA ligase ribozymes have been selected *in vitro*, the class I ligase is in the minority of species that employ the substrate 3´-hydroxyl as a nucleophile. The L1 ligase also falls into this class. Given the crystallographic evidence (38), it was hypothesized that this ribozyme achieves regioselectivity via an inhibitory water-mediated contact to the substrate 2´-hydroxyl, which is proposed to quench any alkoxide species that might develop on the 2´-position. Binding this quenching water was proposed to require that the nucleophile be situated within a G•U wobble pair. We observe no analogous interactions in the class I ligase structure; in fact, there is a remarkable dearth of interactions made to the A–1 nucleophile in general. Although we cannot preclude the possibility that the interactions responsible for regioselectivity

have been lost during the course of catalysis due to a local conformation change, our favored model is that selection arises purely as a consequence of Watson-Crick pairing. A–1 and G1 are positioned in canonical base pairs within a continuously stacked A-form duplex, which naturally places the A–1 3´-hydroxyl optimally for in-line attack on the G1 5´-α-phosphate. Reorienting A–1 so as to place its 2´-hydroxyl at the active site, however, would require gross deviation from A-form geometry, which would be energetically disfavored and could perturb the active site.

Recent crystallographic and biochemical work on the class II ligase–which almost exclusively catalyzes the formation of a 2´-5´ phosphodiester linkage (11)–seems to further support the notion that regioselectivity is largely determined by the local sequence context situating the nucelophile and electrophile (47). By examining the L1 and class II crystal structures, the authors were able to alter the regioselectivity of each ribozyme simply by changing the sequence of the base pairs abutting the ligation junction. Interestingly, in the L1 ligase, conversion of the G•U wobble pair orienting the nucleophile to a Watson-Crick pair has little perceptible effect on the regioselectivity, calling into question the proposed wobble-specific water-quenching mechanism. For both ribozymes, the most efficient conversion to the opposite selectivity was observed when mutations at the ligation junction were augmented by mutation of the base pair downstream of the ligation junction, implying that the positioning of the electrophile by adjoining bases plays a role in determining regiospecificity. In the class I ligase, however, sequences downstream of the ligation junction do not play an essential role for regiospecificity. Primer extension with exogenous GTP, which occurs without downstream sequences with which to position the electrophile, maintains selection for the 3´ -hydroxyl nucleophile (16).

Also of note is that the L1 and class II ligases both position their electrophiles via a *trans* sugar edge/Hoogsteen A•G base pair, not via Watson-Crick pairing. This is identical to the G2•A11 pair observed in the class I ligase (Figure 15B, *left*). It is interesting to consider that, had the register by which J1/2 docks into P2 been offset by one base–aligning A11 with G1 rather than with G2–the class I ligase might have emerged as a 2′-5′ selective enzyme as well.

Conclusion

By identifying the residues at the active site, the ligase crystal structure we report here will facilitate directed examination of the catalytic mechanism of RNA-catalyzed RNA polymerization. Our model also provides insights into how known polymerase ribozymes recognize primer-template duplexes, the feature most in need of improvement for developing a self-replicating polymerase ribozyme (14, 17). This binding activity can now be targeted more explicitly in future design and selection experiments.

Acknowledgements

We thank A. Ferré-D'Amaré, J. Doudna, R. Batey, and K. Zhou for the generous donation of reagents and protocols; K. Rajashankar for assistance in phasing; E. Duguid, V. Malashkevich, C. Drennan, T. Schwartz, U. RajBhandary, J. Cochrane, M. Stahley, S. Strobel, M. Robertson, and W. Scott for advice, encouragement, and helpful discussions; and members of the Bartel laboratory for support and comments.

Materials and Methods

U1A Expression and Purification

The U1A A1-98 Y31H/Q36R double mutant (23, 48) was expressed from the plasmid p11U1ADb. To construct this plasmid, the protein-coding sequence from a vector provided by Adrian Ferré-D'Amaré and Jennifer Doudna was inserted into the pET11a (Invitrogen) plasmid, thereby generating a vector that expressed more efficiently than its predecessor. For protein expression, a saturated culture of *E. coli* BL21(DE3) cells transformed with p11U1ADb was diluted 1:1000 into LB supplemented with 100 μg/mL ampicillin and grown in baffled flasks at 37º C with vigorous shaking. Protein overexpression was induced by the addition of 0.5 mM IPTG when cultures reached an OD_{600} of 0.7. Growth was continued at 37º C in the same fashion for another 2.5 hours, and cells were harvested by centrifugation and stored at –80º C before use. Pellets from a two-liter culture were resuspended in 40 mL of the lysis buffer described in (49) and lysed by two passages through a French press at 4º C. Thereafter, purification was performed essentially as in (23), following a detailed protocol provided by Kaihong Zhou of Jennifer Doudna's laboratory. Column fractions from the final purification step were assayed individually for the presence of contaminating ribonuclease by overnight incubation with ^{32}P body-labeled RNA and visualized by gel electrophoresis and phosphorimaging. Fractions lacking nuclease were pooled, dialyzed into storage buffer (23), concentrated to 15 g/L and stored at –80º C before use.

Crystallization Constructs

The crystals producing the native dataset were of an improved class I ligase (21) (Figure 1A), modified at the end of P5 with four additional base pairs terminating in the U1A-binding

loop (Figure 1B). DNA representing the unligated form of this ribozyme was subcloned into pUC19 (New England Biolabs) under a T7 transcription promoter, followed by a genomic hepatitis delta virus (HDV) self-cleaving ribozyme (50) and an *EarI* restriction site, yielding the plasmid p307HU. The HDV sequence cleaves itself from the ribozyme 3′ terminus, thereby producing a homogenous ribozyme 3′ end. The relevant DNA sequence of the insert was

GCGTAATACGACTCACTATA**G**GAACACTATACTACTGGATAATCAAAGACAAATCTGCCCG
AAGGGCTTGAGAACATACCC*ATTGCACTCCG*GGTATGCAGAGGTGGCAGCCTCCGGTGGGTTAA
AACCCAACGTTCTCAACAATAGTG**A**GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGG
GCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACCCAC

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

A second construct, pH307HP, was used to transcribe RNA representing the post-ligation product of the ribozyme. To ensure a homogenous 5′ terminus with the 5′-hydroxyl resembling that of the synthetic substrate, the transcript began with a hammerhead (HH) self-cleaving ribozyme (50), which excises itself from the ribozyme. The relevant DNA sequence of the insert was

GCGTAATACGACTCACTATAGGGAGATTCCTACTGGACTGATGAGTCCGTGAGGACGAAAC
GGTACCCGGTACCGTC**T**CCAGTAGGAACACTATACTACTGGATAATCAAAGACAAATCTGCCCG
AAGGGCTTGAGAACATACCC*ATTGCACTCC*GGGTATGCAGAGGTGGCAGCCTCCGGTGGGTTAA
AACCCAACGTTCTCAACAATAGTG**A**GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGG
GCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACCCAC

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

RNA Synthesis and Purification.

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

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

from unreacted ribozyme (130 nt), material was loaded onto multiple preparative 6% polyacrylamide TBE/urea gels (same dimensions as above, 0.5 mg RNA/gel), and subjected to 12-15 hours of electrophoresis. Product was visualized and purified as for the unligated RNA. Following the final precipitation, pellets were air-dried to remove residual ethanol, resuspended in minimal water, filtered to remove any particulate matter (0.22μm Centrex filters, Schleicher and Schuell), and stored at –20º C.

RNA used to prepare derivative crystals was transcribed as the ligated product species from the pH307HP plasmid and treated to remove the 2′-3′ cyclic phosphate left by cleavage of the HDV ribozyme. (This material was prepared in the course of exploring an alternate crystal form for which crystallization was sensitive to the state at the 3′ terminus.) Following HDV processing, RNA was desalted, diluted to 10 μM in water and heat/cooled (85º C and 37º C, 5 minutes each). The cyclic phosphate was removed by T4 polynucleotide kinase (PNK) treatment, using "method iii" described in (51), except that RNA was diluted to a final concentration of 5 μM, and PNK (New England Biolabs) was at 0.5 U/μL. After six hours at 37º C, the reaction was quenched with EDTA, extracted with phenol and chloroform, and ethanol precipitated. The pellet was resuspended in water, and material from up to 5mL of initial transcription was separated on a single preparative-scale denaturing 6% polyacrylamide TBE/urea gel. After electrophoresis, purification was identical to that described above for ligated p307HU-transcribed material.

To inspect for full removal of the 2′-3′ cyclic phosphate, we developed an assay appropriate for larger RNAs, which cannot be evaluated using the gel-mobility-shift assays useful for shorter RNA species. Following T4 PNK treatment, 25 pmol of RNA was desalted and incubated at 2.5 μM final concentration with 750 U of Yeast Poly(A) Polymerase (USB), in

the manufacturer's supplied buffer, supplemented with 1 mM ATP. After 30 minutes at 37º C, reactions were quenched with EDTA, extracted with phenol, diluted with urea, boiled and separated on a diagnostic-scale 6% polyacrylamide TBE/urea gel. RNA molecules terminating in a *cis*-diol were extended by poly(A) addition to produce longer species, whereas RNA not treated with PNK did not shift (Figure 17A). As expected, this assay was sensitive to the covalent state of both the 2′ and 3′ hydroxyls: opening the cyclic phosphate by acid-catalyzed hydrolysis (resulting in a mixture of 2′- and 3′ linear mono-phosphates (52)) produced RNAs that were unsuitable substrates for poly(A) extension, whereas further alkaline-phosphatase treatment (Roche) to remove the linear phosphate (52) allowed poly(A) extension of nearly all material (Figure 17B).

Crystallization

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

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

often with an intractable habit (inseparable clusters or stacks of plates). Only ~1% experiments performed in this fashion yielded usable crystals.

A systematic screen of additive compounds (Hampton Research) revealed that addition of 50 mM KCl to the crystallization mixture slightly diminished the degree of crystal clustering, whereas 100 mM KCl ablated nucleation altogether. We therefore exploited these properties, using microseeding to obtain morphologically tractable, diffraction-quality crystals more reproducibly. Seed-producing crystal clusters were grown by mixing 0.5 μL Ligase:U1A complex with an equal volume of 22–26% (v/v) MPD, 50 mM sodium cacodylate (pH 6.0), 40 mM magnesium acetate, 50 mM KCl and 1 mM spermine, and equilibrating over 0.6 mL of this precipitation mixture. Drops containing viable crystals were stabilized by bringing them to 36% MPD, keeping all other buffer components (including those derived from the RNA-annealing and U1A-storage buffers) isotonic. Crystal clusters were crushed using a Seed-Bead kit (Hampton Research), and the resultant seed suspensions were serially diluted in the same stabilization buffer. Diffraction-quality crystals were obtained by using these serially diluted seed stocks directly as precipitant under conditions that would otherwise prohibit crystal nucleation: 0.7 μL of Ligase:U1A complex was prepared in annealing buffer supplemented with 100 mM KCl, mixed with 0.5 μL of the seed-stock dilution, and the drops were equilibrated over 0.6 mL wells of 22–26% MPD. Using this method, roughly one out of every three drops produced at least one high-quality crystal (50–100 μm in all dimensions) within a week. For cryoprotection, drops bearing suitable crystals were brought to 30% (v/v) MPD, again keeping all other buffer compositions isotonic, and incubated against a well of 30% MPD for 2–24 hours before mounting in a nylon loop and plunging directly into liquid nitrogen.

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

Data Collection and Processing

Native and derivative data sets were collected at NE-CAT beamlines 8-BM and 24-ID at the Advanced Photon Source (APS), respectively, the latter aided by the technical advice of K.R. Rajashankar. All data were integrated and scaled using the HKL2000 software suite (53). Initial phases were obtained from a single isomorphous crystal as a "pseudo-native," and an iridium hexamine-soaked crystal as a derivative, similar to the method described in (24) (Table 1). Nineteen initial iridium sites were found using SHELXD, and a preliminary round of solvent flattening was performed in SHELXE (54); both processes were aided by the hkl2map graphical interface (55). The resultant initial electron density map (Figure 2B) contained clear RNA A-form density, as well as conspicuous proteinaceous features from U1A.

Model Building and Refinement

Two monomers corresponding to the U1A protein bound to its cognate loop and a seven base-pair stem (analogous to P5 in the crystallization construct) were located the original experimental density using the program MOLREP (56) in the CCP4 program suite (57), using a search model derived from the hairpin ribozyme-inhibitor crystal structure (58) (PDB ID: 1HP6). Further rounds of solvent flattening were performed in the program SHARP (59), using σA-weighted, combined phases from the initial heavy atom sites and this partial molecular replacement solution, iteratively adjusting the solvent content by hand and inspecting the quality of the resultant electron density map. This yielded modestly improved maps, restoring some phosphate and base-density which had been truncated by the initial SHELXE output.

An initial model was built into this experimental 3.36Å density by real-space refinement in the program COOT (60), starting with 3–4-nucleotide stretches of A-form polycytidine and building into areas with clear RNA-backbone density (38). Each round of poly(C) model building was followed by a round of restrained, individual-site, isotropic Atomic Displacement Parameter (ADP) refinement with automated bulk-solvent correction (termed the "individual ADP" strategy in PHENIX (61)). Refinement was against experimental structure factor amplitudes and phases, using a maximum likelihood target. Given the limited resolution of the data, target geometric weights ("wxc_scale" in PHENIX) were set to a relatively restrictive setting of 0.05 and B-factors were initially refined per-residue. The "wxu_scale" was kept constant at 1.0 during all refinement rounds (61). In early stages, ADP refinement was followed by simulated annealing (5000K to 300K in 100K steps), also implemented in PHENIX. The two monomers were built individually and refined without the use of noncrystallographic symmetry (NCS) averaging. Once ~60% of the RNA sequence had been placed in this fashion,

experimental phases were discarded and ADP refinement with automated bulk solvent correction continued using model-based phases and native amplitudes at 2.99Å, now allowing refinement of individual atomic B-factors. Inspection of the $|F_{obs}|$ - $|F_{calc}|$ difference maps from early rounds of refinement at higher resolution resulted in the unambiguous assignment of individual bases as purine or pyrimidine, and in many cases guanosines could be distinguished from adenosines. This, along with the position of the U1A protein and P5 stem, allowed establishment of the RNA sequence register. In later stages of refinement, NCS averaging was used for residues 20–28, where the electron density map was noisiest, and the resulting B-factors the highest (Figure 3).

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

After placement of all 137 RNA nucleotides for each ligase monomer and all but the first 6-7 protein residues in each U1A protomer, solvent atoms were placed into peaks of 4σ or greater in the resultant $|F_{obs}|$ - $|F_{calc}|$ difference map and further refined; this process of refinement and solvent placement was reiterated until the R_{free} value reached a local minimum. Owing to the resolution limit, magnesium and water atoms could not be individually refined in a chemically sensible manner, so partially- or fully hydrated magnesium clusters were treated as individual monomers during refinement, having defined model residues with ideal bond lengths and geometry. In the final rounds of refinement, the bond distances (2.07Å) and angle (90°) between backbone phosphate nonbridging oxygens at positions A31 and A32 were fixed with respect to the magnesium ion they coordinate (Figures 7B, 13A and 13C); all other clusters were unrestrained relative to RNA or protein. Side-chain atoms for a number of the U1A residues

have been removed in cases for which density in our final $2|F_{obs}| - |F_{calc}|$ maps was apparent for the backbone only. Despite having not enforced global NCS averaging, the two copies of the RNA are nearly identical: excluding residues in the U1A loop, their all-atom RMSD is 0.92Å. Coordinates for the final model have been made available (PDB ID: 3HHN).

Simulated annealing OMIT maps (Figure 13B, D) were calculated in PHENIX, without additional ADP refinement. Simulated annealing composite OMIT maps (Figure 2C) were calculated in CNS (62), with blocks of 5% unit cell volume omitted from each calculation. Though not quoted in table S1, the R_{work} and R_{free} values calculated in CNS are in good agreement with those calculated in PHENIX (23.0 and 25.7%, respectively).

Real space R-values, σA and Luzzati estimated coordinate errors (Table 1) were calculated in CNS. The lengths of potential hydrogen bonds (Table 2) and metal ion contacts (Table 3) were measured without modification of our model. However, in order to measure hydrogen bond angles (Table 2), hydrogens were added automatically using MolProbity (63); where necessary, 2′-hydroxyls were rotated by hand into the optimal hydrogen-bonding orientation. Structural alignments (Figures 4, 9, 11–13) were made using the least-squares all-atom alignment function in COOT (60). All structure figures were made with the program PyMOL (64).

Biochemical Experiments

Mutant ribozymes were generated from the U1A-modified construct, p307HU (QuickChange mutagenesis kit, Stratagene). For kinetic assays, body-labeled RNA was heated and cooled (80° C, 22° C, 5 minutes each) in water, and the reaction was initiated by the simultaneous addition of buffer, salts and an excess of oligonucleotide substrate (final

concentrations, 1 μM ligase, 2 μM substrate, 50 mM Tris, pH 7.0, 10 mM $MgCl_2$, 50 mM KCl, 1.0 μM EDTA). Reactions were incubated at 22°C, and aliquots were removed at time points and quenched by mixing into a two-fold excess of gel loading buffer (8 M urea, 120 mM EDTA, trace xylene cyanol and bromphenol blue). Samples were boiled and separated on 6% or 20% polyacrylamide TBE/urea sequencing gels, imaged and quantified by phosphorimaging (Fujifilm BAS-2500). For each time point, the fraction product was measured as F_P = Product/(Product + Reactant), and fitted to the curve:

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

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

Nucleotide analog interference mapping (NAIM) was performed as described (21, 39), starting with a pool of randomly phosphorothioate-modified RNAs and allowing them to label themselves with ^{32}P-labeled substrate RNA, under solvent and time constraints that selected for only the most active molecules.

Fab-Ligase Crystal Structure

Selection and analysis of the ligase Fab, as well as crystallization and structure solution of it complexed with the ligase shall be presented elsewhere. The coordinates of the ligase complexed with the antibody fragment have been made available (PDB ID: 3IVK).

References

1. G. F. Joyce, L. E. Orgel. (1999) Prospects for Understanding the Origin of the RNA World.In *The RNA World, Second Edition,* R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds., pp. 49-77. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. P. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science.* **289**, 920.
3. L. E. Orgel. (2004) Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol.* **39**, 99.
4. H. B. White, 3rd. (1976) Coenzymes as fossils of an earlier metabolic state. *J Mol Evol.* **7**, 101.
5. D. P. Bartel. (1999) Re-creating an RNA Replicase.In *The RNA World, Second Edition,* R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds., pp. 143-162. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
6. M. D. Been, T. R. Cech. (1988) RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the Tetrahymena ribozyme. *Science.* **239**, 1412.
7. P. S. Kay, T. Inoue. (1987) Catalysis of splicing-related reactions between dinucleotides by a ribozyme. *Nature.* **327**, 343.
8. J. A. Doudna, J. W. Szostak. (1989) RNA-catalysed synthesis of complementary-strand RNA. *Nature.* **339**, 519.
9. R. Green, J. W. Szostak. (1992) Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science.* **258**, 1910.
10. T. A. Lincoln, G. F. Joyce. (2009) Self-sustained replication of an RNA enzyme. *Science.* **323**, 1229.
11. E. H. Ekland, J. W. Szostak, D. P. Bartel. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science.* **269**, 364.
12. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science.* **292**, 1319.
13. M. S. Lawrence, D. P. Bartel. (2005) New ligase-derived RNA polymerase ribozymes. *RNA.* **11**, 1173.

14. H. S. Zaher, P. J. Unrau. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA.* **13**, 1017.
15. D. P. Bartel, J. W. Szostak. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science.* **261**, 1411.
16. E. H. Ekland, D. P. Bartel. (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature.* **382**, 373.
17. M. S. Lawrence, D. P. Bartel. (2003) Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry.* **42**, 8748.
18. B. M. Paegel, G. F. Joyce. (2008) Darwinian evolution on a chip. *PLoS Biol.* **6**, e85.
19. M. C. Wright, G. F. Joyce. (1997) Continuous in vitro evolution of catalytic function. *Science.* **276**, 614.
20. E. H. Ekland, D. P. Bartel. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231.
21. S. C. Bagby, N. H. Bergman, D. M. Shechner, C. Yen, D. P. Bartel. (2009) A class I ligase ribozyme with reduced Mg2+ dependence: Selection, sequence analysis, and identification of functional tertiary interactions. *RNA.* **15**, 2129.
22. M. E. Glasner, N. H. Bergman, D. P. Bartel. (2002) Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry.* **41**, 8103.
23. A. R. Ferré-D'Amaré, J. A. Doudna. (2000) Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J Mol Biol.* **295**, 541.
24. J. C. Cochrane, S. V. Lipchock, S. A. Strobel. (2007) Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. *Chem Biol.* **14**, 97.
25. J. D. Ye *et al.* (2008) Synthetic antibodies for specific recognition and crystallization of structured RNA. *Proc Natl Acad Sci U S A.* **105**, 82.
26. N. H. Bergman, N. C. Lau, V. Lehnert, E. Westhof, D. P. Bartel. (2004) The three-dimensional architecture of the class I ligase ribozyme. *RNA.* **10**, 176.
27. G. J. Quigley, A. Rich. (1976) Structural domains of transfer RNA molecules. *Science.* **194**, 796.

28. M. Martick, W. G. Scott. (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell.* **126**, 309.
29. H. W. Pley, K. M. Flaherty, D. B. McKay. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature.* **372**, 68.
30. P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, T. A. Steitz. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci U S A.* **98**, 4899.
31. U. F. Müller, D. P. Bartel. (2003) Substrate 2'-hydroxyl groups required for ribozyme-catalyzed polymerization. *Chem Biol.* **10**, 799.
32. L. Su, L. Chen, M. Egli, J. M. Berger, A. Rich. (1999) Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat Struct Biol.* **6**, 285.
33. B. T. Wimberly *et al.* (2000) Structure of the 30S ribosomal subunit. *Nature.* **407**, 327.
34. B. S. Schuwirth *et al.* (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science.* **310**, 827.
35. G. F. Joyce. (2004) Directed evolution of nucleic acid enzymes. *Annu Rev Biochem.* **73**, 791.
36. M. R. Stahley, S. A. Strobel. (2005) Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science.* **309**, 1587.
37. N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle. (2008) Crystal structure of a self-spliced group II intron. *Science.* **320**, 77.
38. M. P. Robertson, W. G. Scott. (2007) The structural basis of ribozyme-catalyzed RNA assembly. *Science.* **315**, 1549.
39. S. A. Strobel. (1999) A chemogenetic approach to RNA function/structure analysis. *Curr Opin Struct Biol.* **9**, 346.
40. M. E. Glasner, C. C. Yen, E. H. Ekland, D. P. Bartel. (2000) Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. *Biochemistry.* **39**, 15556.
41. N. B. Leontis, E. Westhof. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA.* **7**, 499.
42. T. A. Steitz. (1998) A mechanism for all polymerases. *Nature.* **391**, 231.

43. N. Sträter, W. N. Lipscomb, T. Klabunde, B. Krebs. (1996) Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. *Angew. Chem. Int. Ed. Engl.* **35**, 2024.
44. C. Castro *et al.* (2009) Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol.* **16**, 212.
45. M. C. Franklin, J. Wang, T. A. Steitz. (2001) Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell.* **105**, 657.
46. P. C. Bevilacqua, R. Yajima. (2006) Nucleobase catalysis in ribozyme mechanism. *Curr Opin Chem Biol.* **10**, 455.
47. J. N. Pitt, A. R. Ferré-D'Amaré. (2009) Structure-guided engineering of the regioselectivity of RNA ligase ribozymes. *J Am Chem Soc.* **131**, 3532.
48. C. Oubridge, N. Ito, C. H. Teo, I. Fearnley, K. Nagai. (1995) Crystallisation of RNA-protein complexes. II. The application of protein engineering for crystallisation of the U1A protein-RNA complex. *J Mol Biol.* **249**, 409.
49. P. Pognonec, H. Kato, H. Sumimoto, M. Kretzschmar, R. G. Roeder. (1991) A quick procedure for purification of functional recombinant proteins over-expressed in E.coli. *Nucleic Acids Res.* **19**, 6650.
50. S. R. Price, N. Ito, C. Oubridge, J. M. Avis, K. Nagai. (1995) Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol.* **249**, 398.
51. H. Schürer, K. Lang, J. Schuster, M. Morl. (2002) A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic Acids Res.* **30**, e56.
52. J. D. Pata, B. R. King, T. A. Steitz. (2002) Assembly, purification and crystallization of an active HIV-1 reverse transcriptase initiation complex. *Nucleic Acids Res.* **30**, 4855.
53. Z. Otwinowski, W. Minor. (1997) Processing of X-ray diffraction data collection in Oscillation Mode. *Methods Enzymol.* **276**, 307.
54. G. M. Sheldrick. (2008) A short history of SHELX. *Acta Crystallogr A.* **64**, 112.
55. T. Pape, T. R. Schneider. (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Cryst.* **37**, 843.
56. A. Vagin, A. Teplyakov. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022.

57. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crysallogr D Biol Crystallogr.* **50**, 760.
58. P. B. Rupert, A. R. Ferré-D'Amaré. (2001) Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature.* **410**, 780.
59. E. DeLaFortelle, G. Bricogne. (1997) Maximum-Likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472.
60. P. Emsley, K. Cowtan. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **60**, 2126.
61. P. D. Adams *et al.* (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr.* **58**, 1948.
62. A. T. Brünger *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr.* **54**, 905.
63. I. W. Davis *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375.
64. W. L. Delano. (2002) The PyMOL Molecular Graphics System. *http://pymol.sourceforge.net.* DeLano Scientific.
65. A. S. Krasilnikov, X. Yang, T. Pan, A. Mondragón. (2003) Crystal structure of the specificity domain of ribonuclease P. *Nature.* **421**, 760.
66. A. R. Ferré-D'Amaré, K. Zhou, J. A. Doudna. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature.* **395**, 567.
67. S. D. Gilbert, R. P. Rambo, D. Van Tyne, R. T. Batey. (2008) Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat Struct Mol Biol.* **15**, 177.
68. C. Lu *et al.* (2008) Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nat Struct Mol Biol.* **15**, 1076.
69. R. T. Batey, S. D. Gilbert, R. K. Montange. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature.* **432**, 411.
70. A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel. (2006) Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature.* **441**, 1167.

71. R. K. Montange, R. T. Batey. (2006) Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature.* **441**, 1172.
72. A. Serganov, L. Huang, D. J. Patel. (2009) Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature.* **458**, 233.
73. C. E. Dann, 3rd *et al.* (2007) Structure and mechanism of a metal-sensing regulatory RNA. *Cell.* **130**, 878.
74. A. Serganov, L. Huang, D. J. Patel. (2008) Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature.* **455**, 1263.
75. M. P. Robertson *et al.* (2005) The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol.* **3**, e5.
76. H. Shi, P. B. Moore. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA.* **6**, 1091.
77. T. Hainzl, S. Huang, A. E. Sauer-Eriksson. (2005) Structural insights into SRP RNA: an induced fit mechanism for SRP assembly. *RNA.* **11**, 1043.
78. J. S. Pfingsten, D. A. Costantino, J. S. Kieft. (2006) Structural basis for ribosome recruitment and manipulation by a viral IRES RNA. *Science.* **314**, 1450.
79. A. S. Krasilnikov, Y. Xiao, T. Pan, A. Mondragón. (2004) Basis for structural diversity in homologous RNAs. *Science.* **306**, 104.
80. J. H. Cate *et al.* (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science.* **273**, 1678.
81. A. Serganov *et al.* (2005) Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat Struct Mol Biol.* **12**, 218.
82. H. Xiao, T. E. Edwards, A. R. Ferré-D'Amaré. (2008) Structural basis for specific, high-affinity tetracycline binding by an in vitro evolved aptamer and artificial riboswitch. *Chem Biol.* **15**, 1125.
83. H. Xiao, H. Murakami, H. Suga, A. R. Ferré-D'Amaré. (2008) Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature.* **454**, 358.
84. S. Rudisser, I. Tinoco, Jr. (2000) Solution structure of Cobalt(III)hexammine complexed to the GAAA tetraloop, and metal-ion binding to G.A mismatches. *J Mol Biol.* **295**, 1211.

Figure Legends

Figure 1. Global architecture of the ligase ribozyme. (**A**) Secondary structure and reaction scheme of a ligase variant with decreased Mg^{2+} dependence (21). It is depicted undergoing ligation, by using the classical secondary-structure representation (26). Red arrows indicate attack by the substrate 3'-hydroxyl on the ribozyme α-phosphate with concomitant loss of pyrophosphate. (**B**) Revised secondary structure of the crystallization construct, reflecting the coaxial stacking and relative domain orientation. Indicated is the ligation junction (thick red dash), backbone phosphates at the active site (yellow dashes), base triples (boxed residues connected with gray lines), and stacking interactions (residues vertically aligned or connected with gray lines terminating in gray bars). Nucleotides numbered as in (A); those in gray were added to facilitate crystallization. Base-pair geometries indicated using nomenclature of (41). (**C**) Ribbon representation of ligase structure, as if peering into the active site (yellow) and ligation junction (red). (**D**) Top-down view, relative to (C).

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

density map, contoured at 1σ; (**C**) a simulated-annealing composite-OMIT map, contoured at 1σ; (**D**) the final refined 3.0Å $2|F_{obs}| - |F_{calc}|$ map, contoured at 1.4σ.

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

Figure 4. Superposition of class I ligase structures independently solved from different crystalline environments. (**A**) Two views of an all-atom alignment between the U1A-bound structure (colored and oriented as in Figure 1C,D) and the 3.15Å structure obtained in complex with a ligase-specific antibody selected through phage-display (gold). Both protein modules bind to L5, which was omitted during the alignment and has been removed from view. Each crystal form contains two copies in its asymmetric unit. Because the four possible pair-wise alignments between each monomer of each crystal form were nearly indistinguishable, an alignment is shown for only one copy from each crystal. Note that the greatest points of divergence between the two models are found in regions involved in crystal-contact formation: P5 (which binds to the protein crystallization modules) and L7 (which is involved in an RNA-protein crystal contact in the U1A-bound structure). (**B**) The relative orientation of the three domains varies slightly between the two crystal forms. The two structures were aligned by the

P1-P2 domain alone, so as to highlight differences in the relative fine positioning of the three helical domains. Each domain is show in isolation: P1–P2 (*left*), P3–P6–P7 (*middle*) and P4–P5 (*right*).

Figure 5. Correlation between calculated and observed solvent accessibilities. (**A**) The calculated solvent-accessible surface area for C4′ carbons (spheres), generated with the program AREAIMOL (57) using a water probe of 1.4Å. These data are projected onto two views of the ligase corresponding approximately to those in Figure 1C and Figure 1D. Positions protected from solvent are in blue; those accessible to solvent are in yellow. (**B**) Plot of the AREAIMOL calculated output (red) compared with the observed solvent protection factors derived from Fe•EDTA hydroxyl radical probing (blue) (26). Both data sets have been plotted as deviations from the mean, in standard deviation units: (value – mean)/σ, as in (65). Fe•EDTA protection factors were multiplied by –1, as positive protection factors imply occlusion from solvent. For the AREAIMOL data, mean = 7.19 $Å^2$, σ = 3.4$Å^2$; for the Fe•EDTA probe data, mean = 1.13, σ = 0.58. Segments spanning residues –7 to 6 and 118 to 121 were not probed by Fe•EDTA. The residues in J1/3 (which interact with the P1 helix in the crystal structure) differ in the construct subjected to Fe•EDTA probing. A lack of correlation in this region may be attributed to this change in sequence, less precise modelling (as suggested by the high temperature factors, Figure 3), or both.

Figure 6. Occlusion of surface area from solvent in known RNA structures. The total solvent-accessible surface area for the present structure (class I ligase, PDB ID: 3HHN) and 23 known RNA structures in the range of 61–174 nt was computed using the program AREAIMOL (57).

The percent surface area buried from solvent exposure (Θ_{OCC}) was calculated by comparing this total surface area (A_{OBS}) by that of a theoretical freely joined RNA chain of equivalent length (A_{FREE}) using the following equation: $\Theta_{OCC} = 100 \times (A_{FREE} - A_{OBS})/A_{FREE}$. For each structure, calculations were performed on a monomer derived from the available PDB file, without bound solvent atoms; for all RNA species using the U1A binding module to facilitate crystallization (including the class I ligase), the protein and its cognate loop were replaced *in silico* with a GNRA tetraloop. A water probe radius of 1.4 Å was used. For comparison, the same calculation was performed on perfect A-form RNAs of the same total length (black diamonds). The data have been subdivided into four classes, as follows. In teal are small, natural self-cleaving ribozymes: hepatitis delta (PDB ID: 1DRZ, (66)), hammerhead (2GOZ, (28)) and hairpin (1HP6, (58)). In gold are the GlmS ribozyme (2NZ4, (24)) and riboswitches: S-adenosylmethionine (SAM) type II (2QWY, (67)), SAM type III (3E5C, (68)), guanosine (1U8D, (69)), thiamine pyrophosphate (2GDI, (70)), SAM type I (2GIS, (71)), flavin mononucleotide (3F4H, (72)), metal ions (2QBZ, (73)) and lysine (3DIL, (74)). In red are other natural structured RNAs or domains: the SARS stem-loop II motif (1XJR, (75)), yeast $tRNA^{Phe}$ (IEHZ, (76)), the SRP S-domain (1Z43, (77)), the *Plautia stali* intestinal virus IRES ribosome-binding domain (2IL9, (78)), B-type (1NBS, (65)) and A-type (1U9S, (79)) RNase P substrate domains and the *T. thermophila* group I intron P4-P6 domain (1GID, (80)). In dark blue are *in vitro*-selected RNAs: the Diels-Alderase ribozyme (1YLS, (81)), the tetracycline aptamer (3EGZ, (82)), the "docked" form of the L1 ligase ribozyme (2O1U, (38)), the aminoacyl synthetase ribozyme ("Flexizyme"), in complex with its substrate (3CUL, (83)), and the class I ligase ribozyme (3HHN). The IRES domain (marked with an asterisk) crystallized as a domain-swapped pseudo dimer with long, flexible, disordered regions between monomers; hence, a portion of surface area that may be

occluded from solvent may have been artificially exposed when we generated a monomer model for solvent accessibility calculations.

Figure 7. Tertiary contacts involving the three longest joining regions. (**A**) Interactions bridging the three domains. (**B**) The path of J1/3. (**C**) Hydrogen bonds of the two A-minor triads (See Figure 12).

Figure 8. Inter-domain contacts involving J3/4. (**A**) Two views of J3/4, showing the triloop enclosed by C40:G44 and highlighting interactions involving G45. For clarity, P2 has been removed from the left view and P4 has been removed from the right view. Dotted black lines denote hydrogen bonds. (**B**) Biochemical support for the role of the C40:G44 Watson-Crick pair. Relative self-ligation rates of ribozymes with the indicated substitutions (yellow) are on the left. (**C**) Expansion and replacement of the J3/4 triloop. Relative self-ligation rates of ribozymes with the indicated substitutions (yellow) are on the right.

Figure 9. The P3-P6-P7 domain is capped by two familiar RNA structural motifs. (**A**) J3/4 forms a GNRA tetraloop-like triloop. C38 and C39 were used to align the last base pairs of P3 with analogous residues in the solution structure of a GAAA tetraloop (PDB ID 1F9L, (84)), shown in gold. Note that, in addition to superposing the C40•G44 base pair, G41 aligns with the analogous residue in the GNRA tetraloop. (**B**) L7 forms a classic uridine-turn (U-turn). U93 and U94 were used to align L7 with 2′-O-Methyl C32 and U33 of the anticodon loop of yeast $tRNA^{Phe}$(PDB ID 1EHZ, (76)), shown in gold. For clarity, the aliphatic group of wybutosine 37 ("YG") has been removed from view.

Figure 10. Base triples observed in the class I ligase structure, each mediated through an adenosine residue. Note that the packing of residues A31 through A33 into adjacent A-U and G-C base pairs is accomplished using the same series of interactions as those employed by A25 through A27 packing into the same sequence. The remaining triples are type I A-minor interactions, although the angle at which A71 packs into C86•G105 precludes it from forming a canonical hydrogen bond with its 2′-hydroxyl (30).

Figure 11. The 5′ end of J1/3 mimics a viral ribosomal frame-shifting pseudoknot. (**A**) All-atom alignment between ligase residues A22–A26 in the 5′ half of J1/3 (colored as in Figure 1C) and the beet western yellow virus (BWYV) ribosomal frame-shifting pseudoknot residues A20-A24 (gold, (32)). P1 of the ligase and the remainder of the pseudoknot structure are shown as ghosted cartoons. Note the divergence of residues immediately abutting the alignment (ligase A21 and pseudoknot G19; ligase A27 and pseudoknot A25). (**B**) Side-by-side comparison between the two motifs (colored as in (A)); the BWYV pseudoknot is on the left, the class I ligase on the right. The first residue of this motif makes a type I A-minor interaction with a base-pair in the receptor helix (Figure 10). The phosphate of the following residue (ligase U23; pseudoknot A21) makes a charge-dipole interaction with the guanosine N2 in the receptor duplex. Ligase residues U23–C24 (pseudoknot residues A21–C22) make joint hydrogen bonds to the 2'-hydroxyl in the docking strand helix, for which U23 uses its exocyclic carbonyl and the pseudoknot A21 uses its N7 and N2. Ligase A25–A26 (pseudoknot residues A23–A24) form the first two base-triples of the A-minor triad motif (Figures 7C, 10 and 12). For clarity, in this

second base-triple only the nucleobase for the receptor guanosine (ligase G–3; pseudoknot G4) is shown.

Figure 12. The "A-minor triad" motif. (**A**) Overview of the path of J1/3, which binds in the minor grooves of P1 and P6. The view is rotated ~60° relative to that in Figure 7B. Residues comprising each of the two iterations of the A-minor triad are highlighted and numbered. (**B**) Wall-eyed stereogram of an all-atom alignment between the two motifs in J1/3. The helices bound by each are ghosted. The 5′-motif (bound to P1) is in gold; the 3′-motif (bound to P6) is in navy. The inner-sphere-coordinated magnesium ion bound by the 3′-motif is in orange. Orange dashes denote inner sphere contacts to this metal; black dashes denote the closest approach between moieties that are proposed to make outer-sphere metal contacts and the metal center (Figures 7B, 13A). For clarity, waters have been removed. Note both the divergence in the positions of the preceding cytidine residue (C24 and C30), and the proximity of the G28 and U34 major-groove carbonyls to one another. (**C**) An A-minor triad in the *T. thermophilus* 16S ribosomal RNA (33). A solvent atom (of unassigned chemistry in the PDB) is shown as an orange sphere; dashes to it are drawn in analogy to those in (B). Note that the first base pair into which this motif docks is a C:G pair (C291:G309), whereas both those observed in the ligase are A:U pairs (A–4:U16 and A73:U84), demonstrating that recognition of this position is independent of the identity of the Watson-Crick base pair.

Figure 13. Architecture of the active site. (**A**) The active site, as viewed from the ligation junction, with P1-P2 removed for clarity. (**B**) Interactions near G1:C12, which is analogous to the NTP-template pair during polymerization (12-14, 16). Meshes are simulated-annealing omit

maps in which active-site nucleotides (gray, contoured at 2σ) or the hydrated metal cluster (aqua, 4σ) were excluded from map calculations. (**C**) Stereograph of the active site. Black dashes indicate hydrogen bonds; magenta dashes indicate proximity between A29 and C30 phosphate oxygens and the ligation junction (red). Mesh represents a simulated-annealing omit map (4.5σ) in which the hydrated metal was excluded from map calculations. (**D**) Mean interference values (±SD) from three α-phosphorothiolate NAIM experiments. The secondary structure is aligned above. Interference values were truncated at the detection limit, 6.0 (21, 39). Missing positions are those modified to facilitate crystallization (hashes) or too close to the termini to measure. (**E**) Models corresponding to pre-ligation complexes for the L1 ligase ribozyme (*left*, (38)) and class I ligase ribozyme (*right*), or prior to primer extension for the RB69 DNA-polymerase (*middle*, (45)), in absence of the reaction electrophile. Coloring is in analogy to (B). Where present, the primer nucleophile is colored red; divalent metal ions are shown as orange spheres. For alignment, the ligase A7:U13 was changed to C:G pair.

Figure 14. Kinetics of C47U and C30U ligase active-site mutants. (**A**) Typical self-ligation time courses of ^{32}P body-labeled wild type, C30U and C47U mutant ligase ribozymes. (**B**) Plots for reaction time courses. Error bars indicate the standard deviation for three replicate experiments.

Figure 15. Interactions with J1/2. (**A**) Two views of J1/2 and its environs, rotated ~180° relative to one another. Nucleotides in the P1-P2 and P4-P5 domains are shown as ghosted cartoons. (**B**) Detail of interactions made with bases in J1/2. Black dashes denote hydrogen bonds. (*Left*) G2 makes a *trans* sugar edge/Hoogsteen base pair with A11 (41). (*Middle*) A3

makes a type I A-minor interaction (30) with G46:C113, the first base pair in P4. (*Right*) A4 stacks on A114, which in turn stacks on the G46:C114 base pair.

Figure 16. Transition-state stabilization by polymerases built from either protein or RNA. (**A**) Catalysis by proteinaceous polymerases (43, 44). Indicated are bonds formed or broken during the transition state (red arrows), coordination of catalytic metal ions, M_A and M_B (blue solid lines), and an active-site acid (A•••H). (**B**) Model for catalysis by the ligase ribozyme. Notation as in (A), with the addition of a hydrogen bond between C47 N4 and the leaving group (dashed gray line). Some magnesium ligands are not specified; for those that are, relative orientations are unknown. A proposed contact to the reactive phosphate pro-R_P oxygen (40) and two speculative contacts implied by NAIM are in blue. Metal ion and coordinations not supported (or refuted) by structural or biochemical evidence are in gray.

Figure 17. The covalent state of the ribozyme 3′ terminus, probed using poly(A) polymerase (PAP). (**A**) Removal of the terminal 2′-3′ cyclic phosphate by T4 polynucleotide kinase (PNK (51)) generates a substrate suitable for extension by PAP. Shown is a denaturing polyacrylamide gel stained with ethidium bromide. Ligase-A_n denotes the 3′-polyadenylated ligase. (**B**) PAP extension requires full removal of the terminal phosphate. Neither untreated RNA bearing a 2′-3′cyclic phosphate (left lanes), nor acid-decyclized RNA bearing a mixture of 2′- and 3′-linear monophosphates (center lanes, (52)) were extended by PAP, whereas subsequent removal of the terminal phosphates by alkaline phosphatase treatment (right lanes) generated a substrate for PAP extension. RNA tested here was ^{32}P body labeled but was otherwise identical to that used in **A**.

Table 1

Crystallographic statistics for the class I ligase product

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

[a]Values in parenthesis refer to the highest resolution shell. [b]5% of all data were excluded from refinement for R_{Free} calculations. [c]Luzzati and σA as calculated in CNS (62); Maximum likelihood as calculated in PHENIX (61). [d]Numbers in parenthesis refer to the cross-validation test set.

Table 2

Lengths and estimated angles for non-solvent-mediated, non-Watson-Crick hydrogen bonds, grouped by secondary structure. All quoted values are the average of observations for each of the two RNA copies in the asymmetric unit; errors report the difference between individual monomer values and the average. Distances are those between hydrogen bond donor and acceptor nuclei. Angles are those centered on the hydrogen atom participating in each bond (*i.e.* the angle: Donor─H•••Acceptor), as modeled in MolProbity (63).

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

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

Table 3

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

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

[a]See Figs. 7B, 13A, 13C. [b]See Fig. 13B.

Figure 1

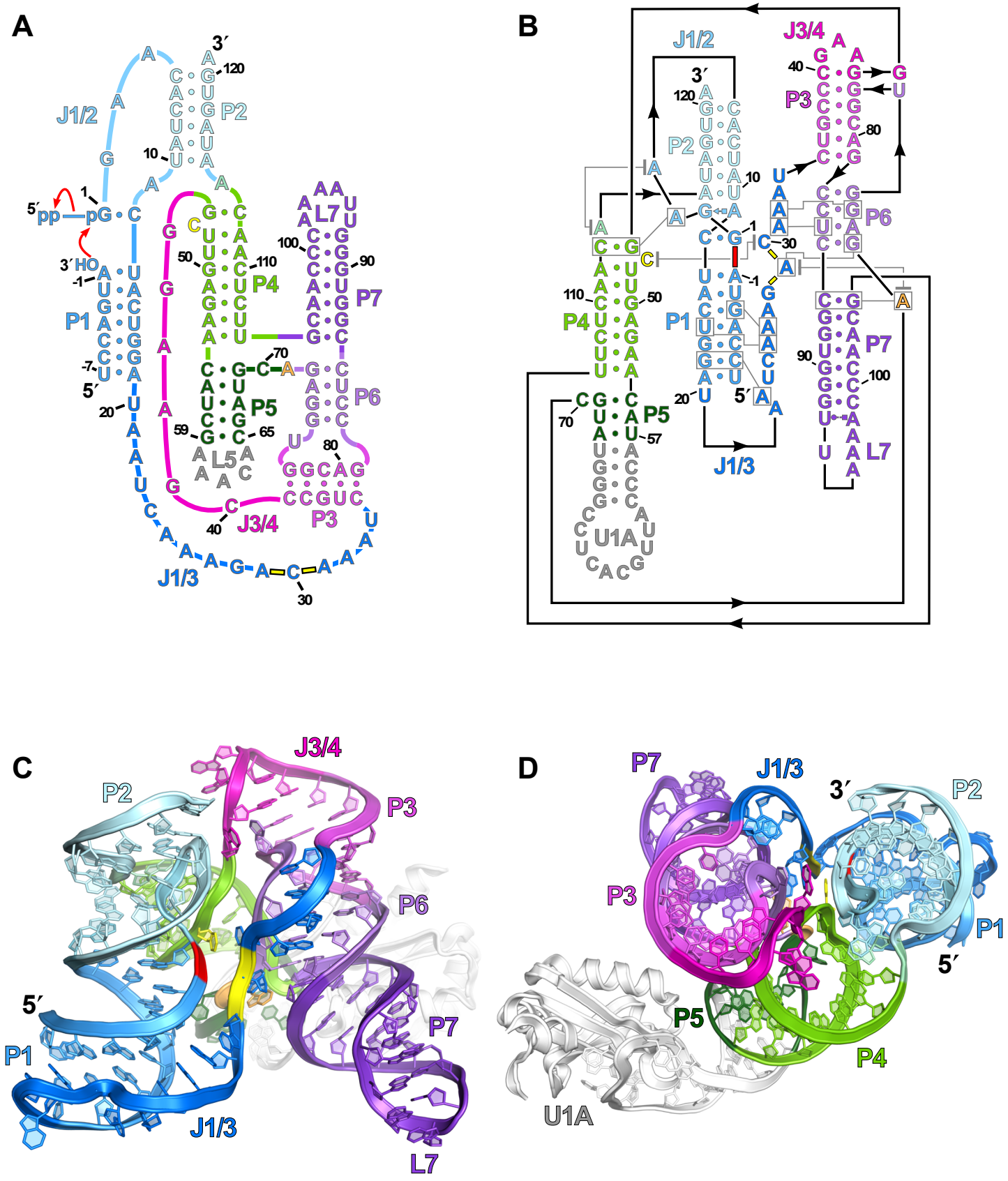


Figure 2

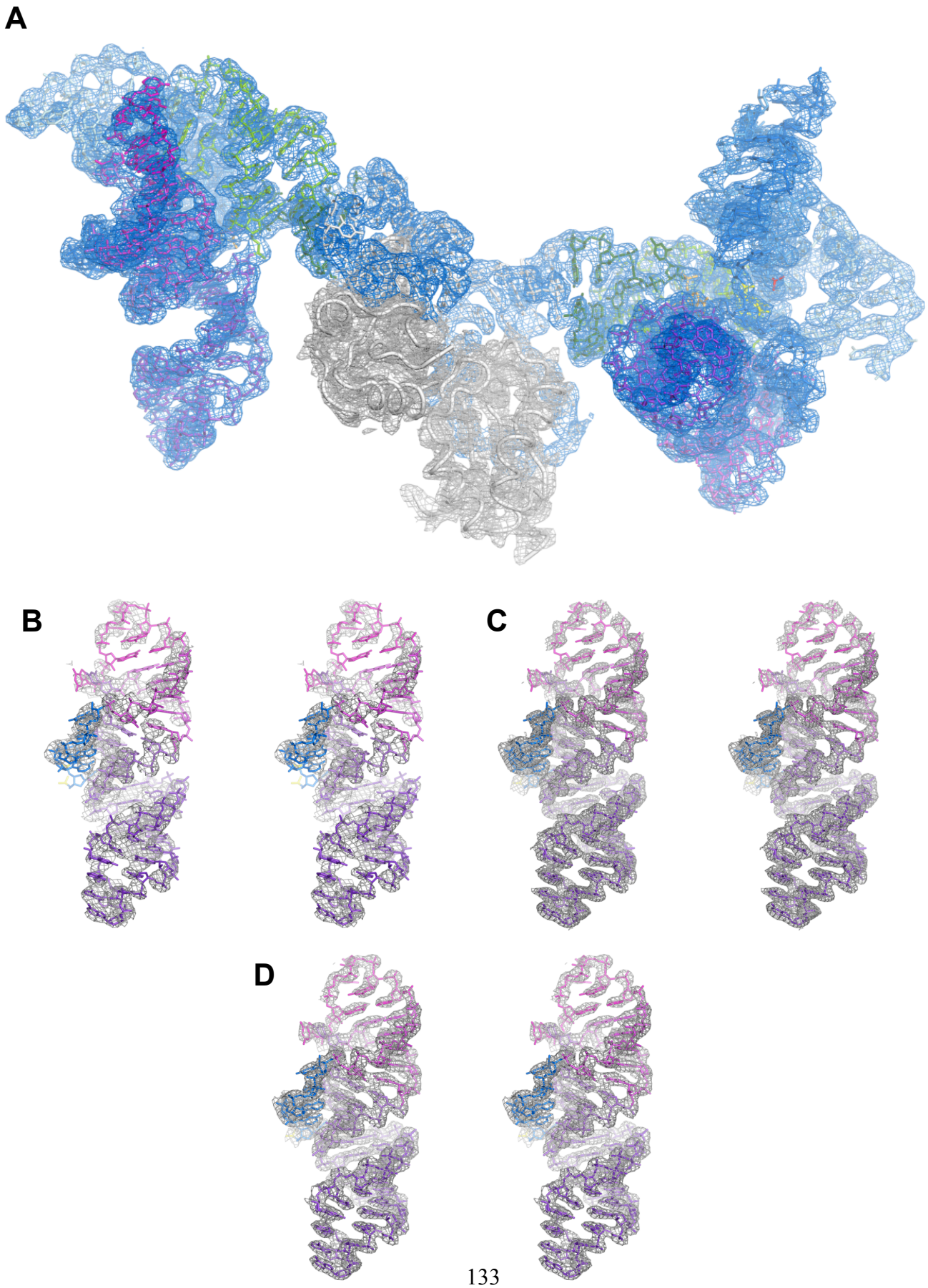


Figure 3

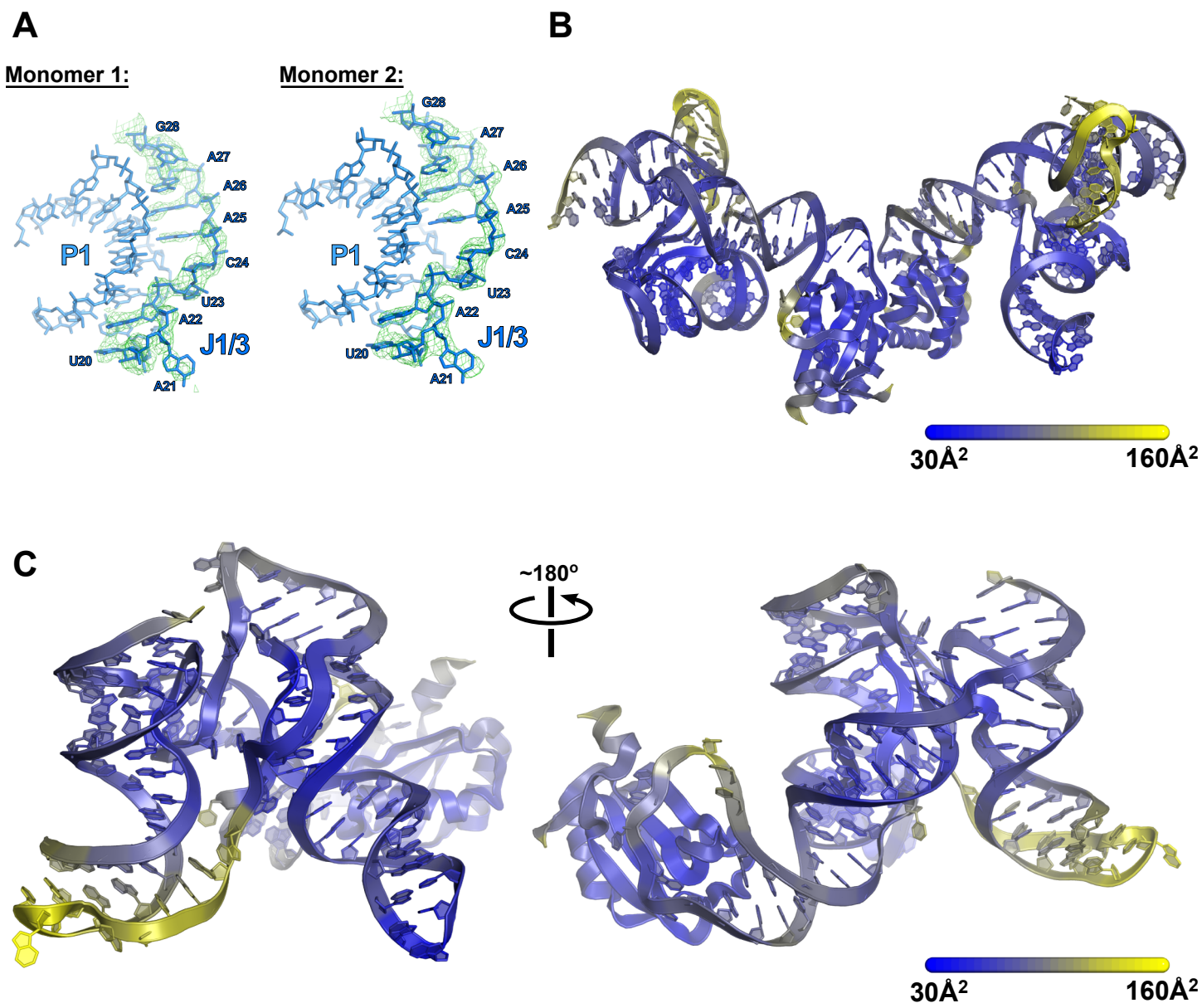

A
Monomer 1:
Monomer 2:
G28
A27
A26
A25
P1
C24
U23
A22
J1/3
U20
A21
B
30Å²
160Å²
C
~180°
30Å²
160Å²

Figure 4

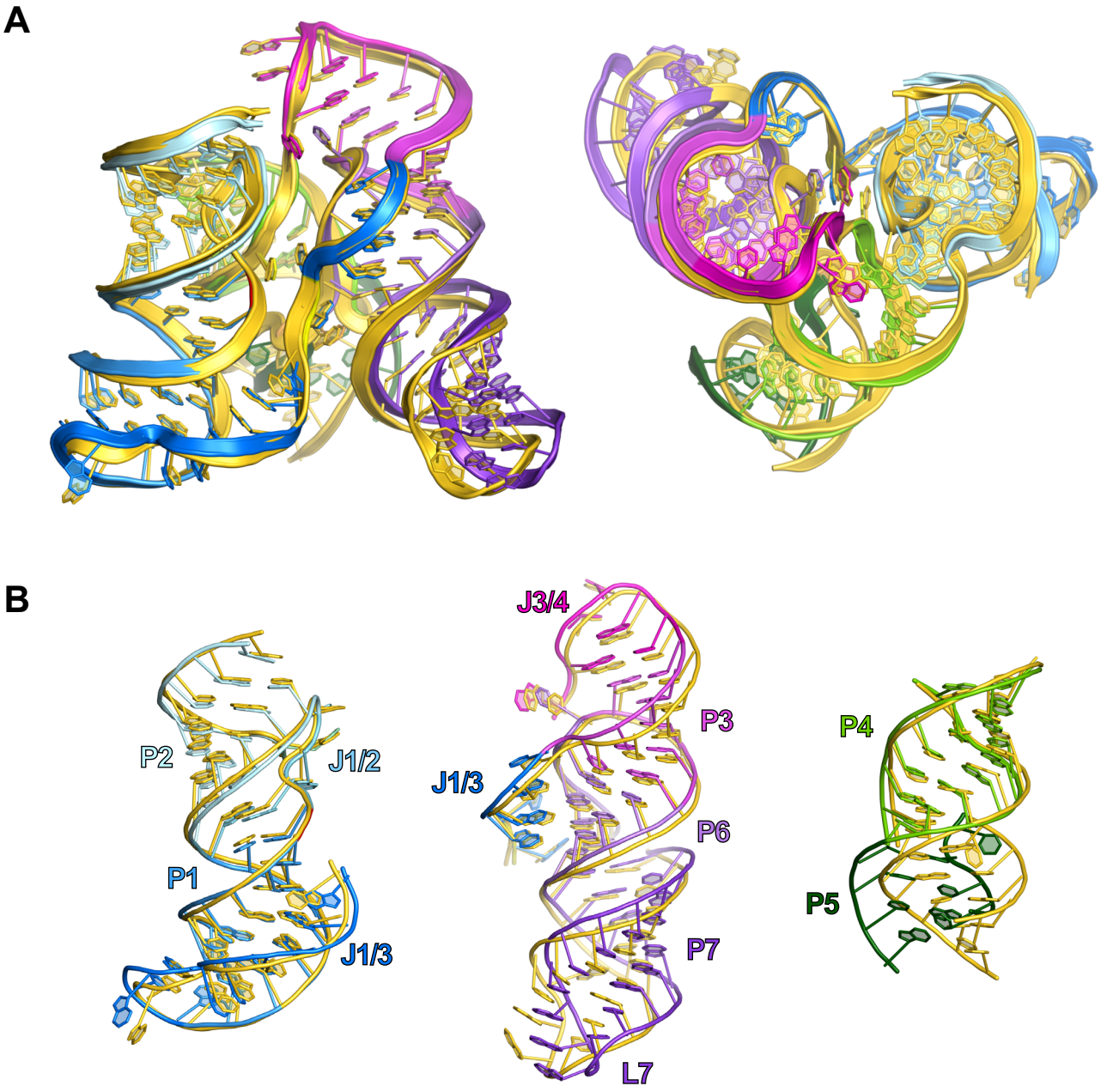
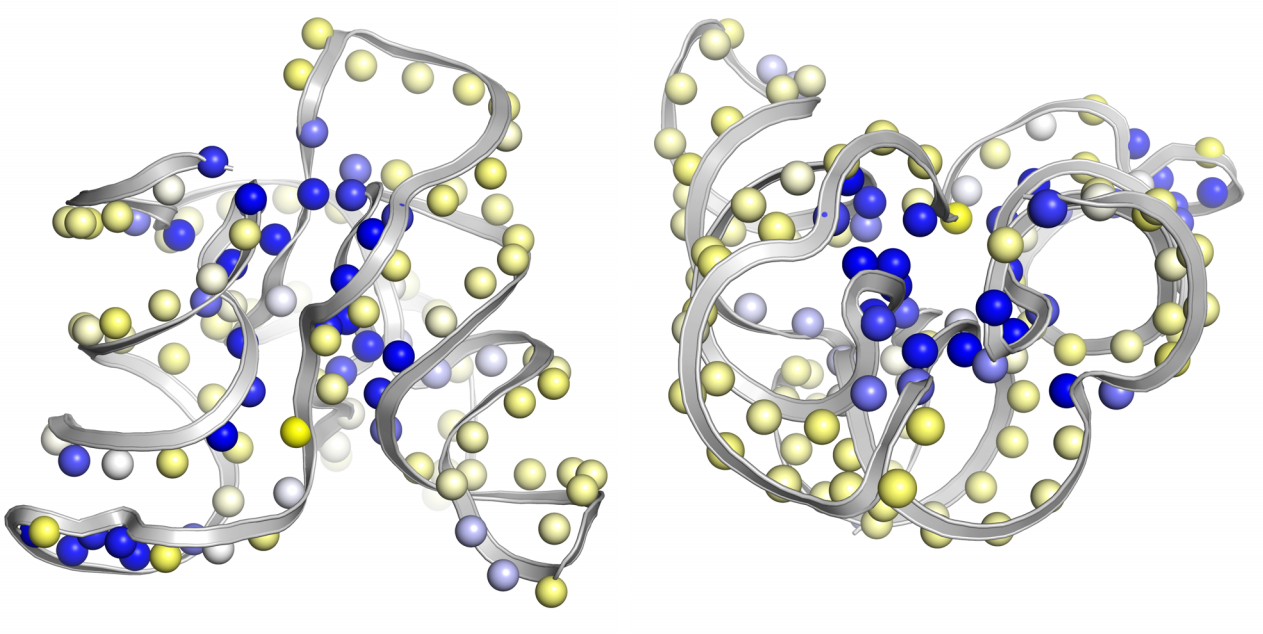


Figure 5

A



B

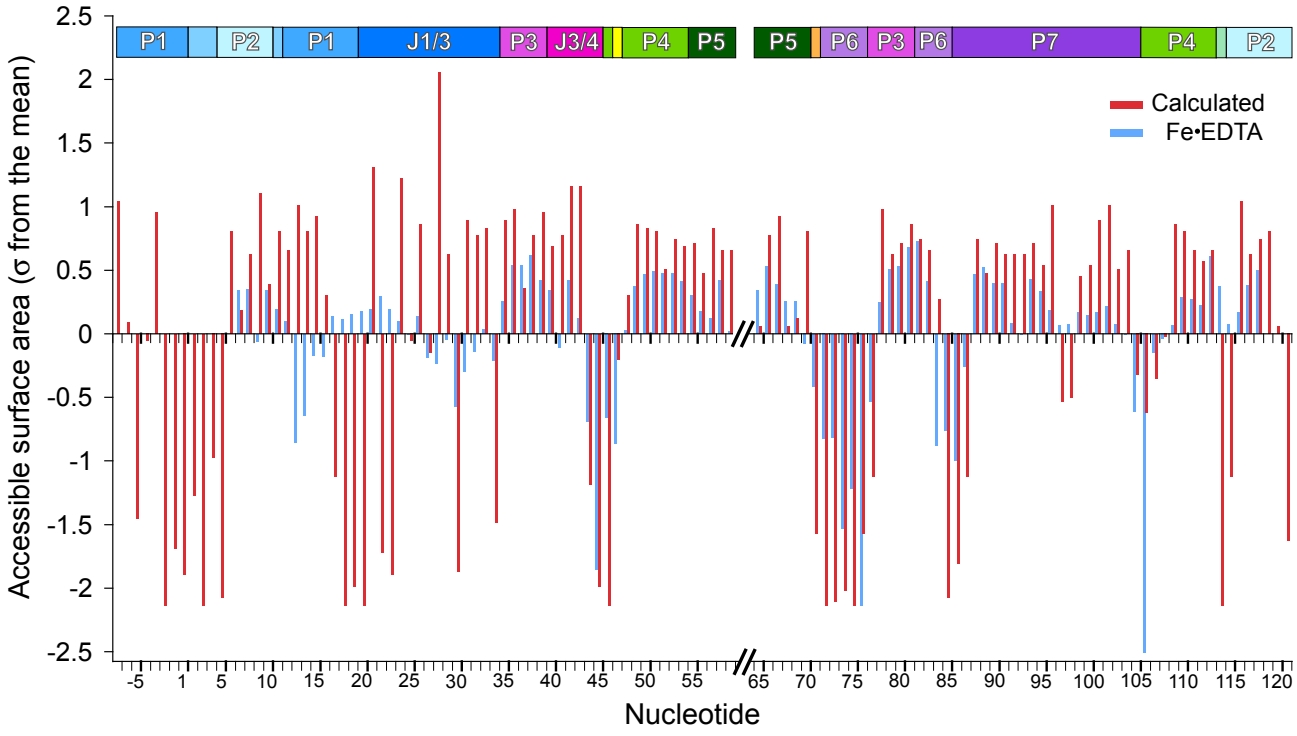

P1
P2
P1
J1/3
P3
J3/4
P4
P5
P5
P6
P3
P6
P7
P4
P2
Calculated
Fe•EDTA
Accessible surface area (σ from the mean)
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-5
1
5
10
15
20
25
30
35
40
45
50
55
65
70
75
80
85
90
95
100
105
110
115
120
Nucleotide

Figure 6

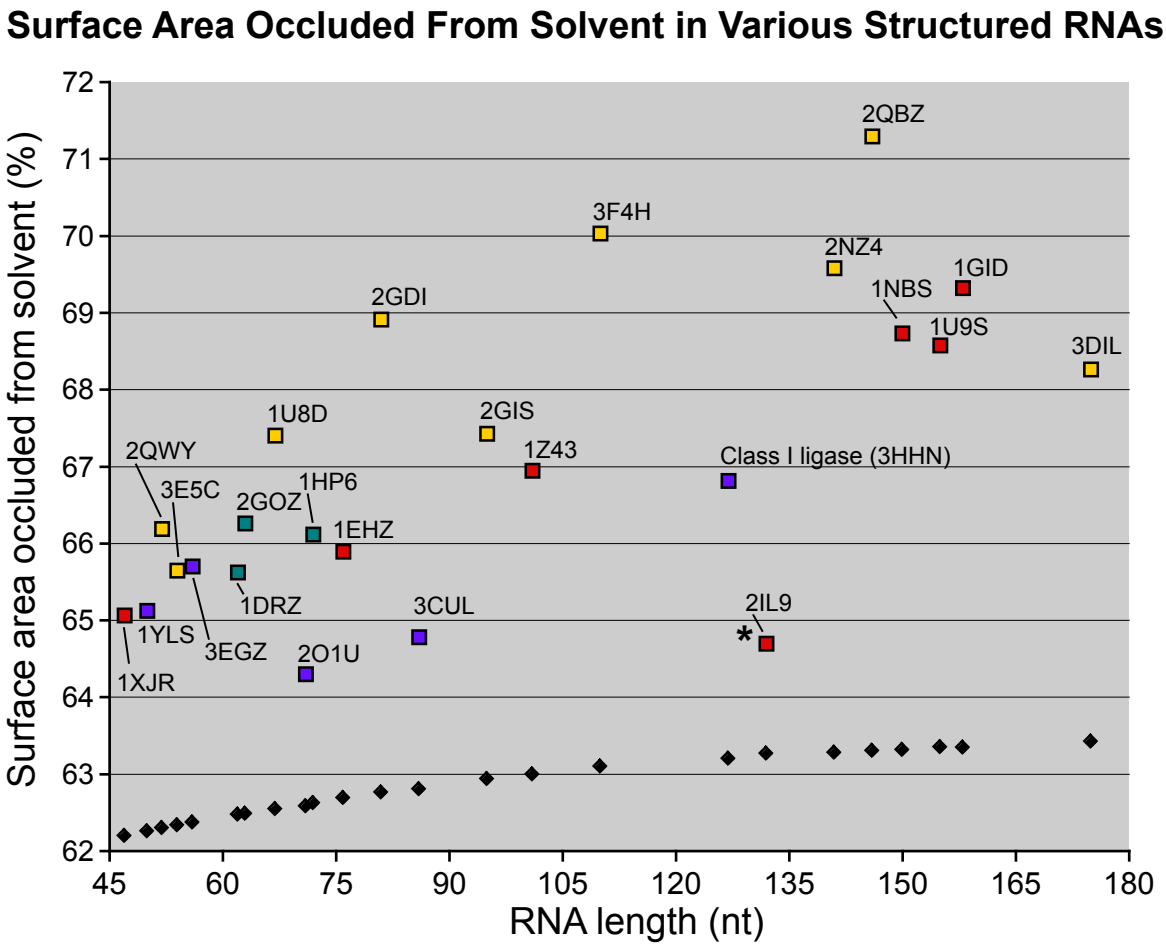

Surface Area Occluded From Solvent in Various Structured RNAs
Surface area occluded from solvent (%)
RNA length (nt)
72
71
70
69
68
67
66
65
64
63
62
45
60
75
90
105
120
135
150
165
180
2QBZ
3F4H
2NZ4
1GID
1NBS
1U9S
3DIL
2GDI
1U8D
2GIS
1Z43
Class I ligase (3HHN)
2QWY
3E5C
2GOZ
1HP6
1EHZ
1DRZ
3CUL
2IL9
*
1YLS
3EGZ
2O1U
1XJR

Figure 7

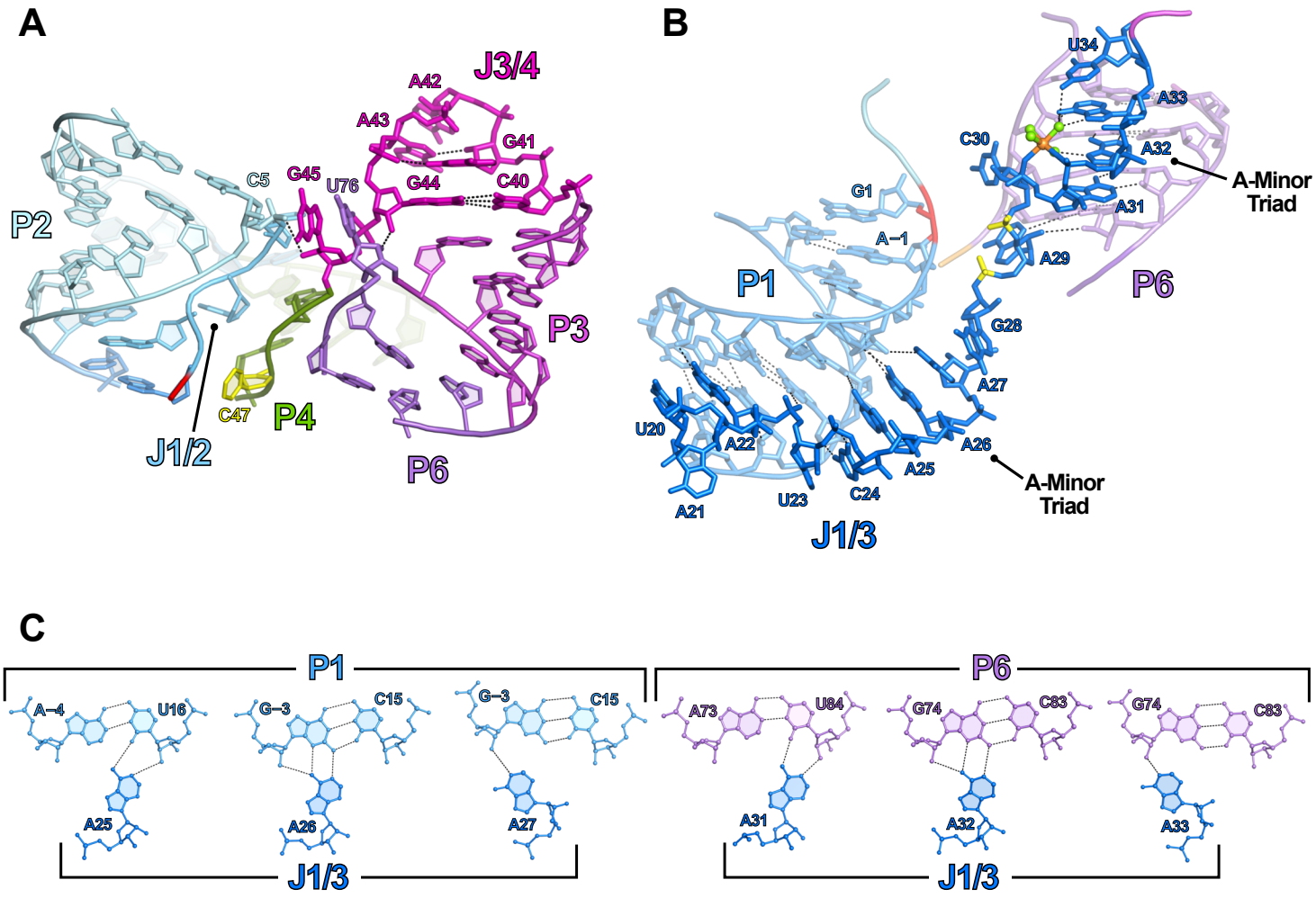

A
J3/4
A42
A43
G41
C5
G45
U76
G44
C40
P2
P3
C47
P4
J1/2
P6
B
U34
A33
C30
A32
A-Minor
Triad
G1
A31
A−1
A29
P1
P6
G28
A27
U20
A22
A26
A25
A-Minor
Triad
U23
C24
A21
J1/3
C
P1
P6
A−4
U16
G−3
C15
G−3
C15
A73
U84
G74
C83
G74
C83
A25
A26
A27
A31
A32
A33
J1/3
J1/3

Figure 8

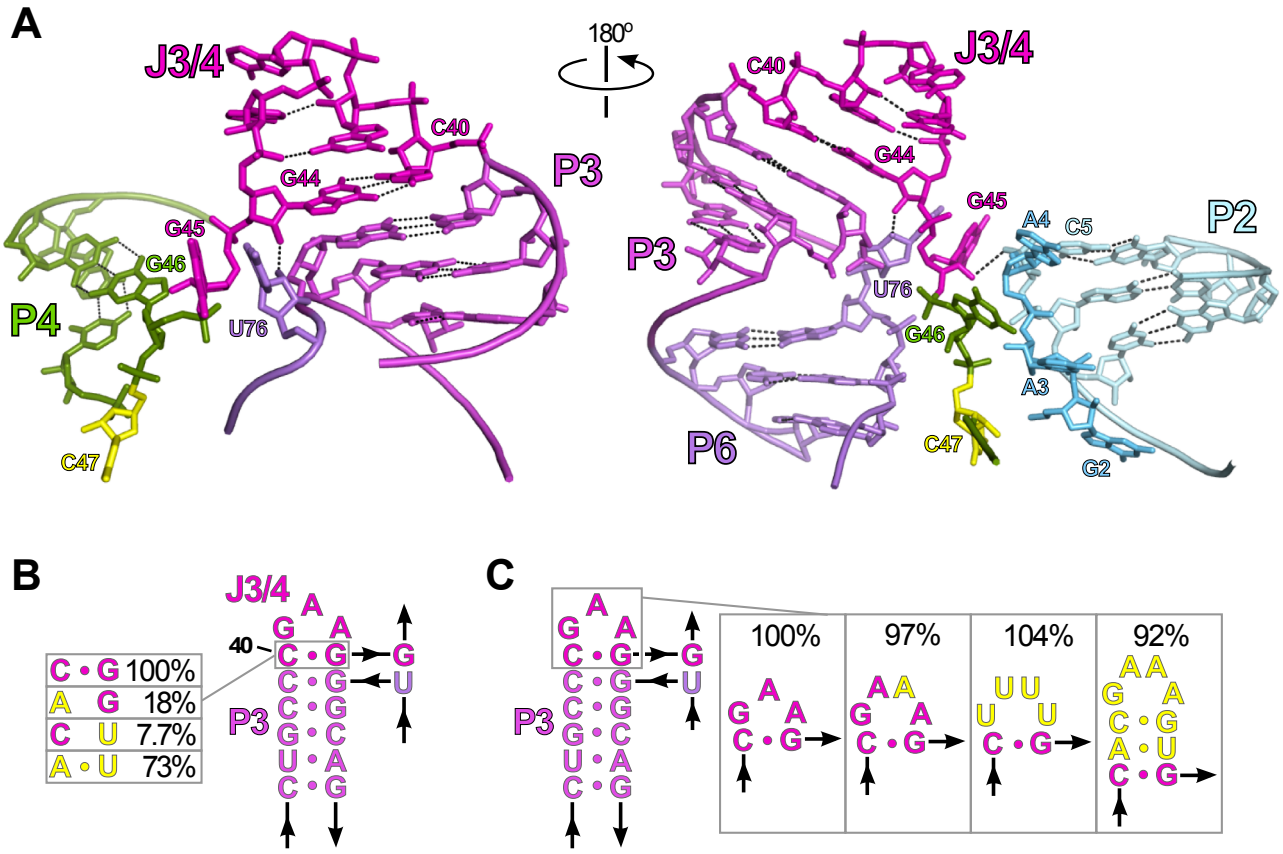

A
J3/4
C40
G44
G45
G46
P4
U76
C47
P3
180°
J3/4
C40
G44
G45
A4
C5
P2
P3
U76
G46
A3
P6
C47
G2
B
J3/4
C • G 100%
A G 18%
C U 7.7%
A • U 73%
40
P3
C
P3
100%
97%
104%
92%

Figure 9

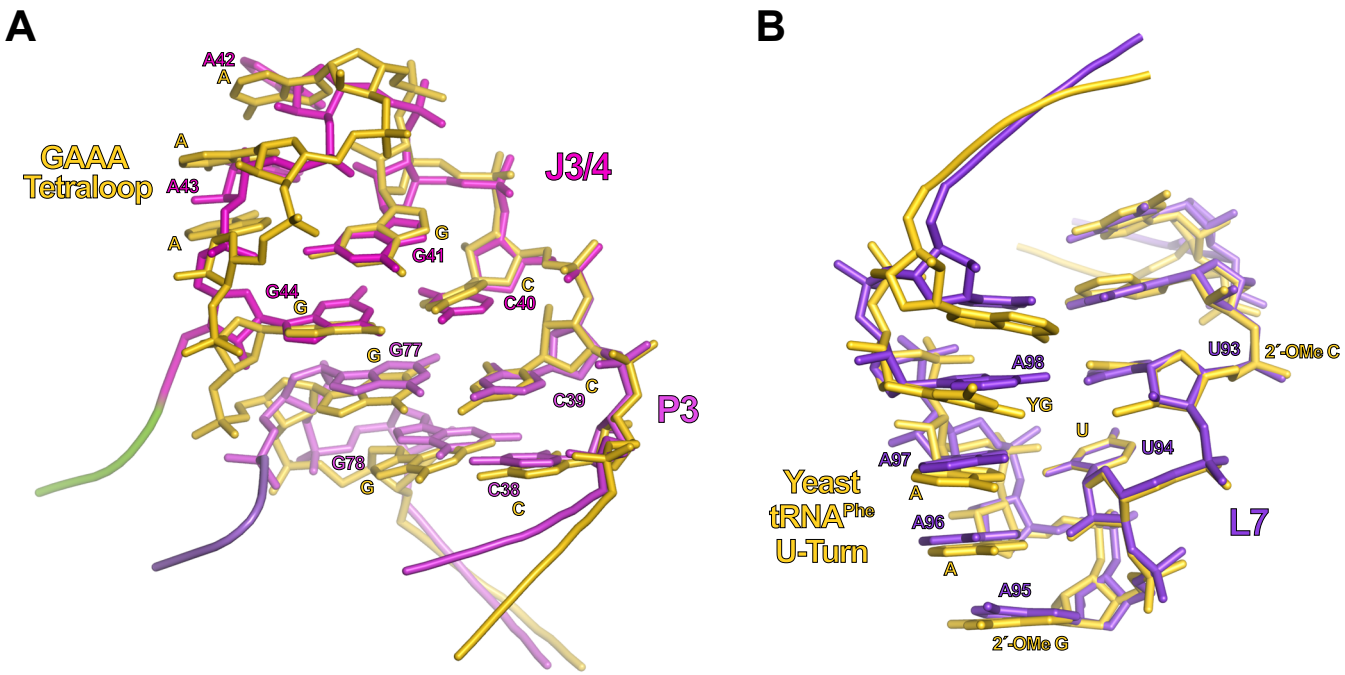

A
A42
A
GAAA
Tetraloop
A
A43
A
J3/4
G
G41
C
C40
G44
G
G
G77
C
C39
P3
G78
G
C38
C
B
A98
U93
2´-OMe C
YG
U
A97
U94
Yeast
tRNAPhe
U-Turn
A
A96
L7
A
A95
2´-OMe G

Figure 10

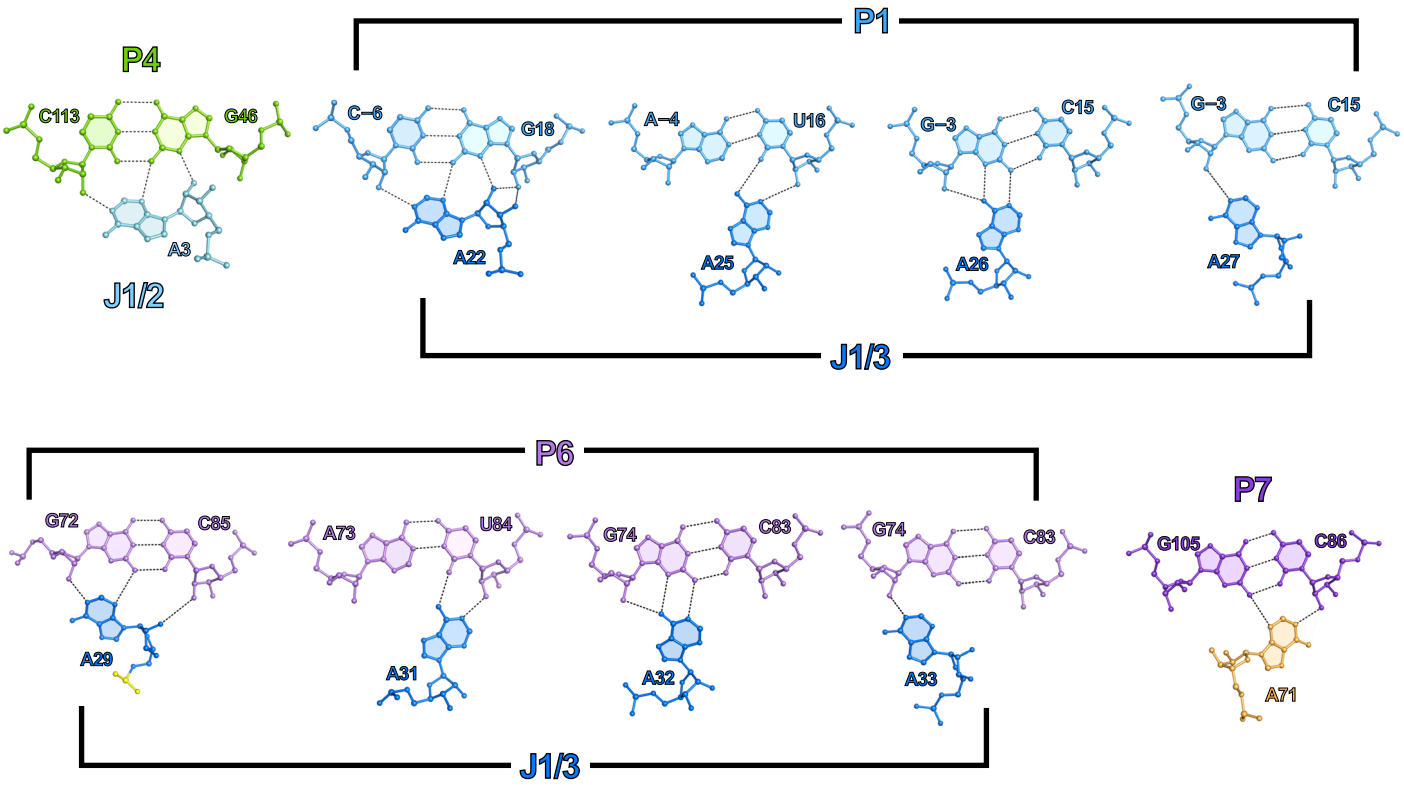

P1
P4
C113
G46
A3
J1/2
C–6
G18
A22
A–4
U16
A25
G–3
C15
A26
G–3
C15
A27
J1/3
P6
P7
G72
C85
A29
A73
U84
A31
G74
C83
A32
G74
C83
A33
G105
C86
A71
J1/3

Figure 11

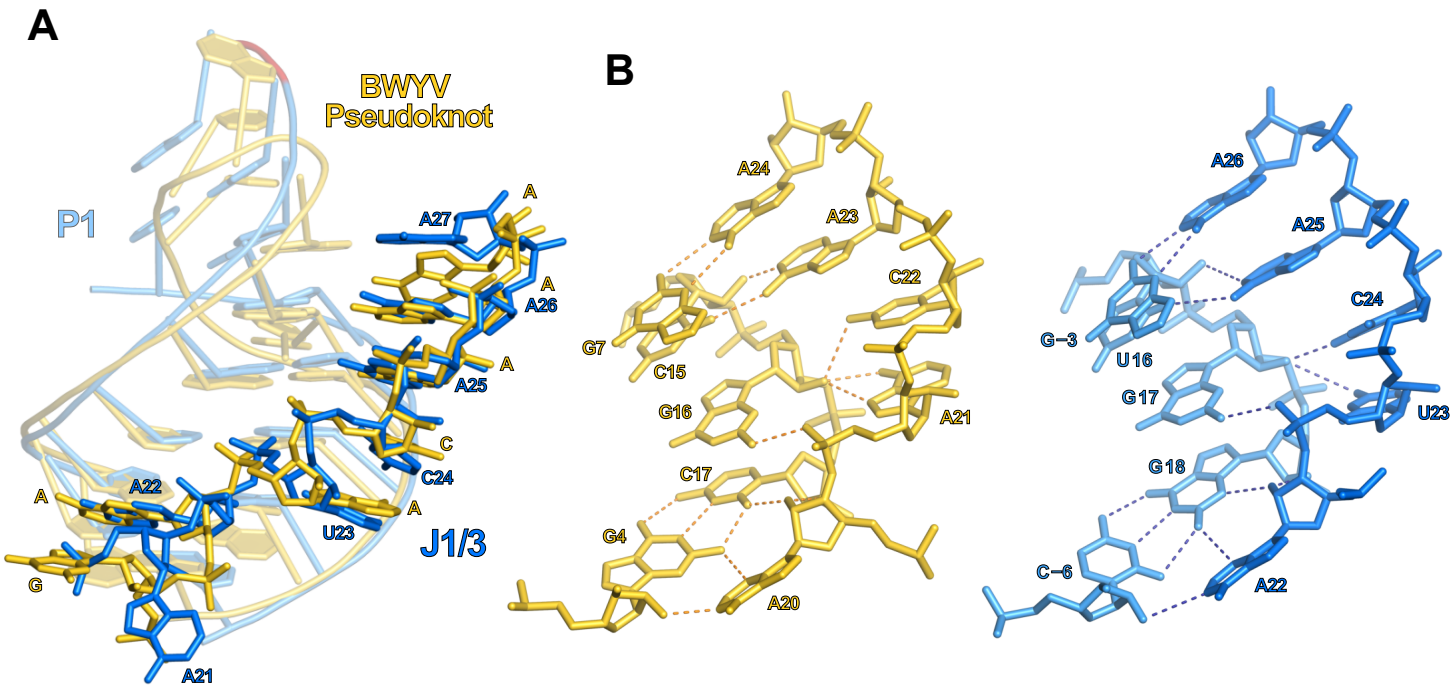

A
BWYV
Pseudoknot
P1
A27
A
A26
A
A25
C
C24
A
U23
J1/3
A
A22
G
A21
B
A24
A23
C22
G7
C15
G16
A21
C17
G4
A20
A26
A25
C24
G−3
U16
G17
U23
G18
C−6
A22

Figure 12

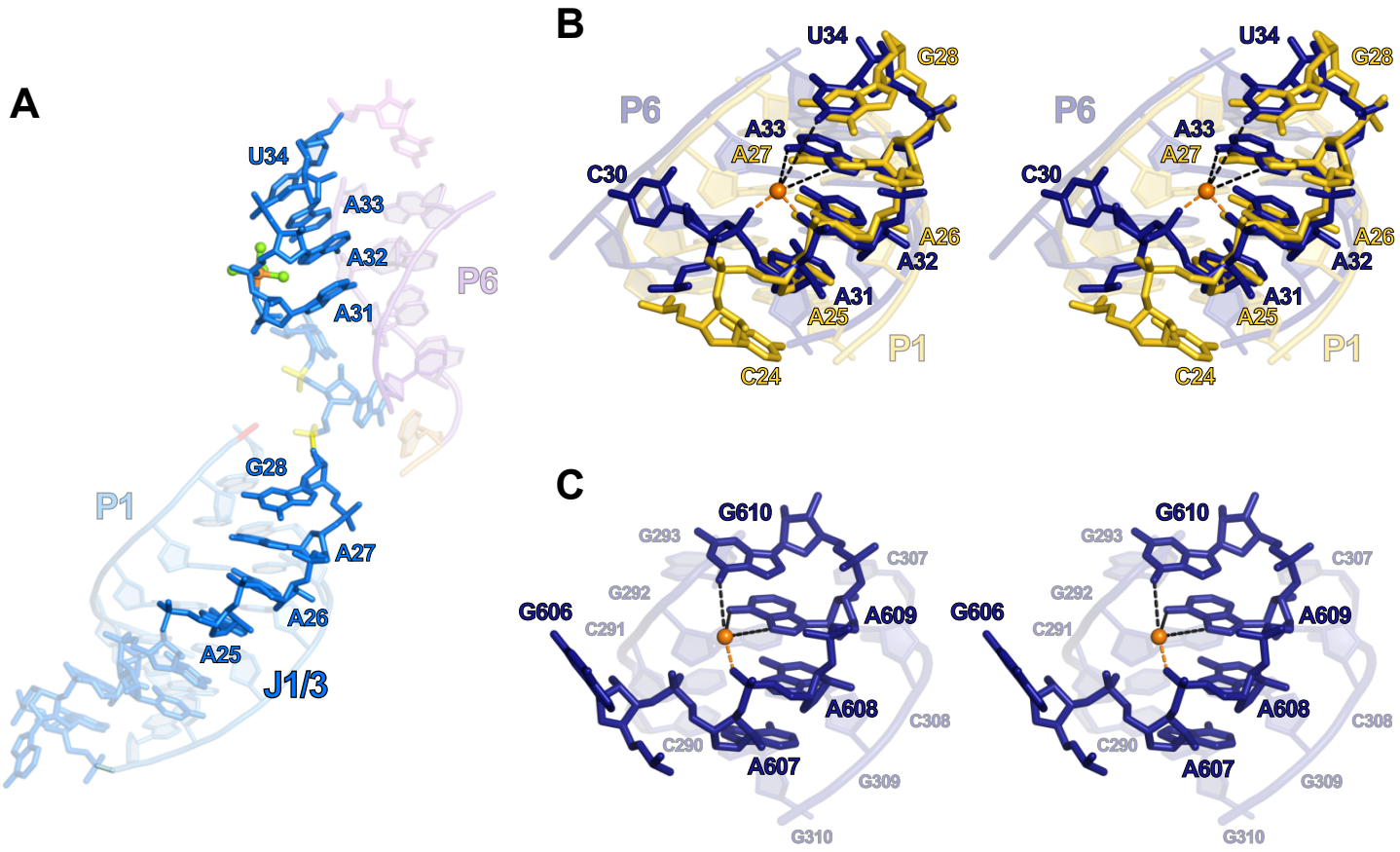
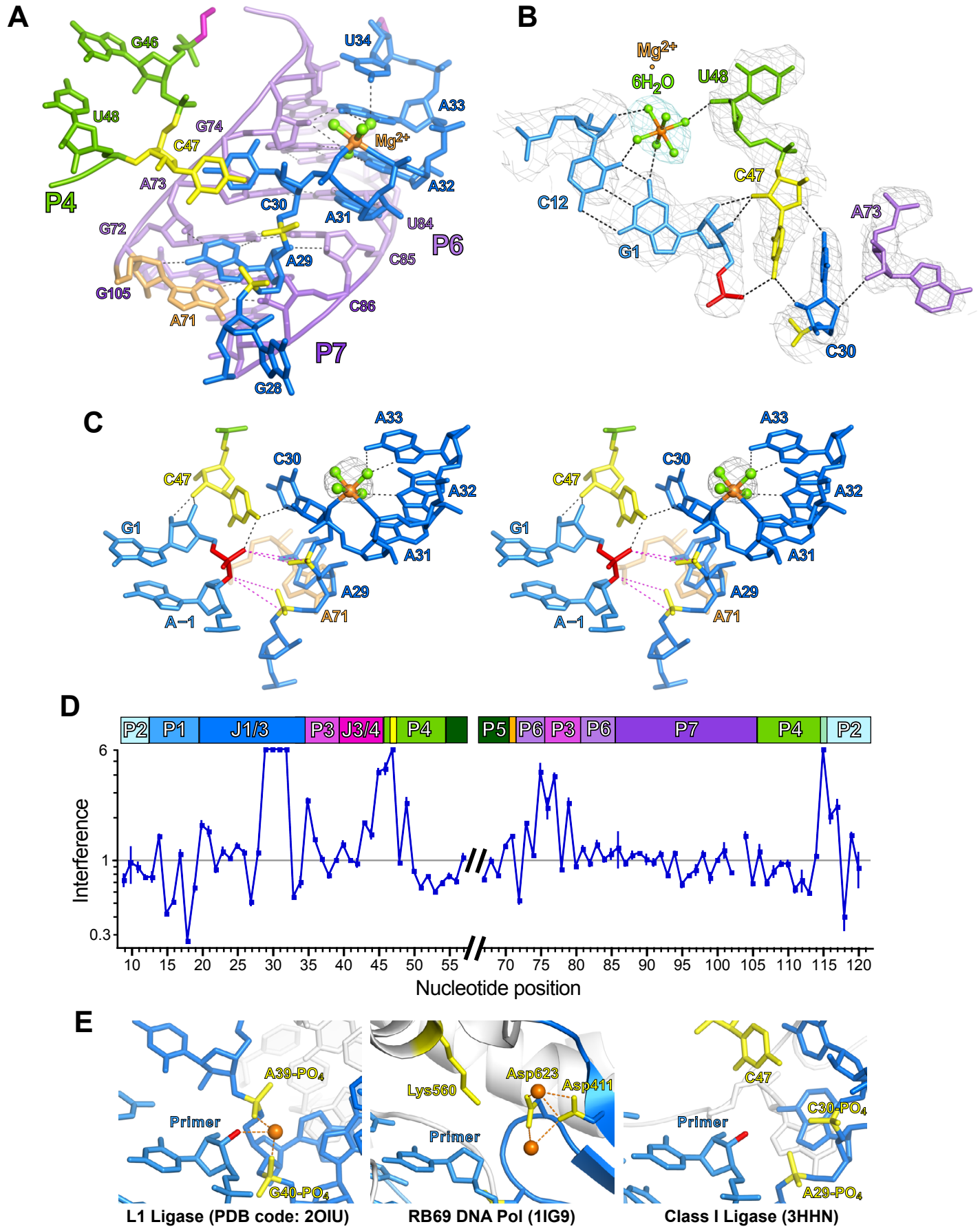

A
U34
A33
A32
A31
P6
G28
P1
A27
A26
A25
J1/3
B
U34
G28
P6
A33
A27
C30
A26
A32
A31
A25
P1
C24
U34
G28
P6
A33
A27
C30
A26
A32
A31
A25
P1
C24
C
G610
G293
C307
G292
G606
C291
A609
A608
C308
C290
A607
G309
G310
G610
G293
C307
G292
G606
C291
A609
A608
C308
C290
A607
G309
G310

Figure 13



L1 Ligase (PDB code: 2OIU) RB69 DNA Pol (1IG9) Class I Ligase (3HHN)

Figure 14

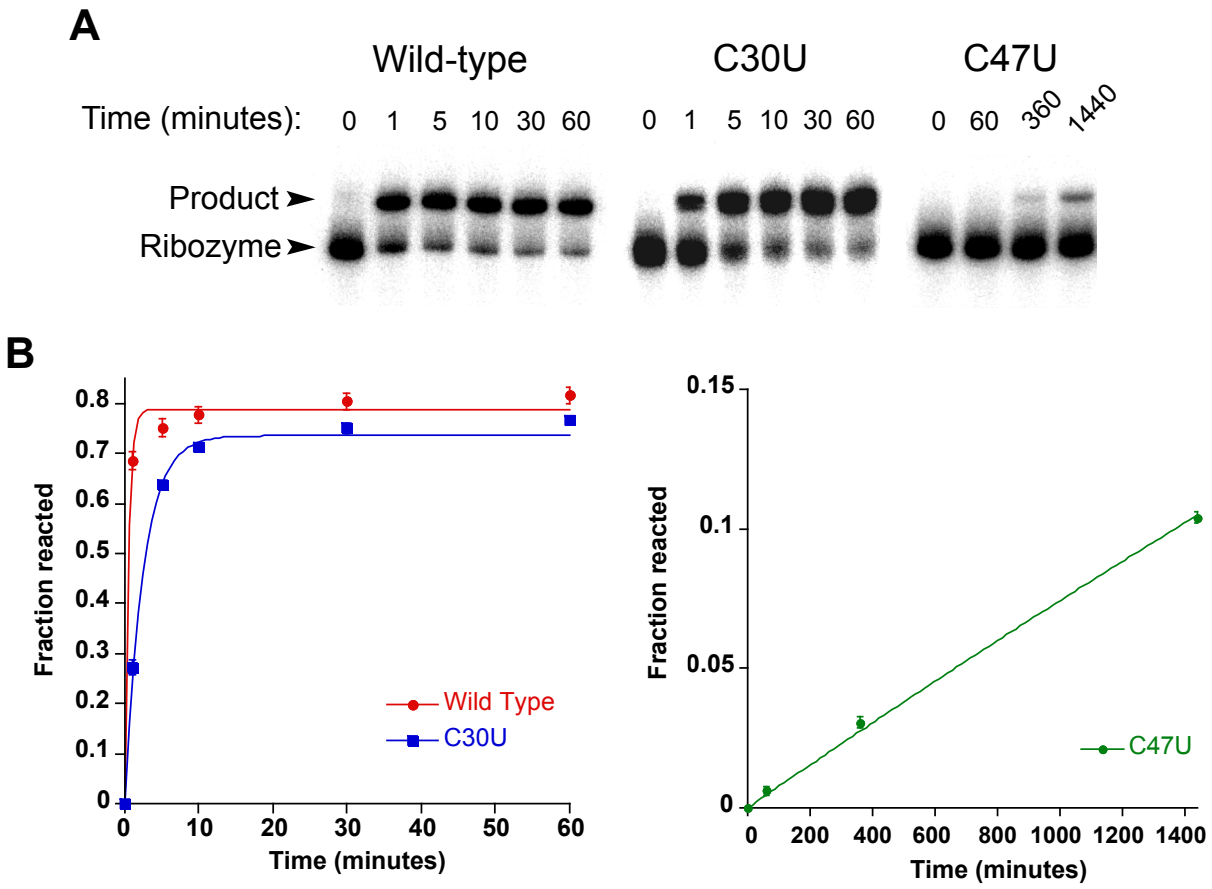

A
Wild-type
C30U
C47U
Time (minutes):
0 1 5 10 30 60
0 1 5 10 30 60
0 60 360 1440
Product
Ribozyme
B
Fraction reacted
Time (minutes)
Wild Type
C30U
C47U

Figure 15

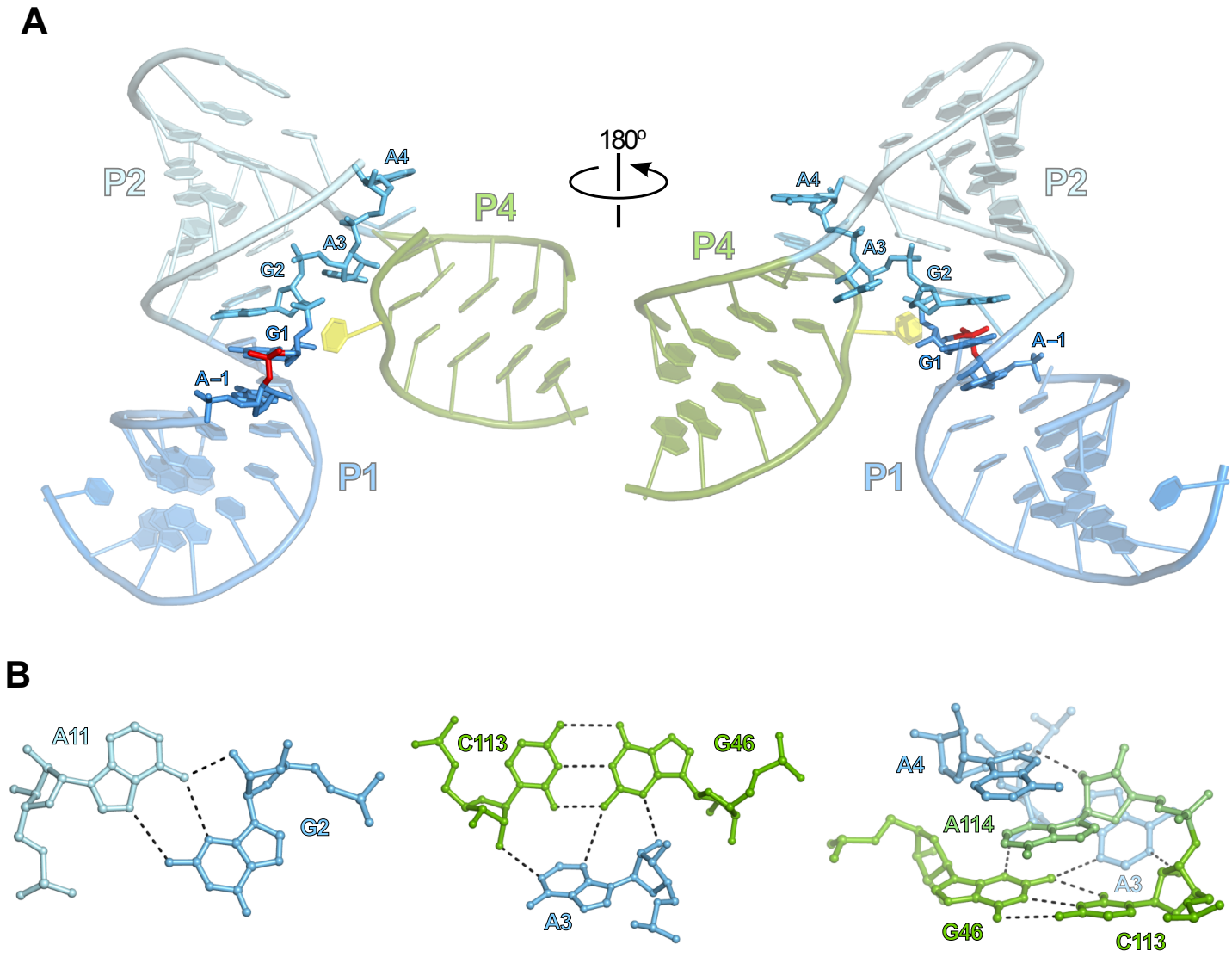

A
P2
A4
P4
A3
G2
G1
A–1
P1
180°
P4
A4
P2
A3
G2
G1
A–1
P1
B
A11
G2
C113
G46
A3
A4
A114
A3
G46
C113

Figure 16

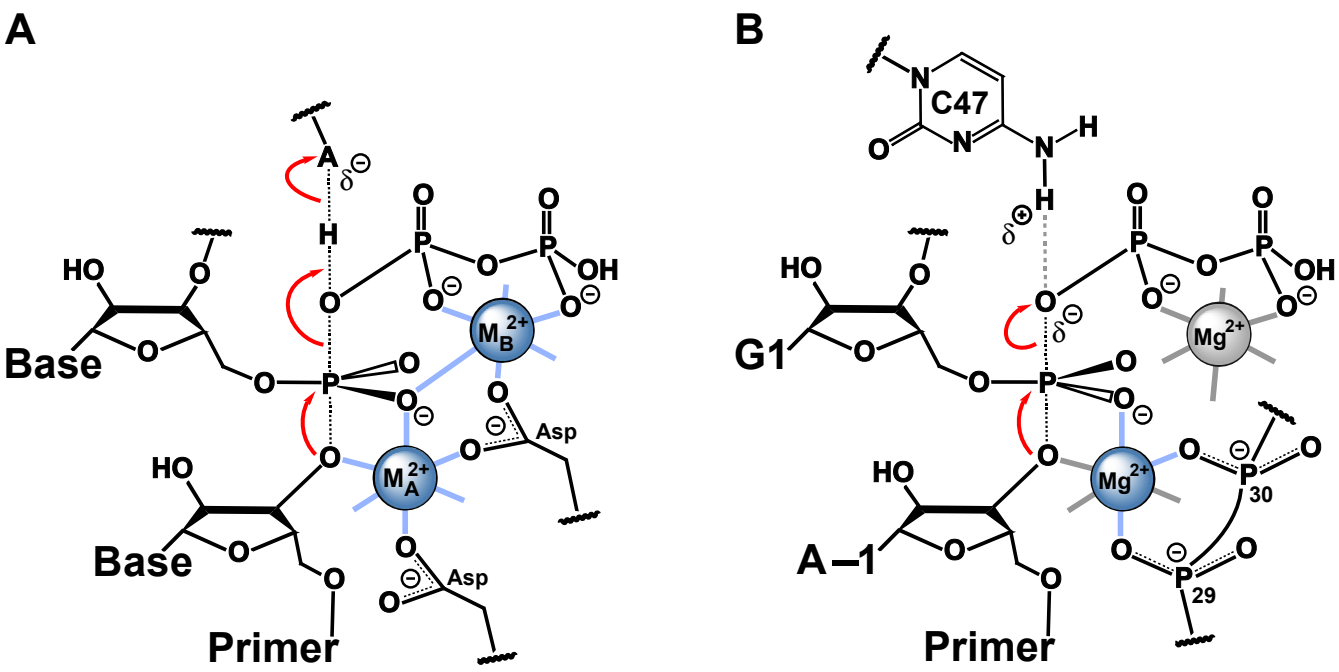
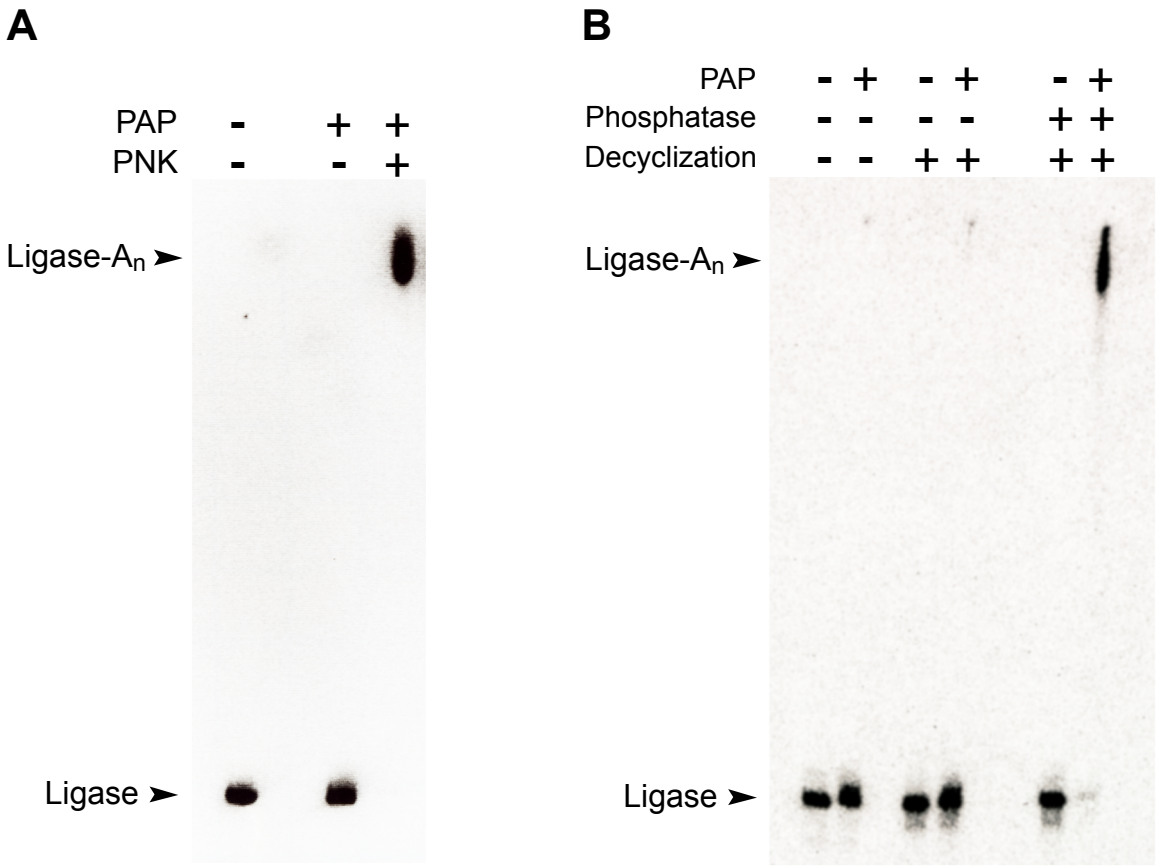


Figure 17


A
PAP - + +
PNK - - +
Ligase-An
Ligase
B
PAP - + - + - +
Phosphatase - - - - + +
Decyclization - - + + + +
Ligase-An
Ligase

The Structural Basis of RNA-Catalyzed RNA Polymerization

This work is based upon research conducted at the Northeastern Collaborative Access Team (NE-CAT) beamlines of the APS, supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

Abstract

According to the RNA world hypothesis, the emergence of early life would have required polymerase ribozymes capable of replicating RNA. Known catalytic RNAs best approximating these primordial replicases use an artificial ribozyme, the Class I ligase, as their catalytic core. We recently solved the crystal structure of the Class I self-ligated product, and here present 3.15-Å structures of ligase complexes trapped prior to catalysis. Two active-site phosphates and the 3′-hydroxyl nucleophile jointly coordinate an essential metal ion. The nucleophile is perfectly positioned for in-line attack on the 5′-triphosphate, which, mediated through an array of solvent interactions, hooks into the major groove of the adjoining RNA duplex. Atomic mutagenesis experiments demonstrate that active site nucleobase and ribose hydroxyl groups directly participate in catalysis. Kinetic isotope experiments imply that these groups function electostatically, perhaps playing a role that in proteinaceous polymerases is performed by a metal ion. These data provide the most detailed structural and mechanistic understanding of an artificial ribozyme to date, and demonstrate that ribozymes and proteins may adopt divergent strategies when catalyzing the same reaction.

Introduction

The transmission and use of genetic information fundamentally depend on the action of template-directed nucleic acid polymerases, enzymes that in modern organisms are built exclusively from protein. Although these polymerases act as the central guardians of genetic information, like all proteins they contain little information themselves: no known biochemical system can polymerize amino acids using a polypeptide template. The act of copying proteins is consequently a complex process involving the participation of many distinct components, including DNA genes, tRNAs, synthetases, ribosomes and so forth. That neither DNA nor protein, the principal biopolymers of information-storage and catalysis, respectively, can be replicated without the action of the other, represents a fundamental paradox underlying the emergence of modern life.

The RNA World hypothesis[1], offers an elegant solution to this paradox, proposing that primordial organisms lacked both DNA genes and protein enzymes, and instead used RNA as the sole biopolymer of information-storage and catalysis. Support for this theory comes from the observation that many of life's most fundamental processes inextricably depend on the action of RNA, its precursors or its derivatives[1-3]. Furthermore, the existence of natural RNA enzymes (ribozymes)[4-6], including the ribosome[7], not only illustrates that the RNA polymer can function as both genome and catalyst, but also implies that some of the oldest extant biological processes were originally built from RNA.

Natural ribozymes are only known to catalyze the cleavage or rearrangement of phosphodiester bonds[8] and the formation of peptide bonds[9]. Hence, if the catalytic scope of RNA is broad enough to support a full metabolism, it is not demonstrated by the capabilities of natural isolates. In contrast, the menagerie of artificial ribozymes produced from *in vitro*

selection experiments has revealed the catalytic purview of RNA to be quite broad: with the exception of free-radical chemistries, almost every common chemical transformation in modern biology is now known to be within the grasp of RNA[10,11]. However, a ribozyme possessing arguably the most fundamental activity required for RNA-based life has yet to be isolated. In the putative organisms of the RNA World, just as in modern life, the processes of copying and utilizing genetic information would have required a nucleic acid polymerase[12].

Several lines of experimentation have sought to demonstrate the ability of RNA to replicate RNA. Natural ribozymes have been contorted into performing very limited template-directed polymerization (reviewed in ref. 12), though the fidelity and processivity of these derivatives precludes their functioning as general replicases. An alternative route has explored the use of short oligonucleotides–rather than mononucleotide triphosphates (NTPs)–as the individual monomers being polymerized[13-16]. In this way, a kind of self-replicating system could be built using a pair of complementary artificial ligase ribozymes[17]. However, since this system is obligatorily bipartite, and lacks sequence universality, it too is unlikely to produce a general replicase ribozyme.

To date, known catalytic RNAs with activities best approximating that required for general RNA replication have been derived from an artificial ribozyme termed the Class I ligase (**Fig. 1A**). This ribozyme was among the first to be isolated by *in vitro* selection[18] and accelerates a reaction functionally identical to a single turnover of template-directed primer-extension as it is performed in modern biology. Namely, it catalyzes the nucleophilic attack by a 3′-hydroxyl on a 5′-triphosphate, requiring that both groups be positioned by Watson-Crick base pairing, and yields a 3′-5′ phosphodiester linakge and pyrophosphate[18-20]. Ligase variants have been engineered to catalyze limited, though highly accurate, template-directed primer-extension

using exogenous NTPs, meaning that the ligase core alone can act as a genuine RNA polymerase ribozyme[21]. Moreover, through further *in vitro* selection the ligase has been co-opted to provide the catalytic engine for several RNA polymerase ribozymes that use NTPs and the information on exogenous templates to extend primers *in trans*[22-24]. Some of these ribozymes can use primer-template (PT) substrates of any sequence[22,24], and although they exhibit varying degrees of processivity[25], with the best templates one isolate can accurately extend primers by up to 22-nt in a 24-hour incubation[24].

Furthermore, the ligase is remarkably fast, accelerating its chemistry nearly 10^9-fold over the uncatalyzed reaction, achieving rates that exceed those of most natural ribozymes and rival those of some proteins[19,20,26]. Its robustness has made the ligase a versatile tool in experiments that model evolution *in vitro*[27-30], and a potential biosensor for medical diagnostics[31]. Given this catalytic prowess, its role in origins-of-life research and the extensive scrutiny with which it has been characterized[26,32-34], the Class I ligase represents a fascinating target for in-depth structural analysis.

We recently reported the crystal structure of the Class I ligase ribozyme in a state corresponding to its self-ligation product[35]. This structure revealed a wealth of knowledge about the overall ribozyme architecture, and identified putative components of the active site. However, the active site visualized in this first structure was incomplete, since the ligase is by necessity a single-turnover catalyst and in its postcatalytic state had already rendered itself inert. Therefore, to develop a deeper understanding of the structural basis for RNA-catalyzed RNA polymerization, we here present crystal structures of the ligase trapped in a catalytically viable preligation state. We have also employed a variety of biochemical techniques to interrogate the catalytic function of groups at the active site. These results have revealed that the ligase employs

a catalytic strategy more complex than those of other artificial ribozymes[36-39], and distinct from that of the proteinaceous polymerases.

Results

Crystallization of Class I ligase preligation complexes

Our efforts to crystallize the preligation complex focused on derivatives of the U1A-binding construct and crystallization conditions previously described for the product species[35]. Since the ligase normally undergoes rapid self-ligation under these crysallization conditions[34,35], several methods were explored that might trap the ribozyme in a precatalytic state without globally perturbing its architecture. While other enzymes have been similarly trapped and crystallized by modifying their respective nucleophiles[8], or leaving groups[40], neither of these methods produced useful crystals of the ligase. Instead, we employed two modifications that each reduce, but do not completely ablate, catalysis. First, we mutated the active site nucleotide C47 to uridine, which reduces the catalytic rate by a factor of $>10^4$ (ref. 35, **Fig. 1A**). Second, we exploited the metal ion dependencies of the ligase: although its rate is $>10^5$ slower in Ca^{2+} than it is in Mg^{2+}, the ribozyme adopts conformations indistinguishable by chemical probing when folded in either metal[33].

Crystals of the C47U mutant preligation complex were obtained under conditions similar to those reported for the product species[35], replacing Mg^{2+} with an inhibitory mixture of Ca^{2+} and Sr^{2+} ions (hereafter, "Ca^{2+}/Sr^{2+}-complex" crystals). Mg^{2+}-bound crystals were obtained by soaking Ca^{2+}/Sr^{2+}-complex crystals in solutions of Mg^{2+} and Sr^{2+} before freezing (hereafter, the "Mg^{2+}/Sr^{2+}-complex" crystals). The structure of each of these crystals was independently solved to 3.15-Å (**Fig. 1B–C**) and refined to R_{work} and R_{free} values of 21.2 and 25.5%, respectively, for the Ca^{2+}/Sr^{2+}-complex, and 19.7 and 24.0%, respectively, for the Mg^{2+}/Sr^{2+}-complex (**Table 1**). Although parallel soaking experiments were performed in the absence of Sr^{2+}, the resulting crystals suffered from prohibitively high static disorder. Regardless, we do not anticipate that

the presence of Sr^{2+} in the Mg^{2+}/Sr^{2+}-complex dramatically perturbed the ribozyme structure: Sr^{2+} is only a mild ligase inhibitor[33], having a <5-fold effect on catalysis under the crystallization conditions. Furthermore, we observe no peaks corresponding to Sr^{2+} density in the resulting electron density maps.

Comparison of the preligation and product structures

The overall conformation of the ligase is remarkably similar in the pre- and post-catalytic states, being more dramatically altered by binding different metals than it is during the course of catalysis (**Fig. 2A–B**, *left*). Excluding nucleotides in the U1A-binding loop, the Mg^{2+}/Sr^{2+}-complex and product structures superpose with an all-atom root mean square deviation (r.m.s.d.) of 0.31-Å, less than the mean precision of either structure (ref. 35, **Table 1**). The r.m.s.d. between the Ca^{2+}/Sr^{2+}-complex and product structures is more than twice as high, 0.76-Å, with the majority of perturbations lying in the P1–P2 domain or near the active site (**Fig. 2A**). Residues 12–25, comprising the 3′-end of helix P1 and the 5′-end of J1/3 (**Fig. 1A**), were particularly affected, suffering from a high degree of crystallographic disorder in the Ca^{2+}/Sr^{2+}-complex structure. Replacement of Ca^{2+} with Mg^{2+} appears to relieve this disorder, reducing the average temperature factors in the region from 164.6 $Å^2$ in the Ca^{2+}/Sr^{2+}-complex to 117.1 $Å^2$ in the Mg^{2+}/Sr^{2+}-complex (**Fig. 2C**).

Reconstitution in Ca^{2+} also has a dramatic effect on the active site architecture (**Fig. 2A**, *right*). Although nucleotides immediately abutting the ligation junction occupy similar conformations in the Ca^{2+}/Sr^{2+} and product structures, the residues to which they pair and those comprising the active site are more substantially altered. Widening of the P2 helix appears to weaken the base pair between G1 and the template residue C12, increasing the distance between

their C1′-carbons by 0.3-Å and rotating the C12 base 24º away from ideal pairing geometry. The G1 ribose is concomitantly shifted ~1-Å away from its position in the product structure, which, owing to the steric effects of the 5′-triphosphate, pushes the U47 base out of the active site by the same distance. Other structural perturbations abut sites of inner-sphere metal ion coordination. Hence, widening of metal binding sites in the P6 A-minor triad[35] and near the four-way junction centered at A71, can be justified as steric effects incurred by Ca^{2+}, which is ~0.3-Å larger than Mg^{2+} (ref 41). None of the above structural changes are observed in the active site of the Mg^{2+}/Sr^{2+}-complex (**Fig. 2B**, *right*).

Metal ions in the active site

Although biochemical data[34,35] support a model in which the active site A29 and C30 backbone phosphates coordinate a catalytic metal ion, electron density for this metal was not observed in the original product structure[35]. This is not the case for the preligation complexes. During refinement of each structure, $|F_o| - |F_c|$ difference fourier maps prominently featured spherical or oblate positive peaks centered within the active site (**Fig. 3A–B**). The 7.8σ difference peak in the Ca^{2+}/Sr^{2+}-complex is perfectly suited to accommodate a partially hydrated Ca^{2+} ion (**Fig. 3A**). Docking a $Ca^{2+}\bullet 3H_2O$ complex into this peak positions the Ca^{2+} center 2.4-Å from the A29 pro-*Sp* and C30 pro-*Rp* oxygens, and from the A–1 3′-hydroxyl nucleophile. A29 is concomitantly shifted 1.6-Å away from its position in the product structure (**Fig. 2A**), its phosphate rotated so as to stabilize the pro-*Sp* oxygen coordination contact. The corresponding 4.8σ difference peak in the Mg^{2+}/Sr^{2+}-maps would represent a less ideal Mg^{2+} binding site, positioning the metal center 2.5-Å from the A–1 3′-OH and C30 pro-*Rp* oxygens and ~2.7-Å from each of the A29 nonbridging phosphate oxygens (**Fig. 3B**). Considering the mean precision

of our data (**Table 1**), and the maximum and mean inner-sphere bonding distances for Mg^{2+} (2.45- and 2.07-Å, respectively[42]), it is possible that the active site conformation observed in the Mg^{2+}/Sr^{2+}-complex represents a strong Mg^{2+}-binding site. However, it is also possible that the C47U mutation, the Ca^{2+}-dependent crystallization process, or the constraints of the crystalline environment in general have weakened a more ideal active site Mg^{2+}-binding pocket. Moreover, that the A29 pro-*Sp* oxygen would play such a prominent role in both preligation structures–directly coordinating Ca^{2+} and being as close to Mg^{2+} as is the pro-*Rp*–is something of a curiosity, since phosphorothioate nucleotide analog inteference mapping (NAIM) implicated the A29 pro-*Rp* oxygen as being functionally significant[34]. However, as NAIM cannot interrogate pro-Sp oxygens[43,44], we have no direct data about the role it plays in catalysis. Likewise, the A29 pro-*Rp* interference may be caused by steric effects, or the conformations observed in our structures may be crystallographic artifacts.

Of note, in neither preligation structure is the central active site metal ion close enough to the 5′-α-phosphate nonbridging oxygens to form an inner-sphere coordination contact. Likewise, based on these structures it seems unlikely that would such a contact be formed at the transition-state. Hence this metal differs from the canonical Metal A of proteinaceous polymerases, which, in addition to activating the reaction nucleophile, is thought to stabilize the developing negative charge at the α-phosphate during the transition state[45,46]. It has been proposed that a similar contact is being made in the active site of the ligase, since the ribozyme is significantly inhibited by phosphorothioate (oxygen-to-sulfur) substitution at the 5′-pro-*Rp* position[32]. Formation of such a contact would require only minor adjustments to the conformation of the 5′-nucleotide, and may still be possible. It is also possible that the observed

phosphorothioate inhibition is caused by ablation of an outer-sphere contact[47], which could be easily accommodated in either structure.

In neither structure do we observe a metal ion analogous to the canonical Metal B. In proteaceous polymerases this metal is positioned ~4-Å from metal A and ligands the nonbridging oxygens on the α- and β-phosphates[45,46]. While metal ions are bound the β-phosphate oxygens in both preligation structures (**Fig. 3A–B**), they are at distances of 6.5–7.0-Å from the Metal A analog, and bind with geometries drastically different from Metal B. In the Mg^{2+}/Sr^{2+}-complex, the metal ion bound to the β-phosphate coordinates a well-ordered water molecule (**Fig. 3B**). This water may be of functional importance, since is within hydrogen bonding distance of the triphosphate [α,β]-bridging oxygen, the reaction leaving group.

Additional metal ions and water molecules decorate the 5′-triphosphate (summarized for the Mg^{2+}/Sr^{2+}-complex in **Fig. 3C**), and may function structurally to stabilize its conformation. Of these, the most prominent is a partially hydrated metal ion bound to the Hoogsteen face of G1 and G2. Metal hydrate clusters are observed near this position in each of the preligation structures, and are reminiscent of a similar fully hydrated Mg^{2+} observed in the product[35]. Each of these metal hydrate complex binds with a different geometry, and in the preligation complexes this geometric variation appears to have an intimate effect on the conformation of the adjacent 5′-triphosphate, as described below.

Conformation of the 5′-triphosphate

In each of the preligation structures, the 5′-triphosphate forms a "hooked" conformation, curling at nearly a right angle as it docks the β- and γ-phosphates just above the G1 base (**Fig. 3A–B**). This conformation is dramatically different from that adopted by (d)NTPs in the active sites of

proteinaceous polymerases, wherein the triphosphate is pulled away from the plane of the adjoining base[48]. Moreover, the (d)NTPs in these structures stereotypically present polydentate Mg^{2+} binding sites, jointly coordinating metal ions with two or more of their phosphate groups[49]. In contrast, the 5′-triphosphate in the Class I ligase forms no such polydentate binding sites, though two metals are bound via bridged interactions with other functional groups in the ribozyme core. The only bidentate interaction is to a water jointly coordinated by the 5′- and [β,γ]-bridging oxygens in the Mg2+/Sr2+-complex structure (**Fig. 3C**).

In the Mg^{2+}/Sr^{2+}-complex structure, docking of the 5′-triphosphate places its α-phosphate 3-Å from the substrate 3′-hydroxyl. The angle formed by the 3′-hydroxyl nucleophile, α-phosphorus electrophile and [α,β]-bridging oxygen leaving group is 176º, extremely close to the ideal value of 180º adopted during in-line nucleophilic attack. Moreover, pulling the β- and γ-phosphates proximal to the G1 base exposes the [α,β]-bridging oxygen, presenting it towards potentially stabilizing functional groups in the active site. In the Ca^{2+}/Sr^{2+}-structure, however, the α-phosphate is positioned 3.8-Å from the 3′-hydoxyl nucleophile. Concomitant rotation of the β- and γ-phosphates towards the major groove contorts the angle between nucleophile, electrophile and leaving group to 154º, and skews the leaving group oxygen away from the active site (**Fig. 3D**).

The relative positioning of the 5′-triphosphate may be explained in part as a consequence of interactions made with the metal hydrate cluster near G2. In the Mg^{2+}/Sr^{2+}-complex, the bound $Mg^{2+}\bullet 4H_2O$ cluster is shifted away from the plane of the G2 base, pushed downward towards G1. This orientation facilitates close approach by the 5′-γ-phosphate, which joins G1 and G2 in liganding the metal cluster (**Fig. 3B–C**). If protonated, the γ-phosphate would also be able to hydrogen bond with the G2 N7. In the Ca^{2+}/Sr^{2+}-complex, the G2 O6 provides a direct

inner-sphere ligand for bound the Ca^{2+}-hydrate cluster (modeled as Ca^{2+}•$5H_2O$, **Methods**). Pulling this bulky complex close to the plane of the G2 base occludes the G2 Hoogsteen face, and as a consequence, the γ-phosphate cannot form the metal ion contact observed in the Mg^{2+}/Sr^{2+}-complex.

It hence seemed plausible that contacts to the G2 base may aid in positioning the leaving group and might therefore contribute to catalytic efficiency. To test this hypothesis, we compared the efficiency of a primer-extending (PE) ligase variant in which nucleotides G1–A4 have been removed (construct "t307"), to one that lacks G1 but retains G2–A4 ("GAA-t307, **Fig. 1A**, inset). Neither construct can self-ligate, and both can catalyze single-nucleotide extension of primers *in trans*, using exogenous GTP as a substrate[21,32]. At pH 7.0, the t307 construct exhibited a k_{cat}^{GTP} of 0.10min^{-1}, a K_M^{GTP} of 3.5 mM and a k_{cat}/K_M^{GTP} of 30 $M^{-1}min^{-1}$, consistent with previous observations[32]. Restoring residue G2–and hence the potential to form the solvent-mediated G2-triphosphate interaction–resulted in modest improvements in Michaelis-Menten parameters (**Fig. 3E**). The GAA-t307 construct exhibited a k_{cat}^{GTP} of 0.11min^{-1}, a K_M^{GTP} of 2.8 mM and a k_{cat}/K_M^{GTP} of 40.8 $M^{-1}min^{-1}$, an overall improvement in catalytic efficiency of ~36%. It should be noted that restoring the G2 base to the GTP substrate, rather than to the enzyme, results in a more substantial (36-fold) improvement in k_{cat}/K_M^{GTP} (ref. 32). Moreover, since G2 caps the P2 helix by base pairing with A11, it is possible that restoration of G2 benefits the enzyme by decreasing its conformational flexibility, rather than by improving GTP-binding.

The C47 nucleobase participates in catalysis

In the ligase product structure, the C47 N4 exocyclic amine is positioned 3.1-Å from the ligation junction pro-*Sp* oxygen[35]. It was therefore speculated that during catalysis C47 might stabilize

the leaving group oxygen through hydrogen bonding, a role similar to that played by the active site general acid of proteinaceous polymerases[35,50]. Observing the structure of the Mg^{2+}/Sr^{2+} preligation complex, this model now seems unlikely: while the U47 O4 is positioned 3.7-Å from the 5′-α-phosphate pro-*Rp* oxygen, it is 4.6-Å away from the leaving group oxygen. In the absence of a substantial conformation change during catalysis, the relative orientation of these groups is not expected to change. Rather, as the ligation reaction follows its trajectory, the C47 amine and α-phosphate oxygen will be brought in closer proximity, strengthening the hydrogen bond between them. Hence, C47 might function not by stabilizing the developing negative charge on the leaving group, but by stabilizing the transition state geometry, a role similar to that of the canonical Metal B in proteinaceous polymerases[45,46]. If, however, the ribozyme undergoes a conformation change during ligation, C47 might still be brought into close proximity with the leaving group oxygen. In this context, it could function as an electrostatic or a general acid catalyst.

To distinguish between these possible functions, we sought to biochemically dissect the role C47 plays during catalysis. Insight into the mechanism of natural self-cleaving ribozymes has been gleaned from exogenous base rescue experiments, wherein the nucleobases at critical active-site positions are removed, and the ability of exogenous compounds to rescue these abasic variants is examined[51-53]. To test if such a strategy would work for the ligase, we used chemically modified oligonucleotides and DNA-splinted ligation to assemble a ligase construct bearing an abasic residue at position 47 ("C47Abasic,"). Self-ligation activity of this construct is reduced by $>10^5$-fold relative to the unmodified ribozyme. However, in the presence of exogenous cytosine, activity of the C47Abasic construct was restored 33-fold, while a parallel experiment rescued the C47U mutant only by ~2.2 fold (**Fig. 4A**). Therefore, of the rescue

observed with the C47Abasic construct, ~15-fold can be attributed to the exogenous base functionally mimicking C47[51]. Ideally, this result could be followed by a systematic screen of additional exogenous compounds, from which the specific cytosine functional groups could be interrogated, or analyzed by Brønsted analysis[54]. However, none of the cytosine analogs we tested, nor the general acid-base catalyst imidazole, could specifically rescue the C47Abasic ligase (**Fig. 4A**).

Given the intractability of further exogenous rescue experiments, we instead turned to traditional and so-called atomic mutagenesis strategies to assess the catalytic contribution of functional groups on the C47 base. Replacing C47 with a series of natural or modified bases generated ligase variants with activities spanning six orders of magnitude (**Fig. 4B**). The activity of some variants was surprising. For example, in the C47A mutant, activity was reduced a modest 30-fold relative to the wild type, a far weaker effect than that observed when a critical active site cytosine in the HDV ribozyme is mutated to adenosine[52,55]. Adenosine retains amine and imine groups in a configuration similar to that of cytosine, and hence the activity of this variant implies that the C47 O2 carbonyl does not contribute to catalysis. Moreover, unlike the HDV ribozyme, the ligase active site is malleable enough to accommodate a purine at this critical position. Another curious observation came from a variant in which C47 is replaced with 4-thiouridine (^{4S}U). This construct exhibited biphasic kinetics with rates diminished 4.4-fold and 10^4-fold, relative to the wild type. While the slow phase might be expected, given the similar characters of uridine and ^{4S}U, we offer no explanation of the remarkably fast burst phase. Most other variants exhibited activities that were less controversial. Relative to the wild type, the C47U and C47G mutants were diminished by $>10^4$ and $>10^5$, respectively, while deleting residue 47 altogether ("C47Δ") reduced activity to nearly background levels.

While these constructs illustrate the importance of the C47 nucleobase, they alone do not report the catalytic contributions of individual functional groups on this base. In particular, many of these alterations simultaneously changes the protonation and tautomeric states of groups analogous to cytosine N3 and N4, and hence do not discriminate the effect of individually altering either group. To more precisely investigate the contribution each functional group makes to catalysis, we incorporated the cytosine analogs zebularine (which lacks the N4 amino group but is otherwise identical to cytosine) and pyridin-2-one (which lacks both the cytosine N4 and N3 groups) at position 47. Activities of these zebularine ("C47Zeb") and pyridin-2-one ("C47P2o") variants were 635-fold and 175-fold diminished, respectively, relative to that of the unmodified ribozyme. These data are revealing in several regards. First, they demonstrate that the C47 N4 amine, and not its N3 imine, participates in catalysis, since ablation of the former group is strongly inhibitory, but further ablation of the latter group is, if anything, beneficial. Second, the data quantitatively assess the catalytic contribution of the C47 N4 amine: an ~200-fold drop in rate is more than might be anticipated from the loss of a single hydrogen bond, but less than that expected from the loss of a general acid catalyst. Finally, since the C47U variant is >500-fold slower than C47P2o, loss of activity in the C47U mutant must in part be due to the addition of an inhibitory group, and not merely the ablation of an enhancing group. It seems probable that the electronegative U47 O4 carbonyl inhibits a role normally played by the electropositive C47 N4 amine; in the C47P2o construct, this role could be performed *in trans* by an exogenous water molecule.

A single proton transfer during the ligase transition state

We next sought to determine the specific catalytic function of the C47 N4 amine. This group might act electrostatically, through the donation of hydrogen bonds, or as a general acid catalyst, through the donation of protons. To distinguish between these models, we first examined the pH dependence of self-ligation activity by the C47P2o construct. Were the C47 N4 amine a proton shuttle, then its removal would ablate a titratable functional group and hence alter the dependence of ligase catalysis on pH. Between pH 6.0 and 8.5, self-ligation activity of the C47P2o mutant exhibited log-linear pH dependence with a slope of one (**Fig. 4C**), identical to the behavior observed for the wild type ligase[33]. For both ribozymes, the deviation from log-linearity at high pH is attributed to deprotonation of the nucleobases, which would presumably destabilize structure. Since data at lower pH values were collected under conditions in which chemistry is rate limiting[26,33], a log-linear pH dependence implies that a single deprotonation event ultimately determines the rate of ligase catalysis. This deprotonation has been previously attributed to the activation of the 3′-hydroxyl nucleophile, and the C47P2o data further support this claim. Removal of the N4 group, a different potential proton shuttle, does not affect the pH dependence.

To more directly measure the number of protons in transit during the transition state, we performed proton inventory experiments[50,56] on the wild type ligase and some of its derivatives (**Methods**). The proton inventory technique is a solvent isotope effect–experiment in which the dependence of the ligation rate constant on the mole fraction of D_2O is measured. The order of a polynomial that best describes this dependence corresponds to the number of protons transferred during the transition state: a linear fit implies one-proton-transfer, while a second-order polynomial implies two-proton-transfer, and so forth. Data for the wild type ligase fit well to a

line with a slope (Φ) of 0.19±0.01 (**Fig. 4D**, *left; solid line*). Though a linear proton inventory could be interpreted as resulting from a complex agglomeration of many microscopic steps[56], or from a rate-limiting folding step[57], the most parsimonious interpretation of the present data is that the ligase transfers a single proton during catalysis. Moreover, since the fit is linear, the inverse of this slope (1/Φ) corresponds to the solvent deuterium kinetic isotope effect (SDKIE)[56]. The unmodified ligase therefore exhibits a SDKIE of 5.2±0.3, comparable to those observed for some natural proteinaceous polymerases[50,58] and consistent with a transition state that strongly depends on proton transfer[56]. The best-fit two-proton-transfer model for these data is not substantially different, yielding SDKIE values of 4.1±0.4 and 1.15±0.1 for the individual protons transferred (**Fig. 4D**, *left; dotted line*). An SDKIE of 1.0 signifies a complete lack of rate-determining proton transfer[56]. Therefore, were the ligase to shuttle two protons during its transition state, the second of these would be a weak determinant of the overall rate, and hence a minor component of its catalytic mechanism.

In a similar manner, proton inventories for the C47P2o and C47U variants could only be convincingly fit to single-proton transfer models, yielding SDKIE values of 2.3±0.1 and 2.9±0.1, respectively (**Fig. 4D**, *middle and right*). Hence, neither the removal of the C47 N4 amino group, nor its replacement with an inhibitory carbonyl group changes the number of protons in transit during ligation. It is notable that each of these variants exhibits an SDKIE lower than that of the wild type, and might imply that the transition states they stabilize depend less strongly on proton transfer. The data might therefore support the model in which C47 functions electrostatically, stabilizing the 5′-α-phosphate transition state geometry by donation of a hydrogen-bond.

The C30 2′-hydroxyl participates in catalysis

In addition to its proximity to the ligation junction, in our product structure the C47 N4 amine was located within hydrogen-bonding distance of the C30 2′-hydroxyl[35]. This interaction is in part a consequence of the metal bound by the P6 A-minor triad, which juxtaposes the two groups by pulling the C30 ribose into an unusual 3′-exo conformation[59]. During crystallographic refinement, attempts to fix C30 in the more standard 3′-endo conformation resulted in higher R_{free} values and characteristic difference peaks in $|F_o| - |F_c|$ maps (*data not shown*). In the Mg^{2+}/Sr^{2+}-complex preligation structure, we observed that the C30 2′-hydroxyl is also 3.5-Å from a well-ordered water molecule positioned 2.7-Å from the leaving group oxygen (**Fig. 5A**). Given these observations, we sought to explore whether or not the C30 2′-hydroxyl plays a direct role in catalysis.

Using DNA-splinted ligation, we assembled a ligase variant bearing a 2′-Deoxyribose at residue 30 ("2′-Deoxy C30"). This alteration is not expected to be structurally disruptive, since the 2′-endo conformation typically adopted by 2′-deoxyribonucleosides is quite similar to the 3′-exo conformation C30 adopts in our crystal structures[59]. Self-ligation activity by the 2′-Deoxy C30 variant was reduced ~18-fold relative to the wild type. This construct exhibited a linear proton inventory with a measured SDKIE of 3.1±0.2, a value intermediate between those measured for the wild type and C47 variants (**Fig. 5B**). These data imply that the C30 2′-hydroxyl contributes to catalysis, but that it is not involved in proton-transfer.

Therefore, during catalysis the C30 2′-hydroxyl–much like the C47 N4 amine–may function electrostatically, acting as a hydrogen-bond acceptor, a donor, or both. To discriminate between these possible roles, we generated ligase variants in which the C30 2′-hydroxyl was replaced with isosteric 2′-fluoro or 2′-amino groups. A 2′-fluoro group could function as a

hydrogen bond acceptor but not a donor; the inverse is true of a 2′-amino substitution. We first examined the effects of these alterations within the context of the primer-extending (PE)-ligase construct, t307 (**Fig. 1A, 5C**). The 2′-deoxy C30 alteration reduced PE-activity ~113-fold, a more dramatic effect than that observed in the self-ligase, perhaps implying that the C30 2′-hydroxyl plays an additional role when binding exogenous NTPs *in trans*. The 2′-fluoro substitution diminished activity even further, ~2200-fold relative to the wild type. This effect parallels that observed in the C47U and C47P2o variants: removal of a functional group allows its replacement by water, but an isosteric, electronegative substitution is severely inhibitory. In contrast, the 2′-amino substitution restored primer extension activity to a level exceeding that of the 2′-deoxy variant, being only ~30-fold below that of the wild type (**Fig. 5C**).

We next repeated these experiments in the context of the self-ligase. While the absolute influence on activity was less dramatic than that observed in the PE-ligase, in the self-ligase this series of 2′-modifications exhibited the same relative behavior: the 2′-deoxy variant was slow, the 2′-fluoro variant slower still, and the 2′-amino group restored activity (**Fig. 5D**, *left*). Together, these data imply that, in the context of either self-ligation or primer-extension, the C30 2′-hydroxyl functions as a catalytically important hydrogen bond donor.

Interactions between active-site functional groups

Given their proximity, we reasoned that the C47 N4 and C30 2′-hydroxyl groups might functionally interact during catalysis. For example, were the C47 N4 to donate a critical hydrogen bond to the C30 2′-hydroxyl, then ablation of the C30 2′-hydroxyl should produce a different effect in the presence and absence of the C47 N4 amine, and *vice versa*. To explore this possibility, we generated C47P2o and C47U self-ligase variants bearing the same series of C30

2′-modifications described above. Each series of variants exhibited the same overall trend in behavior exhibited in the wild type ligase. Namely, ablating the C30 hydroxyl was detrimental, and this loss of activity could be exacerbated by its replacement with a 2′-fluoro group, or rescued by its replacement with a 2′-amino group (**Fig. 5D**, *middle and right*). However, the magnitude of these effects depended on the identity of the C47 base. In the C47P2o variants, which lack the C47 N4 amine, C30 2′-deoxy and 2′-fluoro substitutions resulted in a loss of activity ~10–20-fold more severe than in the wild type C47 background. Likewise, rescue by the C30 2′-amino group was ~10-fold less effective in the C47P2o background than it is in the context of wild type C47 (**Fig. 5D**, *middle*). The opposite effect is observed in the C47U variants, which replace the N4 amine with an inhibitory carbonyl group. Here, activity of the C47U C30 2′-fluoro variant was only 5.9-fold reduced below that of the unmodified C47U ligase, while the C30 2′-amino modification restored activity to ~82% that of C47U (**Fig. 5D**, *right*).

Were the C47 and C30 functional groups acting independently, then the effect of altering one would be independent of the status of the other. The data indicate that this is not the case, as evidenced by the schema in **Fig. 5E**. The left scheme summarizes the reduction in self-ligation rates incurred through the sequential removal of activating groups: the C47 N4 amine and the C30 2′-hydroxyl. The right scheme summarizes the reduction in self-ligation rates incurred by the sequential replacement of active site functionalities with inhibitory groups: the U47 O4 carbonyl and the C30 2′-fluoro.

Comparing the relative effects of single and double deletions (*i.e.,* by comparing horizontal or vertical arrows to one another) in the inactivation scheme, we observed that ablation of one functional group becomes six-fold more detrimental when the other functional

group has already been removed (**Fig. 5E**, *left*). In contrast, in the deactivation scheme, we observed that the addition of one inhibitory group becomes 50-fold less detrimental when the other group has already been replaced (**Fig. 5E**, *right*). Thus, in the absence of the C30 2′-hydroxyl, catalysis becomes six-fold more dependent on the presence of the C47 N4 amine, and *vice versa*. Conversely, when catalysis is quelled by the addition of an inhibitory group at either position, it becomes 50-fold less dependent on the function of the other. Taken together, these data imply that the C47 N4 exocyclic amine donates a hydrogen-bond to the C30 2′-hydroxyl, which itself must furthermore donate a hydrogen-bond during catalysis.

Discussion

Based on past[34,35] and present results, we here propose a model for transition state stabilization by the Class I ligase (**Fig. 6**). In this model, the A29 and C30 nonbridging phosphate oxygens coordinate a catalytic Mg^{2+} cofactor that activates the 3′-hydroxyl for nucleophilic attack. A binding site for this metal was implied by the previous biochemical data[34,35] and the metal itself is observed in the preligation crystal structures (**Fig. 3A–B)**. Furthermore, two RNA functional groups–the C47 N4 exocylic amine and the C30 2′-hydroxyl–participate in networks of hydrogen bonds that stabilize both the transition state geometry and the leaving group. The C47 amine contributes directly to geometric stabilization, forming a hydrogen bond with the 5′-α-phosphate pro-*Rp* nonbridging oxygen that grows stronger as the reaction proceeds from the ground state to the transition state. This proposition is supported by atomic mutagenesis experiments, which clearly demonstrate that catalysis depends on the partial positive charge presented by the C47 N4 group (**Fig. 4B**). In addition, these data suggest that C47 participates in a second critical hydrogen bond with the C30 2′-hydroxyl (**Fig. 5D–E**), a group that itself contributes a critical partial positive charge during catalysis (**Fig. 5C**). We propose that these interactions aid in leaving-group stabilization, positioning a water molecule near the [α,β]-bridging oxygen. This water can be modeled in the Mg^{2+}/Sr^{2+}-complex structure (**Fig. 3B–C**), although direct interpretation of this model should be limited, given the resolution limit of our data.

The C47 and C30 functional groups are proposed to act through hydrogen bonding, and not through proton transfer, because the proton inventory is unchanged by the removal of either group, and because ablation of the C47 N4 amine has no effect on pH-dependence (**Fig. 4C–D, 5B**). In contrast, removal of a general acid catalyst from the active sites of proteinaceous

polymerases changes both of these properties[50]. In our model, then, we propose that a network of hydrogen bonds confers a partial positive charge on the bound water molecule, allowing it to stabilize the developing negative charge on the leaving group. In theory, it is also possible that interactions between C47 and C30 might allow this water to function as a specific acid that directly protonates the leaving group. Considering the kinetic isotope data, specific acid catalysis of this sort could not happen in concert with nucleophilic attack, nor could it be rate limiting.

It has previously been hypothesized that the ligase employs a two metal-ion mechanism similar to those of proteinaceous polymerases and the self-splicing introns[33,35]. We now know that the ligase mechanism bears some resemblance to its natural counterparts, since it uses an acid-bound Mg^{2+} cofactor to activate the 3′-hydroxyl for nucleophilic attack. This metal is similar to the canonical Metal A of proteinaceous polymerases. However, the two are not directly equivalent, since, in addition to activating the nucleophile, Metal A is known to interact with the α-phosphate nonbridging oxygen[45,46,60]. An analogous contact is not observed in the ligase structures, though it is supported by phosphorothioate interference data[32]. It is therefore possible that a small conformation change facilitates this contact during catalysis, or that the interference data are caused by an outer-sphere coordination effect.

In almost all other regards, the Class I ligase seems to have found a catalytic strategy dramatically different from its natural counterparts. Notably, the ribozyme appears to lack Metal B[45,46], a universal feature of the active sites of natural polymerases that is not paralleled by any of the numerous solvent atoms decorating the ligase 5′-triphosphate (**Fig. 3A–C**). In proteinaceous polymerases, Metal B is thought to enter the active site in complex with incoming NTPs and exit in complex with the pyrophosphate product; during catalysis it helps stabilize the

transition state geometry and charge distribution at the α-phosphate[45,46]. Its absence in the ligase may be linked to the unusual conformation adopted by the G1 5′-triphosphate, which would provide a poor binding site for this metal. Since RNA is more adept at forming metal ion binding sites than is protein, it should be substantially easier for an evolving ribozyme to explore different potential ways of using metal ion interactions to stabilize an NTP substrate. As such, the Class I ligase has apparently been evolved to strip the 5′-triphosphate of Metal B, contorting it into the hooked geometry and stabilizing it through metal-mediated interactions with other groups on the enzyme. This triphosphate conformation concomitantly pulls the negatively charged β- and γ-phosphates away from the reaction center, exposing the leaving group oxygen toward catalytic moieties at the active site.

That the ligase might accelerate phosphoryl-transfer via a one-metal-ion mechanism, and that functional groups on the enzyme perform roles typically played by a second metal ion, is not without precedent. A similar proposition has been made in the case of ββα-me and HUH nucleases, which use an active site metal cofactor near the 3′-hydroxyl leaving group, but which lack one near the 5′-phosphate[61]. A comparison of these enzymes to those that use a two-metal-ion strategy led to the conclusion that active-site amino acids have supplanted the role normally ascribed to the absent metal. Similarly, the ligase appears to have been replaced this metal with the C47 base. While the use of catalytic nucleobases is now known to occur in the active sites of natural catalytic RNAs, the ligase C47 is the first example of this phenomenon in an artificial ribozyme.

Although crystal structures have been solved of other artificial ribozymes in their postcatalytic states[37,39], as apoenzymes[36,38], or bound to their substrates[38] or products[36], the Class I ligase is the first artificial ribozyme for which two steps of a catalytic cycle have been

visualized crystallographically. This allows us to compare the proposed ligase mechanism to those of its natural counterparts, which have been similarly analyzed (reviewed in ref. 62). In many regards, the most direct comparison is with the HDV ribozyme, which also uses a metal ion cofactor to deprotonate its nucleophile and an active site cytosine to stabilize the leaving group[63]. In HDV, however, this cytosine almost certainly functions as an active site acid[62,64], having an N3 group with a pKa near neutrality[65,66]. In contrast, the ligase uses a cytosine N4 amine, and does not perform general acid-base catalysis. Like the ligase, the GlmS ribozyme is thought to make catalytically critical hydrogen bonds with a nucleobase exocyclic amine (the G57 N2 amine), and a hydroxyl group (the Glucosamine-6-Phosphate C1-hydroxyl)[67]. However, both of these groups are thought to stabilize the scissile phosphate transition state geometry, and not the leaving group. Finally, the ligase's extensive use of electrostatics is reminiscent of the Hairpin ribozyme, which accelerates self-cleavage by forming more hydrogen bonds with the scissile phosphate at the transition state than it does at either ground state[68]. A water molecule might also play a catalytic role for the Hairpin[69], though it is proposed to function as a specific base, activating the reaction nucleophile. Moreover, unlike the ligase, the Hairpin ribozyme catalysis does not depend on a metal ion catalytic cofactor[70].

Compared to the Class I ligase, most artificial ribozymes achieve far lower rate enhancements, and are likewise thought to achieve catalysis by substantially simpler means[36-39]. Hence, it is tempting to propose that the relatively complicated catalytic strategy of the Class I ligase (**Fig. 6**) might be tied to its remarkable catalytic power. Further exploration of the ligase mechanism may facilitate a greater understanding not only of the differences between natural and artificial ribozymes, but of RNA catalysis in general. Likewise, the extension of these

findings to the ligase-derived polymerase ribozymes[22-24] may help in future efforts to isolate an RNA replicase ribozyme.

Acknowledgements

We are deeply indebted to F. Eckstein for the donation of reagents, R.A. Grant, D. Lim, T. U. Schwartz and K.R. Rajashankar for their assistance with data collection and processing, H. Mackie, E. B. Roesch and J. De Luca for their advice on oligonucleotide synthesis, J. Chen, A. Ricardo, K. Frederick, N. Yoder and W. K. Johnston for their help with HPLC purification, E. Spooner for performing mass spectrometry, U. L. RajBhandary, C.L. Drennan, J. A. Piccirilli, J. W. Szostak and members of the Bartel laboratory for thoughtful comments and stimulating discussions.

Materials and Methods

Sample synthesis for crystallography. Except for the C47U active site mutation, the crystallization construct used in these experiments was identical to the improved Class I ligase U1A-binding variant described previously[35]. The transcription template for this mutant ribozyme, p307HU_C47U, was generated from the wild type template p307HU[35] using the QuickChange mutagenesis kit (Stratagene), and linearized by *EarI* digestion before use. RNA was synthesized in 1–3 mL T7 *in vitro* transcription reactions, processed and purified essentially as described previously[35], but without removal of the terminal 2′-3′ cyclic phosphate. Purified, desalted RNA was concentrated to ~200 μM in deionized water and stored at –20ºC before use.

The ligase substrate oligonucleotide was purchased from Dharmacon and deprotected according to the manufacturer's specifications. To eliminate contaminants remaining from solid phase synthesis, the sample was further purified on a preparative denaturing 0.5x TBE 20% polyacrylamide gel. The corresponding band was visualized by UV-shadowing and excised. RNA was passively eluted into 300 mM NaCl overnight at 4ºC, precipitated with five volumes of ethanol, dried, resuspended into deionized water and stored at –20ºC before use.

The U1A A1-98 Y31H/Q36R double mutant[71,72] was expressed and purified as described previously[35].

Crystallization and data collection. The annealing protocol used to prepare RNA samples for crystallization was designed to parallel that used during the original ligase selection[18] and in subsequent biochemical experiments[21,26,32-34]. Approximately 200 μM unreacted C47U ligase RNA was heated (80ºC, 5 minutes) and cooled (22ºC, 10 minutes) in water, and then mixed with a 1.1-fold excess of substrate oligonucleotide in calcium annealing buffer (final composition, 5

mM MES, pH 6.0, 10 mM $CaCl_2$, 1 mM DTT), and incubated at 22ºC for an additional 15 minutes. U1A was added at a 1:1 molar ratio of ligase:protein, bringing the final complex concentration to 5.0 g/L. This mixture was incubated at 22ºC for 45–60 minutes, mixing periodically, and then centrifuged at 13,000xg for 1 minute prior to setting up crystallization experiments.

Experiments using the C47U construct and Mg^{2+}-containing annealing and precipitation buffers, similar to those reported for ligase product crystals[35], failed to yield crystalline matter. Moreover, wild type ligase•substrate complexes in which the nucleophile was blocked with a 3′-O-methyl group, or in which the 5′-triphosphate was replaced with a nonhydrolyzable α,β-methylene triphosphate, failed to produce crystals under any condition tested (*data not shown*). Crystals could only be obtained from the unreacted C47U mutant in Ca^{2+}-containing buffers. Initial crystals were grown by hanging drop vapor diffusion, mixing equal volumes of the C47U ligase•substrate•U1A complex sample with mother liquor consisting of 14% Methyl-2,4-pentane-diol (MPD), 50 mM sodium cacodylate (pH 6.0), 20 mM calcium acetate, 10 mM strontium acetate and 1mM spermine, and equilibrating over 0.6 mL of this same precipitation mixture at 20ºC. Crystal clusters with an interdigitated cubic morphology appeared within a week, and reached maximum size (~100 μm per side) within two weeks.

Diffraction quality crystals were grown using a previously developed microseeding technique[35] that exploits the ability of KCl to ablate *de novo* crystal nucleation. Crystalline clusters were stabilized by progressively increasing the drop concentration of MPD to 30%, holding all other buffer components isotonic. To generate seed stocks, stabilized clusters were crushed with a Seed-Bead (Hampton Research), serially diluted in the same stabilization buffer, and used directly as precipitant in subsequent experiments. To avoid spontaneous nucleation in

these experiments, RNA samples were prepared in calcium annealing buffer supplemented with 100 mM KCl. This ligase•substrate•U1A sample was mixed in equal volumes with the microseed stock, and equilibrated in hanging drops over 0.6 mL of 20–24% MPD. Morphologically homogeneous crystals grown by this method appeared overnight, and were harvested within four days.

For cryoprotection, crystallization drops were brought to 30% MPD under otherwise isotonic conditions, as described for the generation of seed stocks. Calcium-bound crystals were allowed to equilibrate in stabilization buffer for two hours, mounted in nylon loops and plunged directly into liquid nitrogen. To obtain magnesium-bound crystals, drops containing cryostabilized calcium-bound crystals were serially diluted with an isotonic 30% MPD solution in which Mg^{2+} replaced Ca^{2+}. After eight twofold dilutions into this magnesium stabilization buffer (final Mg^{2+}:Ca^{2+}, 256:1) the crystals were equilibrated for two hours more, mounted in nylon loops and plunged into liquid nitrogen. Since the estimated ligation rate of the C47U mutant is ~1.1×10^{-5} min^{-1} at pH 6.0, we anticipate ~0.1–0.3% should have ligated during equilibration with Mg^{2+}.

Data collection, structure determination and refinement. Data were collected at NE-CAT beamlines 24-ID-C and 24-ID-E at the Advanced Photon Source (APS), aided by the expert technical advice of K.R. Rajashankar. All data were indexed, scaled and integrated using the HKL2000 software suite[73], removing 5% of all reflections for R_{free} calculations. The Ca^{2+}/Sr^{2+}- and Mg^{2+}/Sr^{2+}-complex structures were solved, built and refined independently of one another. The two monomers in each asymmetric unit were built separately.

Initial phases were obtained by molecular replacement in PHASER [74], using the U1A-bound Class I ligase product structure (PDB ID: 3HHN) as a search model. The optimal solutions for each search (Z-scores of 26–30) were used as the starting points for further refinement, carried out in PHENIX[75]. Model building was performed in COOT[76]. All refinement steps followed the "individual ADP" strategy in PHENIX[75], alternating rounds of automated bulk-solvent correction, positional and individual atomic B-factor refinement. Target geometric weights (the "wxc_scale" in PHENIX[75]) were held at relatively restrictive values of 0.05–0.01; the "wxu_scale" was held at 1.0 throughout all refinement steps. To remove model bias from the molecular replacement solutions, the first three rounds of refinement also employed rounds of simulated annealing (5000K to 300K, in 100K steps) and rigid body refinement, defining the individual ligase helices (P1, P2, etc…), the U1A protein, its cognate loop, and 3–6-nt subdivisions of the unpaired joining regions (J1/2, J1/3 and J3/4) as independent rigid bodies. Neither NCS-averaging nor TLS was used during either refinement.

The $|F_O| - |F_C|$ difference fourier maps resulting from initial refinement rounds exhibited prominent (7–10σ) peaks in the vicinity of nucleotide G1 (**Fig. 1B–C, 3A–B**), corresponding to the 5′-triphosphate. This 5′-GTP was built by hand in COOT[76], and the resulting model was subjected to another round of simulated annealing in addition to the ADP regimen outlined above. Fully- or partially-hydrated metal ions outside the active site were built next, their degree of hydration assigned by first placing dehydrated metals into strong (>3.5σ) $|F_O| - |F_C|$ difference peaks, and inspecting the resultant temperature factors and difference peaks after a subsequent round of refinement[77,78]. This process was repeated iteratively, until no strong peaks remained. Active site metal ions were placed last, using this same strategy.

Owing to the resolution limits, metal ions and their bound waters could not be individually refined in a chemically sensible manner. So, partially- or fully-hydrated metal clusters were treated as monomers during refinement, having defined model residues with idealized bond lengths and geometries using the SKETCHER program the CCP4 program suite [79]. Magnesium–water bond distances were defined as 2.07 Å; hexacoordinate geometry was used in for all magnesium monomers [41,42]. Calcium–water bond distances were similarly fixed at 2.43 Å. However, owing to the variable coordination geometries of calcium hydrate clusters[41,42], bound waters were not built unless their geometry with respect to the calcium ion was clearly indicated in difference Fourier maps. For the metal-hydrate clusters bound to the P6 A-minor triad (residues A31 and A32 [35]) bond distances were restrained to their ideal values, and the angle was fixed at 90º; none of the other metal hydrates were fixed with regard to the neighboring macromolecule.

Each monomer in the final Mg^{2+}-bound structure contains all 130 nucleotides of the ligase core, all seven nucleotides of the substrate, and all but the first 6–7 amino acids of the U1A protein. The occupancies have been set to zero for protein side chains lacking electron density. Despite high temperature factors in the Ca^{2+}/Sr^{2+}-complex structure (**Fig. 2C**), refining the occupancies of, or altogether removing problematic nucleotides resulted in an increase in the R_{free} value. Hence final structure consists of all nucleotides of the ligase core and substrate.

All structural figures were made in PyMol[80].

Synthetic RNA and DNA oligonucleotides. Most of the RNA oligonucleotides used for the splinted assembly of modified ligases (*see below*) were purchased from Dharmacon. Oligonucleotides bearing the pyridin-2-one and abasic modifications were purchased from Trilink Biotechnologies. Those bearing zebularine, 2′-deoxy isoguanosine and 2′-deoxy 5-methyl isocytosine modifications were purchased from the W.M. Keck Foundation Biotechnology Resource Laboratory at Yale University (hereafter "Yale Keck"), using

phosphoramidites from Glen Research. DNA transcription template oligonucleotides bearing 5′-terminal 2′-O-methyl modifications (*see below*) were also synthesized by the Yale Keck facility. All other DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT).

The reagents from Trilink and IDT did not require deprotection or further purification before use. DNA oligonucleotides from Yale Keck were purified from denaturing 10% polyacrylamide gels as described for the RNA substrate oligonucleotide. RNA oligonucleotides from Yale Keck were deprotected with 55.6% (v/v) triethylamine trihydrofluoride (TEA.3HF) in anhydrous DMSO for 2.5 hr at 65ºC. Reaction mixtures were then briefly chilled, brought to 300 mM sodium acetate and precipitated by addition of five volumes of isopropanol and overnight incubation at –20ºC. Following centrifugation, pellets were washed with absolute ethanol and pelleted twice, then dried and resuspended into deionized water. Oligonucleotides from Dharmacon were deprotected according to that manufacturer's instructions, and in biochemistry assays were used without further purification. The purity and mass of all modified RNA oligonucleotides was confirmed by MALDI mass spectrometry. All reagents were stored at –20ºC until use.

Transcription templates for RNAs used in biochemical assays. Templates for primer-extending variants t307 and GAA-t307 were plasmids generated by QuickChange mutagenesis from pH307HP[35], which brackets the ligase core with 5′-hammerhead (HH) and 3′-hepatitus dela virus (HDV) self-cleaving ribozymes. Excision of these ribozymes ensures that the final ligase species has homogeneous termini[81]. For t307, the relevant sequence of the insert was

GCGTAATACGACTCACTATA*GGGAGAGTAGTATAGTGCTGATGAGTCCGTGAGGACGAAAC*
*GGTACCCGGTACCGT*C**C**ACTATACTACTGGATAATCAAAGACAAATCTGCCCGAAGGGCTTGAG
AACATACCCATTGCACTCCGGGTATGCAGAGGTGGCAGCCTCCGGTGGGTTAAAACCCAACGTT
CTCAACAATAGTG**A***GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATTCCG*
AGGGGACCGTCCCCTCGGTAATGGCGAATGGGACCCAC

where larger letters indicate the T7 promoter, bold nucleotides denote the 5′ and 3′ ends of the mature ligase species, italicized nucleotides denote the 5′-HH and 3′-HDV ribozymes and underlined nucleotides denote the U1A modification in stem P5. For GAA-t307, the relevant sequence of the insert was

GCGTAATACGACTCACTATA*GGGAGAGTATAGTGTTCCTGATGAGTCCGTGAGGACGAAAC*
*GGTACCCGGTACCGTC***G**AACACTATACTACTGGATAATCAAAGACAAATCTGCCCGAAGGGCTT
GAGAACATACCCATTGCACTCCGGGTATGCAGAGGTGGCAGCCTCCGGTGGGTTAAAACCCAAC
GTTCTCAACAATAGTG**A***GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATT*
CCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACCCAC

annotated as above. Plasmids were linearized by digestion with *EarI* endonuclease before use.

Templates for the C47A, C47G and C47Δ mutants, and the "3′-Arm" species used in assembling modified ligases by splinted ligation (*see below*) were PCR products generated from overlapping DNA oligonucleotides[19]. Templates for the active site mutants did not include self-cleaving ribozymes; that for the "3′-Arm" construct consisted of a 5′-HH ribozyme followed by nucleotides 52–121 of the ligase core, appended with the P5 U1A-loop extension. The full sequence of this construct was

GCGTAATACGACTCACTATA*GGGAGATGGGTATGTTCCTGATGAGTCCGTGAGGACGAAAC*
*GGTACCCGGTACCGTC***G**AACATACCCATTGCACTCCGGGTATGCAGAGGTGGCAGCCTCCGGTG
GGTTAAAACCCAACGTTCTCAACAATAGTG**A**

annotated as above.

The template for the "5′-Arm" species used in assembling modified *cis*-ligases by splinted ligation (*see below*) was a DNA oligonucleotide modified with two 2′-O-methyl nucleotides at its 5′-terminus. These modifications have been shown to limit the 3′-

heterogeneity of transcription products[82]. The template oligonucleotide was annealed to a 21-nt T7 promoter top strand DNA oligonucleotide and used as previously described[32,82].

Transcription and purification of RNAs used in biochemical assays. T7 *in vitro* transcription reactions were performed as described[19], in volumes of 50 μL–1 mL, depending on the construct. Templates were included at concentrations of 20 μg/mL in reactions using linearized plasmids, 6 μg/mL for those using PCR products, and 6 μg/mL for those using DNA oligonucleotides. Transcription reactions were quenched with the addition of EDTA, extracted with phenol and chloroform, ethanol precipitated and resuspended into deionized water. Constructs employing self-cleaving ribozymes were subsequently desalted and refolded to allow full processing of the HH or HDV ribozymes, as previously described[35,81]. The terminal 2′-3′ cyclic phosphate was not removed from these species. All RNAs were purified from denaturing 6% 0.5X TBE polyacrylamide gels, passively eluted into 300 mM NaCl at 4ºC overnight, ethanol precipitated, resuspended in deionized water and stored at –20ºC before use.

Assembly of modified ligases. Ligase constructs bearing modified nucleotides were assembled by DNA-splinted ligation[83,84]. To allow modular manipulation of functional groups at all positions near the active site, the ligase was schematically split into five oligonucleotides spanning different regions under interrogation. For the construction of self-ligases (**Fig. 1A**), an *in vitro* transcribed "5′-Arm" oligonucleotide (*see above*) spanning nucleotides 1–23 was used. For the construction of primer-extending ligases (**Fig. 1A**, *inset*) this oligonucleotide was replaced with a chemically synthesized species spanning positions 4–23. Modifications at C30 were introduced on synthetic 10mer oligonucleotides spanning residues 24-33; those at C47 were

introduced on 9mer oligonucleotides spanning residues 43–51. An unmodified intermediary bridging oligonucleotide spanned positions 34–42. The remainder of the ligase (corresponding to residues 52–121, appended with the P5 U1A-loop) was provided by a "3′-Arm" RNA (*see above*). The DNA splint was complementary to residues 13–61, hence extending ten residues upstream of junction between the 5′-Arm and 10mer oligonucleotide, and ten residues downstream of the junction between the 9mer oligonucleotide and the 3′-Arm. As required for splinted ligation, the 9mer, 10mer and bridging oligonucleotides were synthesized bearing 5′-monophosphates. The 3′-Arm was phosphorylated with T4 polynucleotide kinase (PNK, New England Biolabs) prior to gel purification, according to the manufacturer's instructions.

DNA-splinted assembly reactions were performed in a single pot, one-step synthesis using T4 RNA ligase 2^{84} (RNAl2, New England Biolabs). Equimolar concentrations of all five RNA oligonucleotides and the DNA splint were combined in deionized water to give a final concentration of 12 μM. The reaction mixture was annealed by heating and cooling (85ºC and 37ºC, for five minutes each), supplemented with the manufacturer's reaction buffer (1X, final concentration) and incubated at 37ºC for 15 minutes more. RNAl2 was then added to a final concentration of 1 U/μL. After four hours at 37ºC, the reaction was quenched by the addition of EDTA to 60 mM and NaCl to 300 mM, extracted with phenol and ethanol precipitated. Pellets were resuspended in deionized water, brought to 5 M urea and purified from denaturing 6% 0.5X TBE polyacrylamide gels. Bands corresponding to reaction products were visualized by UV-shadowing and excised. RNA was passively eluted into 300 mM NaCl overnight at 4ºC, ethanol precipitated, resuspended into deionized water and stored at –20ºC before use.

Under these reaction conditions, the assembly reaction appears nearly quantitative, as monitored by ethidium bromide-stained gels (**Fig. 7A**). However, due to the inefficiency of

passive gel elution, reactions starting with 2 nmol of oligonucleotide substrates typically yielded 0.5–0.8 nmol of fully assembled product. Unmodified ligase species assembled in this fashion exhibited activities identical to those that had been transcribed whole *in vitro* (**Fig. 7B**), within the experimental error.

Kinetic Assays. Unless otherwise noted, all biochemical assays were performed in triplicate; reported errors are standard deviations. All data were fit using the program KaleidaGraph (Synergy Software). In experiments calculating relative rates, uncertainties were propagated using the equation

$$\sigma_R = R\sqrt{\left(\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2\right)}$$

where A and B are individual rate constants, R is their ratio, and σ_A, σ_B and σ_R are their associated uncertainties.

Kinetic assays used substrate oligonucleotides that had been 5′-radiolabeled using PNK and [^{32}P]γ-ATP, or ribozymes that had been body labeled by addition of [^{32}P]α-UTP in the transcription reaction. Unless otherwise noted, all kinetic assays were performed in 50 mM buffer (sodium cacodylate for pH 6.0–6.5; Tris for pH 7.0–9.0), 10 mM MgCl2, 200 mM KCl and 600 μM EDTA. Compounds used in exogenous rescue experiments were prepared in 50 mM Tris•HCl buffer, pH 8.0, and their pH was confirmed at the reaction concentrations before use.

Self-ligation assays were performed as described previously[33]: the ribozyme was heated (5 minutes, 80ºC) and cooled (5 minutes, 22ºC) in water, and reactions were initiated at 22ºC by the simultaneous addition of buffer, salts and substrate oligonucleotide. The final ligase

concentration was 1 μM; substrate was added to a final concentration of 0.5 μM. In exogenous rescue experiments, the rescuing compound was added concomitantly with the substrate.

Primer extension assays were performed under single-turnover conditions in a manner similar to that described previously[32], with the order of component addition changed slightly. Ribozyme was first heated (5 minutes, 80ºC) and cooled (5 minutes, 22ºC) in water, and then supplemented with buffer, salts and the substrate oligonucleotide. This mixture was incubated at 22ºC for 15 minutes, and reactions were initiated by the addition of NTP. Since NTPs chelate Mg^{2+}, the NTP was added in an equimolar mixture with $MgCl_2$[32]. Final concentrations were 1 μM ligase, 0.5 μM substrate and, unless otherwise noted, 4 mM NTP.

For all reactions, aliquots were taken at the specified time points and rapidly quenched by mixing with an equal volume of gel-loading buffer (8 M urea, 120 mM EDTA, trace bromphenol blue and xylene cyanol). Samples from primer-extension assays, or from self-ligation assays with radiolabeled substrates were separated on denaturing 0.5X TBE 20% polyacylamide gels; those from body-labeled self-ligation assays were separated on denaturing 6% gels. Gels were visualized by phosphorimaging (Fujifilm BAS-2500). For each time point the fraction product was measured as

$$F_P(t) = \frac{P(t)}{(P(t) + R(t))}$$

where $P(t)$ and $R(t)$ are the product and reactant at a given time, respectively, and fit to the equation

$$F_P(t) = F_M(1 - e^{-k_{OBS}t})$$

were t equals time, treating F_M (the maximum fraction reacted) and k_{OBS} (the observed rate constant) as unknowns. Michaelis-Menten parameters were measured from single-turnover

primer-extension reactions in which the concentration of NTP was varied. Observed single-turnover rate constants at each NTP concentration ($k_{OBS}([NTP])$) were measured and fit to the equation

$$k_{OBS}([NTP]) = \frac{k_{CAT}[NTP]}{K_M + [NTP]}$$

with k_{CAT} and K_M treated as unknowns. Experiments were performed at pH 7.0.

Kinetic Isotope Experiments. For solvent isotope effect experiments, all reaction components (ribozymes, substrates, buffers, $MgCl_2$, KCl, EDTA) were reformulated in parallel in H_2O and D_2O. Aliquots of ribozyme and radiolabeled substrate in H_2O were dried overnight in a speed-vac, resuspended in D_2O, quantified by UV-vis spectroscopy and stored at –80ºC until use. Buffers were brought to pL (pH or pD) by titration with the conjugate acid and base of the buffer system. MES (free acid, sodium salt) was used for experiments at pL 6.0 and Tris (hydrochloride, free base) for experiments at pL 8.0. In all cases, the pL was measured using a glass electrode; for buffers in D_2O, the apparent pD was corrected by adding a value of 0.4 to the instrument readings[85].

Other than the experimental variation of H_2O and D_2O, proton inventory experiments were performed under standard self-ligation conditions: 50 mM Buffer (pL 6.0 or 8.0), 10 mM $MgCl_2$, 200 mM KCl, 600 μM EDTA, 1 μM ribozyme and 0.5 μM substrate. Ribozyme samples in D_2O and H_2O were combined in appropriate ratios to produce mixtures in 100%, 80%, 60%, 40%, 20% and 0% D_2O. Substrate oligonucleotides, buffers and salts were similarly combined to obtain a parallel series of reaction start mixes at the same fractions of D_2O. Self-ligation reactions were then performed as described above, heat/cooling the ligase in isolation before the simultaneous addition of all other reaction components. Time points were measured as described

above, and at each molar fraction of D_2O (termed "*n*") the apparent rate constant, k_n, was calculated. Proton inventories were calculated by plotting the ratio of k_n to the apparent rate constant in pure water, k_{H2O}, as a function of *n*, and fitting the data to the modified Gross-Butler equation[56] for either a two-proton-transfer model,

$$\frac{k_n}{k_{H2O}} = (1 - n + n \cdot \Phi_1)(1 - n + n \cdot \Phi_2)$$

or for a one-proton-transfer model,

$$\frac{k_n}{k_{H2O}} = (1 - n + n \cdot \Phi)$$

where each Φ is the inverse of the of the SKIE for an individual ionizable group. Uncertainties were propagated as described above.

References

1. Joyce, G.F. & Orgel, L.E. Prospects for Understanding the Origin of the RNA World. in *The RNA World, Second Edition* (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 49-77 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999).
2. Orgel, L.E. Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* **39**, 99-123 (2004).
3. White, H.B., 3rd. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* **7**, 101-4 (1976).
4. Kruger, K. et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147-57 (1982).
5. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-57 (1983).
6. Zaug, A.J. & Cech, T.R. The intervening sequence RNA of Tetrahymena is an enzyme. *Science* **231**, 470-5 (1986).
7. Steitz, T.A. & Moore, P.B. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci* **28**, 411-8 (2003).
8. Fedor, M.J. & Williamson, J.R. The catalytic diversity of RNAs. *Nat Rev Mol Cell Biol* **6**, 399-412 (2005).
9. Nissen, P., Hansen, J., Ban, N., Moore, P.B. & Steitz, T.A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-30 (2000).
10. Wilson, D.S. & Szostak, J.W. In vitro selection of functional nucleic acids. *Annu Rev Biochem* **68**, 611-47 (1999).
11. Chen, X., Li, N. & Ellington, A.D. Ribozyme catalysis of metabolism in the RNA world. *Chem Biodivers* **4**, 633-55 (2007).
12. Bartel, D.P. Re-creating an RNA Replicase. in *The RNA World, Second Edition* (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 143-162 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999).
13. Doudna, J.A. & Szostak, J.W. RNA-catalysed synthesis of complementary-strand RNA. *Nature* **339**, 519-22 (1989).
14. Green, R. & Szostak, J.W. Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science* **258**, 1910-5 (1992).

15. Kim, D.E. & Joyce, G.F. Cross-catalytic replication of an RNA ligase ribozyme. *Chem Biol* **11**, 1505-12 (2004).
16. Hayden, E.J. & Lehman, N. Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chem Biol* **13**, 909-18 (2006).
17. Lincoln, T.A. & Joyce, G.F. Self-sustained replication of an RNA enzyme. *Science* **323**, 1229-32 (2009).
18. Bartel, D.P. & Szostak, J.W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411-8 (1993).
19. Ekland, E.H. & Bartel, D.P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res* **23**, 3231-8 (1995).
20. Ekland, E.H., Szostak, J.W. & Bartel, D.P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364-70 (1995).
21. Ekland, E.H. & Bartel, D.P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373-6 (1996).
22. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319-25 (2001).
23. Lawrence, M.S. & Bartel, D.P. New ligase-derived RNA polymerase ribozymes. *RNA* **11**, 1173-80 (2005).
24. Zaher, H.S. & Unrau, P.J. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**, 1017-26 (2007).
25. Lawrence, M.S. & Bartel, D.P. Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry* **42**, 8748-55 (2003).
26. Bergman, N.H., Johnston, W.K. & Bartel, D.P. Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry* **39**, 3115-23 (2000).
27. Wright, M.C. & Joyce, G.F. Continuous in vitro evolution of catalytic function. *Science* **276**, 614-7 (1997).
28. Ordoukhanian, P. & Joyce, G.F. A molecular description of the evolution of resistance. *Chem Biol* **6**, 881-9 (1999).
29. Paegel, B.M. & Joyce, G.F. Darwinian evolution on a chip. *PLoS Biol* **6**, e85 (2008).

30. Voytek, S.B. & Joyce, G.F. Niche partitioning in the coevolution of 2 distinct RNA enzymes. *Proc Natl Acad Sci U S A* **106**, 7780-5 (2009).
31. Vaish, N.K. et al. Zeptomole detection of a viral nucleic acid using a target-activated ribozyme. *RNA* **9**, 1058-72 (2003).
32. Glasner, M.E., Yen, C.C., Ekland, E.H. & Bartel, D.P. Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. *Biochemistry* **39**, 15556-62 (2000).
33. Glasner, M.E., Bergman, N.H. & Bartel, D.P. Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry* **41**, 8103-12 (2002).
34. Bagby, S.C., Bergman, N.H., Shechner, D.M., Yen, C. & Bartel, D.P. A class I ligase ribozyme with reduced Mg2+ dependence: Selection, sequence analysis, and identification of functional tertiary interactions. *RNA* **15**, 2129-46 (2009).
35. Shechner, D.M. et al. Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* **326**, 1271-5 (2009).
36. Serganov, A. et al. Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat Struct Mol Biol* **12**, 218-24 (2005).
37. Robertson, M.P. & Scott, W.G. The structural basis of ribozyme-catalyzed RNA assembly. *Science* **315**, 1549-53 (2007).
38. Xiao, H., Murakami, H., Suga, H. & Ferre-D'Amare, A.R. Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* **454**, 358-61 (2008).
39. Pitt, J.N. & Ferre-D'Amare, A.R. Structure-guided engineering of the regioselectivity of RNA ligase ribozymes. *J Am Chem Soc* **131**, 3532-40 (2009).
40. Yin, Y.W. & Steitz, T.A. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393-404 (2004).
41. Harding, M.M. Geometry of metal-ligand interactions in proteins. *Acta Crystallogr D Biol Crystallogr* **57**, 401-11 (2001).
42. Harding, M.M. The geometry of metal-ligand interactions relevant to proteins. II. Angles at the metal atom, additional weak metal-donor interactions. *Acta Crystallogr D Biol Crystallogr* **56**, 857-67 (2000).

43. Strobel, S.A. & Shetty, K. Defining the chemical groups essential for Tetrahymena group I intron function by nucleotide analog interference mapping. *Proc Natl Acad Sci U S A* **94**, 2903-8 (1997).
44. Strobel, S.A. A chemogenetic approach to RNA function/structure analysis. *Curr Opin Struct Biol* **9**, 346-52 (1999).
45. Sträter, N., Lipscomb, W.N., Klabunde, T. & Krebs, B. Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. *Angew. Chem. Int. Ed. Engl.* **35**, 2024-2055 (1996).
46. Doublie, S. & Ellenberger, T. The mechanism of action of T7 DNA polymerase. *Curr Opin Struct Biol* **8**, 704-12 (1998).
47. Basu, S. & Strobel, S.A. Thiophilic metal ion rescue of phosphorothioate interference within the Tetrahymena ribozyme P4-P6 domain. *RNA* **5**, 1399-407 (1999).
48. Patel, P.H., Suzuki, M., Adman, E., Shinkai, A. & Loeb, L.A. Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. *J Mol Biol* **308**, 823-37 (2001).
49. Hsiao, C. et al. Complexes of Nucleic Acids with Group I and II Cations. in *Nucleic Acid-Metal Ion Interactions*, Vol. 1 (ed. Hud, N.V.) 1-38 (RSC Publishing, Cambridge, UK, 2009).
50. Castro, C. et al. Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol* **16**, 212-8 (2009).
51. Peracchi, A., Beigelman, L., Usman, N. & Herschlag, D. Rescue of abasic hammerhead ribozymes by exogenous addition of specific bases. *Proc Natl Acad Sci U S A* **93**, 11522-7 (1996).
52. Perrotta, A.T., Shih, I. & Been, M.D. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science* **286**, 123-6 (1999).
53. Lebruska, L.L., Kuzmine, II & Fedor, M.J. Rescue of an abasic hairpin ribozyme by cationic nucleobases: evidence for a novel mechanism of RNA catalysis. *Chem Biol* **9**, 465-73 (2002).
54. Perrotta, A.T., Wadkins, T.S. & Been, M.D. Chemical rescue, multiple ionizable groups, and general acid-base catalysis in the HDV genomic ribozyme. *RNA* **12**, 1282-91 (2006).

55. Nakano, S., Chadalavada, D.M. & Bevilacqua, P.C. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**, 1493-7 (2000).
56. Venkatasubban, K.S. & Schowen, R.L. The proton inventory technique. *CRC Crit Rev Biochem* **17**, 1-44 (1984).
57. Tinsley, R.A., Harris, D.A. & Walter, N.G. Significant kinetic solvent isotope effects in folding of the catalytic RNA from the hepatitis delta virus. *J Am Chem Soc* **125**, 13972-3 (2003).
58. Castro, C. et al. Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc Natl Acad Sci U S A* **104**, 4267-72 (2007).
59. Saenger, W. in *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
60. Steitz, T.A. & Steitz, J.A. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* **90**, 6498-502 (1993).
61. Yang, W. An equivalent metal ion in one- and two-metal-ion catalysis. *Nat Struct Mol Biol* **15**, 1228-31 (2008).
62. Fedor, M.J. Comparative enzymology and structural biology of RNA self-cleavage. *Annu Rev Biophys* **38**, 271-99 (2009).
63. Ke, A., Zhou, K., Ding, F., Cate, J.H. & Doudna, J.A. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **429**, 201-5 (2004).
64. Das, S.R. & Piccirilli, J.A. General acid catalysis by the hepatitis delta virus ribozyme. *Nat Chem Biol* **1**, 45-52 (2005).
65. Luptak, A., Ferre-D'Amare, A.R., Zhou, K., Zilm, K.W. & Doudna, J.A. Direct pK(a) measurement of the active-site cytosine in a genomic hepatitis delta virus ribozyme. *J Am Chem Soc* **123**, 8447-52 (2001).
66. Gong, B. et al. Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography. *J Am Chem Soc* **129**, 13335-42 (2007).
67. Cochrane, J.C., Lipchock, S.V. & Strobel, S.A. Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. *Chem Biol* **14**, 97-105 (2007).
68. Rupert, P.B., Massey, A.P., Sigurdsson, S.T. & Ferre-D'Amare, A.R. Transition state stabilization by a catalytic RNA. *Science* **298**, 1421-4 (2002).

69. Salter, J., Krucinska, J., Alam, S., Grum-Tokars, V. & Wedekind, J.E. Water in the active site of an all-RNA hairpin ribozyme and effects of Gua8 base variants on the geometry of phosphoryl transfer. *Biochemistry* **45**, 686-700 (2006).
70. Nesbitt, S., Hegg, L.A. & Fedor, M.J. An unusual pH-independent and metal-ion-independent mechanism for hairpin ribozyme catalysis. *Chem Biol* **4**, 619-30 (1997).
71. Oubridge, C., Ito, N., Teo, C.H., Fearnley, I. & Nagai, K. Crystallisation of RNA-protein complexes. II. The application of protein engineering for crystallisation of the U1A protein-RNA complex. *J Mol Biol* **249**, 409-23 (1995).
72. Ferré-D'Amaré, A.R. & Doudna, J.A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J Mol Biol* **295**, 541-56 (2000).
73. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collection in Oscillation Mode. *Methods Enzymol.* **276**, 307-326 (1997).
74. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
75. Adams, P.D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-54 (2002).
76. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
77. Stahley, M.R., Adams, P.L., Wang, J. & Strobel, S.A. Structural metals in the group I intron: a ribozyme with a multiple metal ion core. *J Mol Biol* **372**, 89-102 (2007).
78. Klein, D.J., Moore, P.B. & Steitz, T.A. The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA* **10**, 1366-79 (2004).
79. The CCP4 suite: programs for protein crystallography. *Acta Crysallogr D Biol Crystallogr* **50**, 760-3 (1994).
80. Delano, W.L. The PyMOL Molecular Graphics System. http://pymol.sourceforge.net, DeLano Scientific, Palo Alto, CA, USA (2002).
81. Price, S.R., Ito, N., Oubridge, C., Avis, J.M. & Nagai, K. Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J Mol Biol* **249**, 398-408 (1995).

82. Kao, C., Zheng, M. & Rudisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268-72 (1999).
83. Moore, M.J. & Sharp, P.A. Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites. *Science* **256**, 992-7 (1992).
84. Bullard, D.R. & Bowater, R.P. Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Biochem J* **398**, 135-44 (2006).
85. Glasoe, P.K. & Long, F.A. Use of glass electrodes to measure acidities in deuterium oxide. *J Phys Chem* **64**, 188-190 (1960).
86. Leontis, N.B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499-512 (2001).
87. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crysallogr* **A32**, 922-923 (1976).
88. Rohatgi, R., Bartel, D.P. & Szostak, J.W. Kinetic and mechanistic analysis of nonenzymatic, template-directed oligoribonucleotide ligation. *J Am Chem Soc* **118**, 3332-9 (1996).

Figure Legends

Figure 1. Architecture of the Class I ligase ribozyme preligation complex. (**A**) Secondary structure of the crystallization construct, drawn reflecting the coaxial stacking and relative domain orientation observed in crystal structures[35]. The ribozyme is depicted undergoing ligation: the nucleophile, electrophile and leaving groups are drawn in red. Curved arrows indicate attack by the substrate 3′-hydroxyl on the ribozyme α-phosphate, with concomitant loss of pyrophosphate. Active site backbone phosphates at positions 29 and 30 are shown as yellow boxes, as is residue 47, which has been mutated from cytosine to uridine to trap the ligase prior to catalysis. Nucleotides in gray were added to facilitate crystallization. Residue numbering is as in (ref. 19); base pair geometries are indicated using the nomenclature of Leontis and Westhof[86]. Base triples are boxed in gray and connected with gray lines; stacked residues are vertically aligned or connected with thin gray lines terminating in gray bars. Inset, primer-extending (PE) ligase constructs t307 (*left*) and GAA-t307 (*right*), depicted extending their primers by exogenous GTP. Residue 47 in these constructs has been restored to cytosine. (**B**) Overview of the Ca^{2+}/Sr^{2+}-preligation structure, peering into the active site. The U1A protein and its cognate loop have been removed from view. The 5′-GTP is shown as sticks. Proposed catalytic metal ions, or solvent atoms that appear important for GTP docking, are shown as spheres (orange, metal ions; chartreuse, water). Meshes are simulated-annealing $|F_o| - |F_c|$ OMIT maps in which G1 (magenta, contoured at 5σ) or active site solvent atoms (dark blue, contoured at 4.5σ) were excluded from map calculations. (**C**) Overview of the Mg^{2+}/Sr^{2+}-preligation structure, depicted as in **B**.

Figure 2. Comparison of the preligation and product structures. (**A**) Superposition of the Ca^{2+}/Sr^{2+}-preligation complex and the product structure (PDB ID: 3HHN)[35]. Left, global view of the alignment, depicted as a backbone trace. The U1A protein and its cognate loop have been removed from view. Residues in lavender were poorly ordered. Right, active site view of the same alignment. The product structure is depicted as black lines. The preligation structure is depicted as sticks, colored as in **Fig. 1**; the 2′-hydroxyl nucleophile and G1 phosphorus atoms are in red, G1 phosphate oxygens are in purple. A structural Ca^{2+} ion and water molecules are shown as orange and chartreuse spheres, respectively; other active site solvent atoms have been removed from view. (**B**) Superposition of the Mg^{2+}/Sr^{2+}-preligation complex and the product structure, depicted as in **A**. Alignments were performed using the program LSQKAB[87] in the CCP4 program suite[79], anchored on all RNA atoms outside of the U1A loop. (**C**) Crystallographic disorder in the Ca^{2+}/Sr^{2+} structure. The 5′-end of P1-P2 domain, comprising helix P1 and the first nine residues of J1/3, is shown for each preligation structure. Meshes are simulated-annealing $|F_o| - |F_c|$ OMIT maps, contoured at 3.5σ, in which the residues depicted as sticks were excluded from the map calculations. These residues were colored in lavender in part **A**.

Figure 3. Active site interactions with the 5′-GTP and solvent. (**A**) Stereograph of the active site observed in the Ca^{2+}/Sr^{2+}-preligation structure. Coloring and rendering is as in **Fig. 2**. Inner-sphere coordination interactions with metal ions are depicted as thin sticks; hydrogen bonds are shown as black dashes. Meshes are simulated-annealing $|F_o| - |F_c|$ OMIT maps in which the GTP (magenta, contoured at 5σ) or active site solvent atoms (navy, contoured at 4.5σ) were excluded from map calculations. (**B**) Stereograph of the active site observed in the Mg^{2+}/Sr^{2+}-preligation

structure, depicted as in **A**. (**C**) Schematic summary of solvent interactions with the 5′-GTP observed in the Mg^{2+}/Sr^{2+}-preligation complex. Blue numbers indicate the distance, in angstroms, between atomic centers. Waters hydrating the Mg^{2+} bound by G1 and G2 are represented by black lines. The inner sphere Mg^{2+}–H_2O bond near the β-phosphate was restricted to 2.07Å during refinement (**Methods**). (**D**) Reorientation of the 5′-GTP during metal ion exchange. Solvent atoms have been removed from view. The two preligation structures were superposed by alignment with G2. Indicated distances are between the 3′-hydroxyl nucleophile and α-phosphate of each structure; the indicated angular shift is between the γ-phosphates in each structure and the α-phosphate in the Mg^{2+}/Sr^{2+}-complex. (**E**) Michaelis-Menten kinetics for single-turnover, single-nucleotide primer extension by the t307 (black) and GAA-t307 (gray) constructs.

Figure 4. The C47 nucleobase participates directly in catalysis. (**A**) Self-ligation kinetics for the C47U (*left*) and C47Abasic (*right*) variants, supplemented with exogenous small molecules. The y-axis is the same for both plots. Rescuing compounds were added at 35 mM – near the solubility limit for cytosine. Reaction kinetics at the solubility limits of the other rescuing compounds, or at 400 mM Imidazole, were indistinguishable from those shown here (*data not shown*). (**B**) Self-ligation rates of C47 variants, relative to the wild type. Results are shown on a logarithmic scale. The ^{4S}U (4-thiouridine) variant shows biphasic kinetics; each of the two rates is shown individually. The uridine, abasic, guanosine and delta mutants were measured at pH 8.0 and scaled relative to the wild type rate at that pH; all other variants were measured at pH 6.0. The uncatalyzed rate was measured previously[88]. (**C**) pH dependence of the self-ligation rate of the pyridin-2-one variant. The fit between pH 6.0 and 8.0 is linear. (**D**) Proton

inventories for self-ligation by the wild type (*left*), C47P2o (*middle*) and C47U variants. The y-axis is the same for each plot. n is the molar fraction of D_2O; k_n is the apparent first-order rate contant at a given n, k_o is this rate constant in 100% H_2O. The solid black line indicates the best fit for a single-proton-transfer model; the dashed gray line indicates the best fit for a two-proton-transfer model[56].

Figure 5. The C30 2′-hydroxyl participates directly in catalysis. (**A**) Orthogonal views of the active site observed in the Mg^{2+}/Sr^{2+}-complex, highlighting interactions between C47, C30, a partially hydrated Mg^{2+} ion and the 5′-triphosphate. Coloring and rendering are as in **Fig. 3A**. Red dotted line indicates the assumed line of nucleophilic attack. (**B**) Proton inventory of the 2′-deoxy C30 construct, plotted as in **Fig. 4D.** (**C**) Relative first-order rate constants for single-addition primer-extension by t307 constructs bearing modifications at the C30 2′-hydroxyl. Results are scaled relative to the unmodified construct (C30 2′-OH), and shown on a logarithmic scale. (**D**) Relative self-ligation rate constants for C30 2′-modifications in the context of base variants at position 47. Results are scaled relative to the C30 2′-OH construct for each C47 base, and are shown on a logarithmic scale. (**E**) Two schemes summarizing functional interaction between the C30 2′-hydroxyl and the C47 N4 during self-ligation. In each, numbers above the arrows indicate the fold reduction in first-order rate constants incurred by introducing the given modification. The left scheme summarizes the effects of removing activating groups, the C47 N4 amine and C30 2′-hydroxyl. The right scheme summarizes the effects of adding inhibitory groups, the U47 O4 and C30 2′-fluoro. At pH 6.0, the observed rate of the unmodified ligase (upper left corner of each scheme) was 0.93±0.009 min^{-1}.

Figure 6. A model for transition-state stabilization by the Class I ligase ribozyme. Black dotted lines indicate bonds formed or broken during the transition state; red arrows indicate the reaction direction. Gray dashed lines denote hydrogen bonds. The A29 pro-*Sp* and C30 pro-*Rp* non-bridging phosphate oxygens form inner-sphere contacts with the catalytic metal ion (thick blue lines). The remaining Mg^{2+} ligands (thick gray lines) are presumed to be water (*not shown*).

Figure 7. Assembly of active ligases by DNA-splinted ligation. (**A**) Efficient DNA-splinted assembly of modified t307 species. Leftmost lane: the 3′-Arm oligonucleotide, the longest of the RNAs used in assembly. Middle three lanes: end points (four hr, **Methods**) of DNA-splinted assembly reactions; products are t307 derivatives bearing phosphorothioate modifications at the indicated positions. Rightmost lane: *in vitro*-transcribed, unmodified t307. This material runs slightly faster than the assembled t307 products because it terminates with a 2′-3′ cyclic phosphate and lacks 3′- transcription heterogeneities. Shown is a denaturing polyacrylamide gel stained with ethidium bromide. (**B**) DNA-splinted assembly of active self-ligases. Shown are typical self-ligation timecourses using a 5′-^{32}P-labeled oligonucleotide substrate. Left lanes: T7 *in vitro*-transcribed wild type 307. Right lanes: the equivalent species generated by DNA-splinted assembly. The observed rates of these samples differ by approximately two-fold (*not shown*).

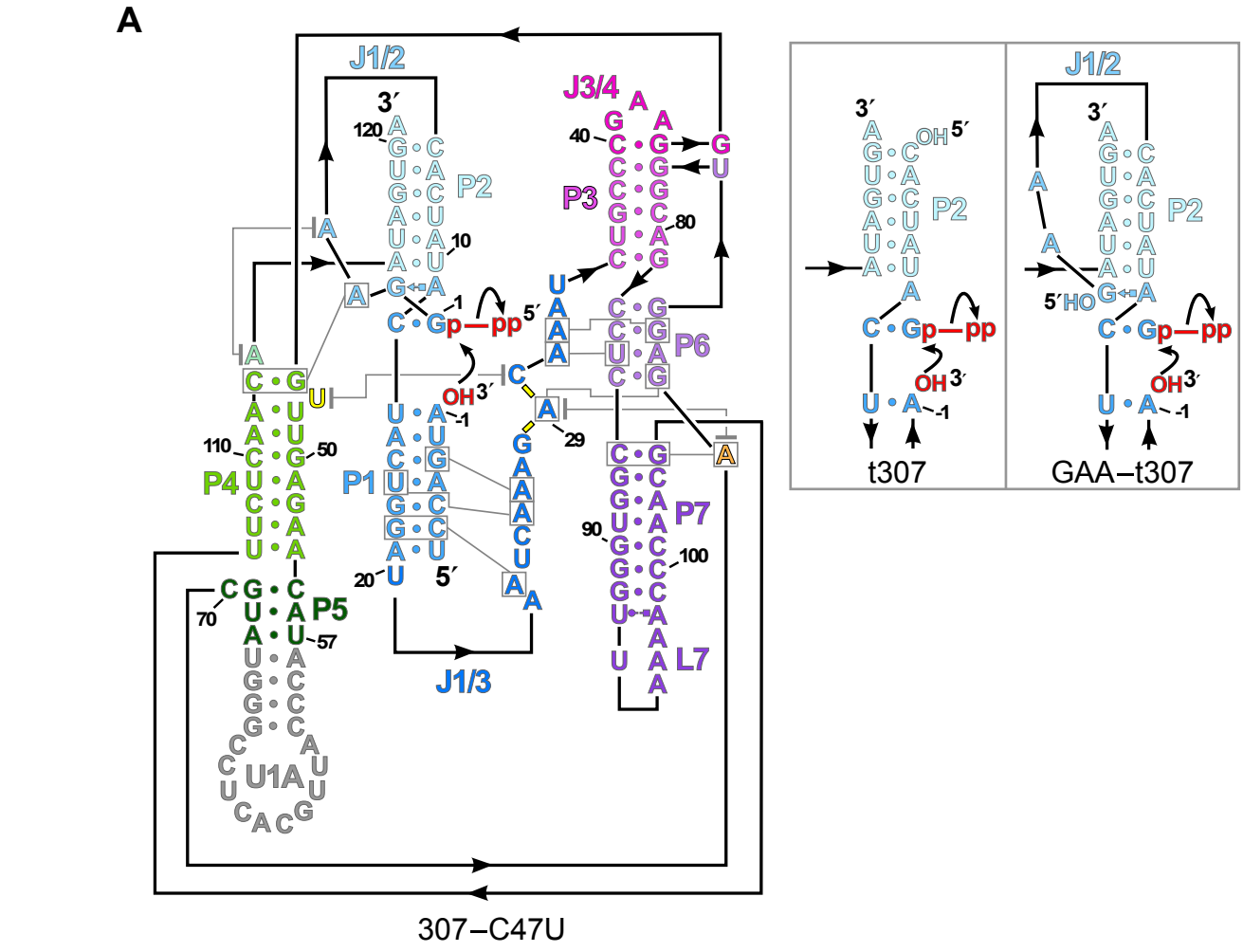
Table 1

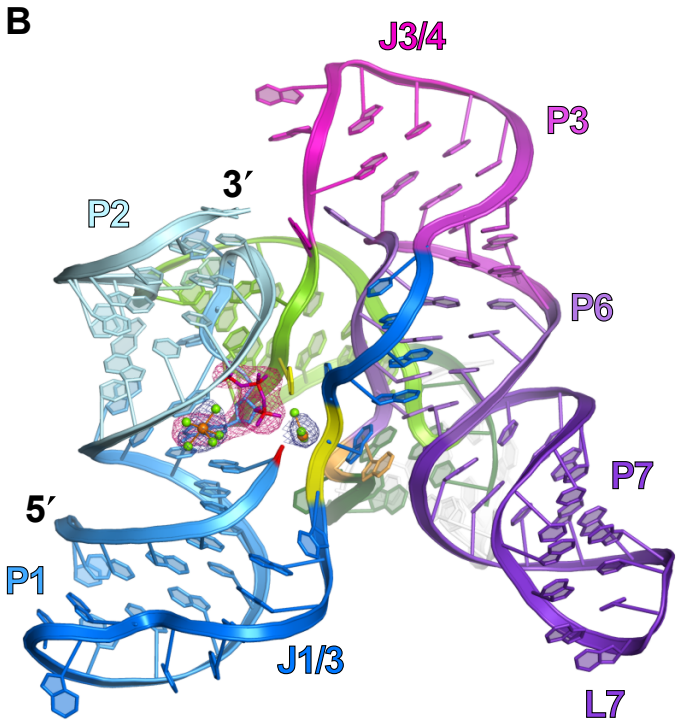
Crystallographic statistics for preligation class I ligase•substrate complexes

Crystal	Ca^{2+}/Sr^{2+}	Mg^{2+}/Sr^{2+}	Mg^{2+} Alone[c]
Data Collection			
Space group	P1	P1	P1
Cell dimensions			
a,b,c (Å),	58.70, 69.99, 71.86	59.19, 70.24, 71.21	59.52, 70.62, 71.21
α,β,γ (°)	99.86, 99.73, 103.68	99.86, 99.34, 103.81	100.65, 99.34, 104.09
Wavelength (Å)	0.9795	0.9795	0.9795
Resolution (Å)[a]	30-3.15 (3.26-3.15)	30-3.15 (3.26-3.15)	40-3.10 (3.21-3.10)
R_{merge} (%)[a]	9.8 (39.8)	13.5(42.9)	9.4 (36.8)
Mean I/σ(I)[a]	13.5 (2.1)	17.5 (3.2)	14.2 (2.4)
Unique Observations[a]	18,190 (1,767)	18,739 (1,858)	19,396 (1810)
Completeness (%)[a]	98.4 (94.1)	99.1 (98.3)	98.2 (91.5)
Redundancy[a]	3.3 (2.9)	7.2 (5.7)	3.7 (3.0)
Refinement			
R_{Work} (%)[b]	21.23	19.74	
R_{Free} (%)[b]	25.45	23.98	
Number of atoms			
RNA	5862	5862	
Protein	1369	1392	
Ca^{2+}/Mg^{2+}	26	46	
Sr^{2+}	0	0	
Water	38	51	
Mean B-factors(Å^2)			
RNA	93.7	78.9	
Protein	73.6	65.9	
M^{2+}	103.2	75.5	
Water	88.0	76.0	
R.m.s. deviations			
Bond distances (Å)	0.015	0.006	
Angles (°)	1.133	1.310	
Estimated coordinate error (Å)[d]	0.47	0.40	

[a]Values in parenthesis refer to the highest resolution shell. [b]5% of all data were excluded from refinement for R_{Free} calculations. [c]Refinement of the Mg^{2+} Alone structure was stopped after the rigid body step. [d]Maximum likelihood, as calculated in PHENIX.

Figure 1


A
J1/2
J3/4
P2
P3
P1
P4
P5
P6
P7
L7
J1/3
307–C47U
t307
GAA–t307


B
J3/4
P3
P2
3′
P6
P7
5′
P1
J1/3
L7

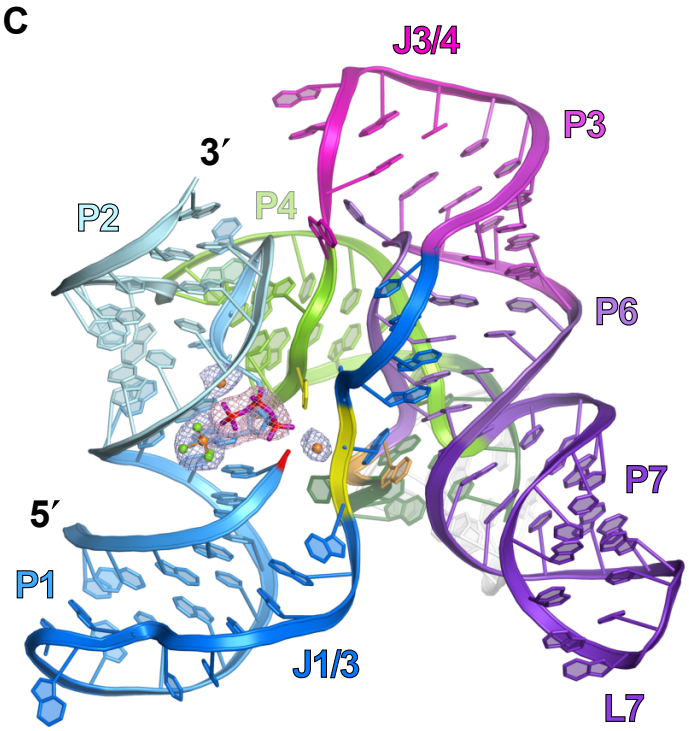

C
J3/4
P3
3′
P2
P4
P6
P7
5′
P1
J1/3
L7

Figure 2

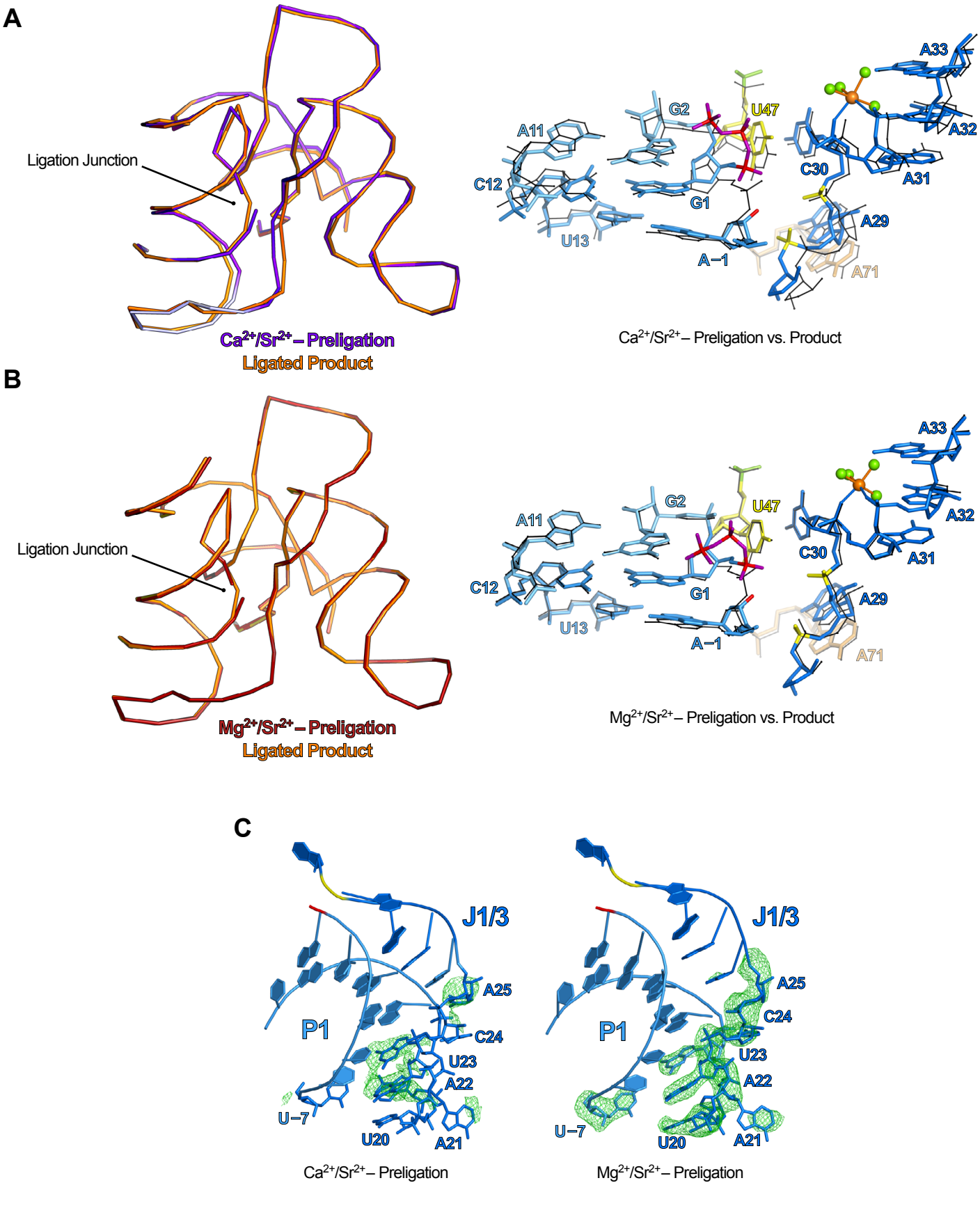

A
Ligation Junction
Ca2+/Sr2+ – Preligation
Ligated Product
A33
G2
U47
A11
A32
C12
C30
A31
G1
A29
U13
A–1
A71
Ca2+/Sr2+ – Preligation vs. Product
B
Ligation Junction
Mg2+/Sr2+ – Preligation
Ligated Product
A33
G2
U47
A11
A32
C12
C30
A31
G1
A29
U13
A–1
A71
Mg2+/Sr2+ – Preligation vs. Product
C
J1/3
A25
P1
C24
U23
A22
U–7
U20
A21
Ca2+/Sr2+ – Preligation
J1/3
A25
C24
P1
U23
A22
U–7
U20
A21
Mg2+/Sr2+ – Preligation

Figure 3

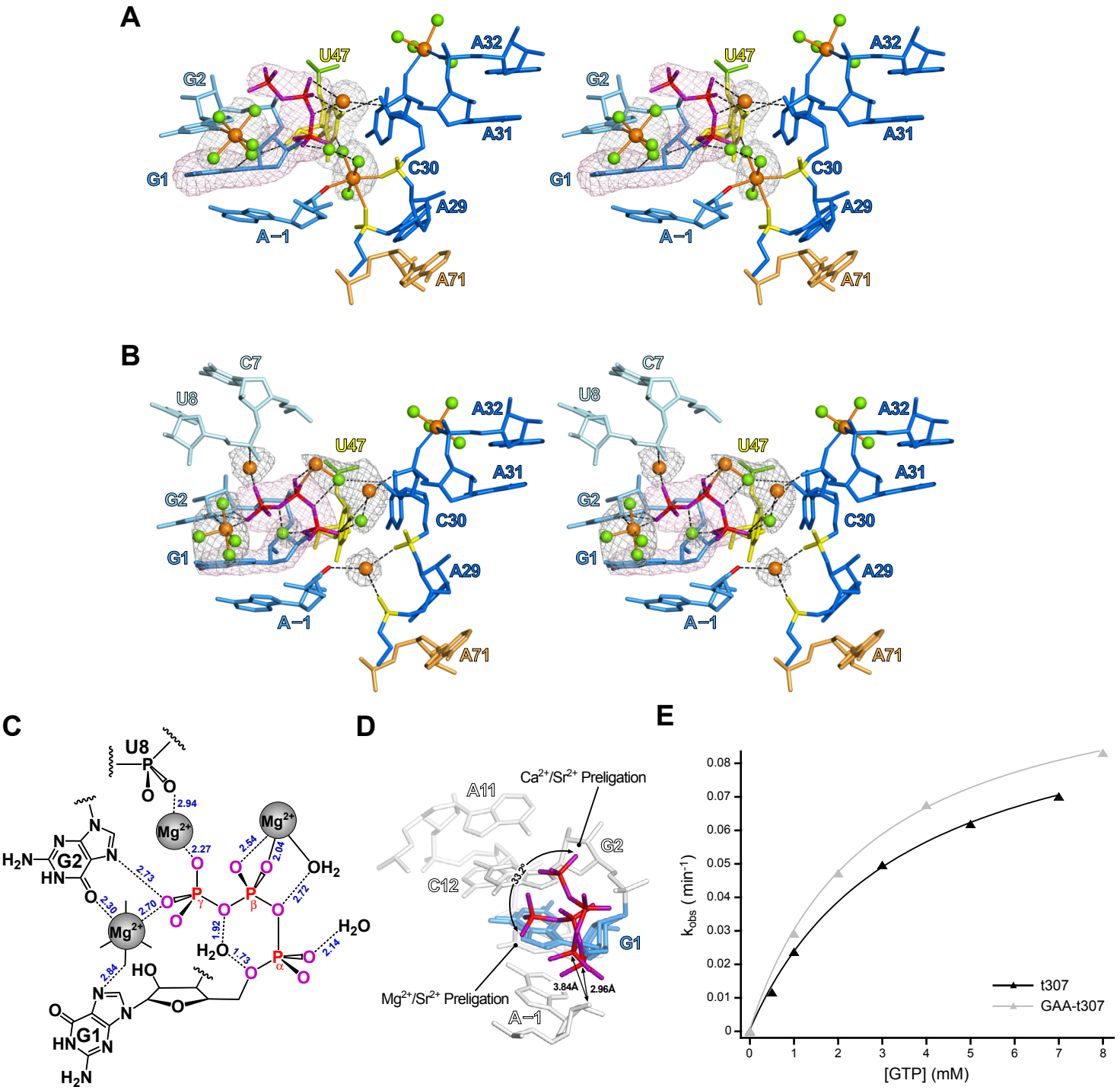

A
U47
A32
G2
A31
G1
C30
A29
A−1
A71
B
C7
U8
C
D
Ca2+/Sr2+ Preligation
A11
C12
Mg2+/Sr2+ Preligation
3.84Å
2.96Å
E
kobs (min-1)
[GTP] (mM)
t307
GAA-t307

Figure 4

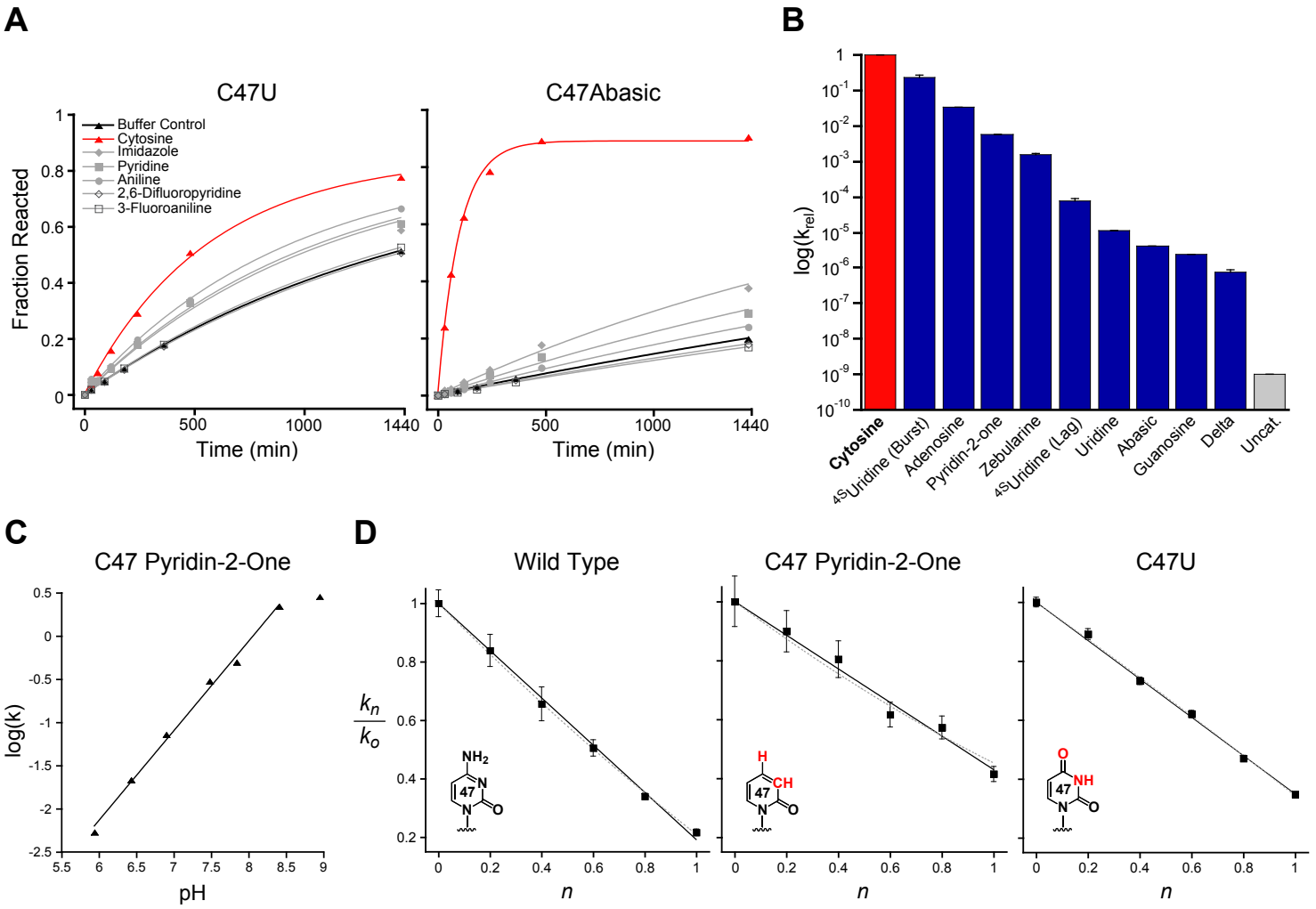

A
C47U
C47Abasic
Buffer Control
Cytosine
Imidazole
Pyridine
Aniline
2,6-Difluoropyridine
3-Fluoroaniline
Fraction Reacted
Time (min)
B
log(k_rel)
Cytosine
Uridine (Burst)
Adenosine
Pyridin-2-one
Zebularine
Uridine (Lag)
Uridine
Abasic
Guanosine
Delta
Uncat.
C
C47 Pyridin-2-One
log(k)
pH
D
Wild Type
C47 Pyridin-2-One
C47U
k_n / k_o
n

Figure 5

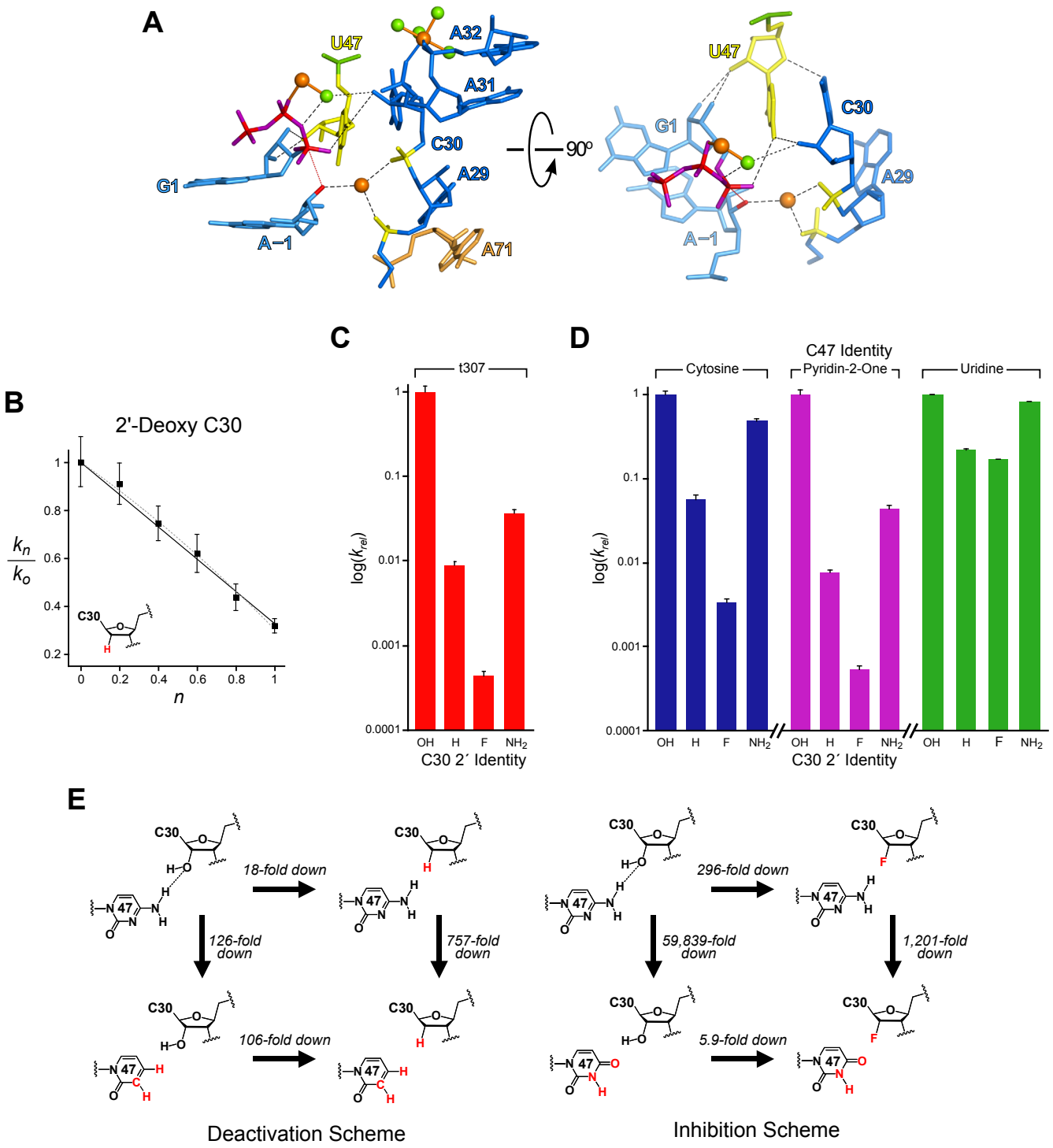

A
U47
A32
A31
C30
A29
G1
A−1
A71
90°
U47
C30
G1
A29
A−1
B
2'-Deoxy C30
k_n/k_o
n
C30
C
t307
log(k_{rel})
OH H F NH2
C30 2´ Identity
D
C47 Identity
Cytosine
Pyridin-2-One
Uridine
log(k_{rel})
OH H F NH2
C30 2´ Identity
E
18-fold down
126-fold down
757-fold down
106-fold down
296-fold down
59,839-fold down
1,201-fold down
5.9-fold down
Deactivation Scheme
Inhibition Scheme

Figure 6

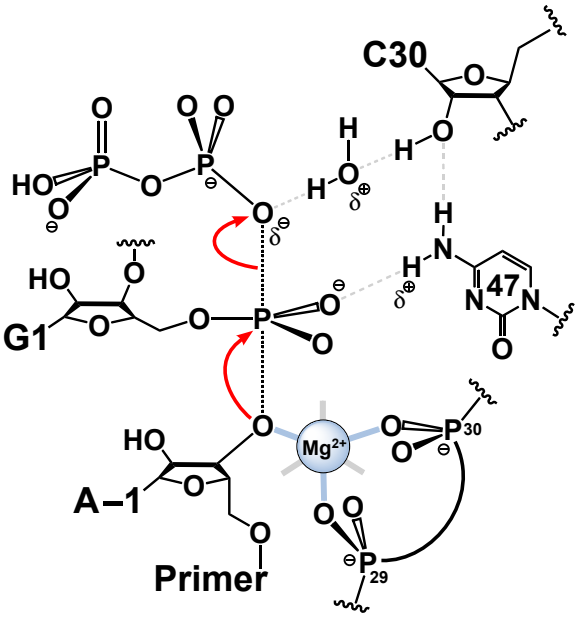
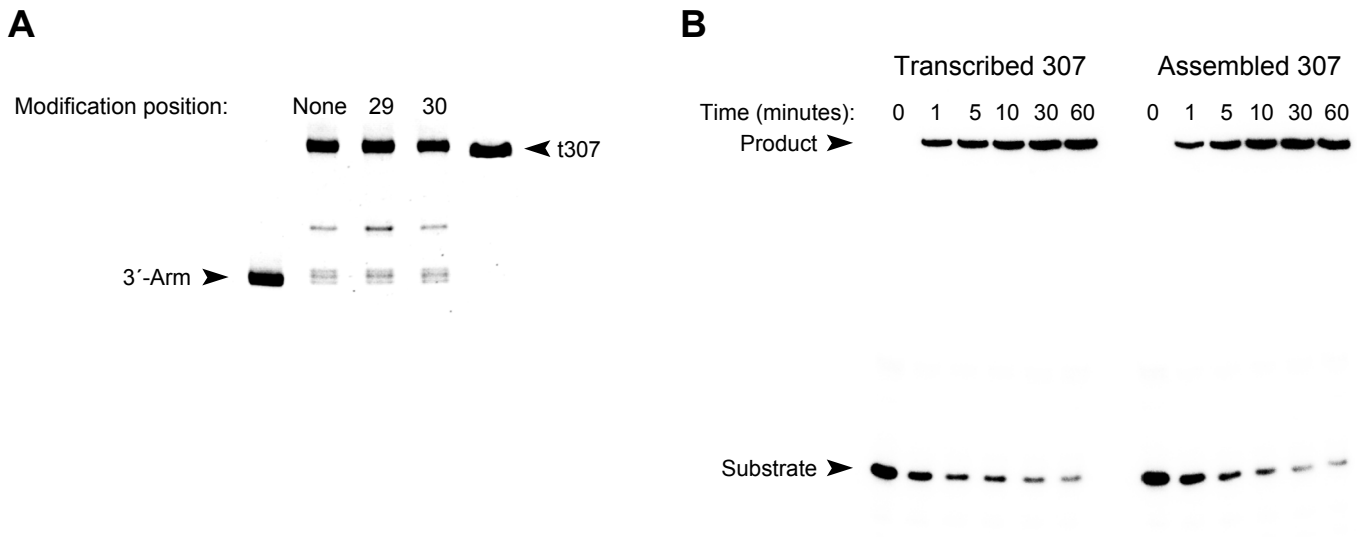


Figure 7


A
Modification position:
None
29
30
t307
3´-Arm
B
Transcribed 307
Assembled 307
Time (minutes):
0 1 5 10 30 60
0 1 5 10 30 60
Product
Substrate

Future Directions

Even prior to the work presented here, the Class I ligase was among the most extensively studied and best characterized artificial ribozymes. Building on this robust enzymological background, crystallographic analyses have now elucidated in atomic detail the global and active site architectures of this remarkable enzyme. These results provided an invaluable starting point for studies that interrogated the functional groups on the ligase that directly participate in catalysis, and have yielded a catalytic model with a level of detail nearly rivaling those of the natural ribozymes. This model is still lacking in some regards, and could be strengthened through further crystallographic and biochemical analyses, as described below. Moreover, a more expansive line of inquiry would harness this mature understanding of the ligase core, using it as the groundwork for dissecting the ligase-derived polymerase ribozymes. The stage is now set for detailed structural and mechanistic analysis of these polymerases, and for more informed efforts at isolating a true replicase ribozyme.

Biochemical and crystallographic approaches to the ligase mechanism

Our current model of the ligase catalytic mechanism is lacking in two regards: an explicit understanding of the catalytic roles played by metal ion cofactors, and the mechanism by which incoming NTP substrates are recognized. Each of these shortcomings could potentially be addressed through a combination of crystallographic and biophysical techniques.

Overall, the crystallographic results presented here support those of previous biochemical experiments (1-3), suggesting that the ligase employs a catalytic Mg^{2+} cofactor. However, these analyses have directly demonstrated neither the presence of this cofactor, nor its role during catalysis. Metal-rescue experiments (4, 5), may be used to address this latter issue. It is known,

for example, that phosphorothioate (oxygen-to-sulfur) substitutions at A29, C30 and the incoming NTP dramatically interfere with ligase catalysis; this might be attributed to the ablation of a catalytic Mg^{2+}-binding site, because sulfur, unlike oxygen, is a poor inner-sphere ligand for "hard" metals like Mg^{2+}. Supplementing these modified ligase constructs with more thiophilic metals, such as Mn^{2+} or Cd^{2+}, may restore the phosphate-metal interaction, and hence rescue activity. The beauty of this technique is that it directly assays for catalytic activity, and not simply the presence or absence of a metal ion. However, its application to the ligase may prove problematic, since this particular ribozyme is strongly inhibited by thiophilic divalent metals (2). Indeed, preliminary metal-rescue attempts have proven unsuccessful (D.M.S., Sarah C. Bagby and D.P.B., unpublished results), though a more extensive search of rescue conditions will be necessary in order to prove this line of inquiry untenable.

While crystallographic assays cannot directly assay the functionality of metal ligands, facile techniques are available that provide a direct visualization of their binding sites. These could be used to probe the A29-C30 site, and furthermore identify other structural or catalytic Mg^{2+} ions that have remained opaque to our analyses due to the resolution of our current structures, the potential for intrinsic disorder therein, and the limitations of biochemical probing by NAIM (3, 6, 7). Since the ligase P1 crystal form is clearly tolerant to buffer exchanges, unreacted ligase crystals could be soaked with lanthanide ions, such as Yb^{3+} or Sm^{3+}. These trivalent metals are close Mg^{2+} isosteres, and have been shown to replace inner-sphere magnesium binding sites in RNA (8, 9). However, unlike magnesium, lanthanide elements exhibit conspicuous anomalous signals at experimentally tractable x-ray wavelengths, a property that has made them powerful tools for the direct visualization of metals in cases where resolution is limiting (10-12). This similarity between lanthanide (III) ions and Mg^{2+} is also exploited in

the terbium footprinting technique (13), wherein substitution of Tb^{3+} for Mg^{2+} induces slow, local cleavage of the RNA backbone near its binding site. A combination of solution-state and crystallographic probes might produce an exhaustive, if not functionally elucidative, map of specific metal binding sites in the ligase core.

The second unresolved mechanistic issue involves the recognition of NTP substrates, since in its *trans* primer-extension context the ligase discriminates cognate from noncognate substrates with fidelities superior to that predicted from the energetics of base pairing alone (14). Proteinaceous polymerases are known to achieve such sequence-independent specificity through a combination of steric complementarity and direct minor-groove contacts with the template•NTP base pair (15, 16). However, no such interactions are observed in the ligase: the sole interactions that bracket the substrate base pair, and which might serve to measure its geometry, involve functional groups on nucleotides 30, 47 and 48 (*see* chapter 1, figure 13B, *for example*). Since individual ablation of any of these groups resulted in an overall loss of activity, but not of fidelity (D.M.S. and D.P.B., unpublished results), cognate discrimination might require the combined efforts of multiple interactions, or contacts peripheral to the active site (17).

A crystallographic approach to this problem would involve solving structures for a systematic set of all possible template•NTP base pairs, which one might attempt either within the current *cis*-ligation context, or by moving to the *trans* primer-extending ligase (1). Each approach has its strengths and weaknesses. Since the work presented here demonstrates that the *cis*-ligase P1 crystal form can tolerate structural perturbations in the active site, this fidelity study could be undertaken without the need to seek out new crystallization hits. However, this approach is limited in its scope: although the template nucleotide could be varied systematically,

T7 transcription requires the 5′-triphosphate-bearing nucleotide to be either a guanosine or an adenosine (18, 19). Hence, half of all possible sequence contexts aren't tractable in this format. Crystallography of the *trans* primer-extending ligase is not limited in this regard, but would suffer from two principal drawbacks. First, crystals of the known P1 form could not be obtained for the *trans*- version of the 307-ligase (20), bound to its oligonucleotide substrate or product (D.M.S. and D.P.B., unpublished results). This might be justified in light of a crystal lattice contact between helix P2 and the J3/4 GAA triloop of a symmetry-related molecule; removing J1/2 might render the P2 helix more flexible with regard to the ligase core, and preclude the formation of this contact. Since the identical crystal contact is also observed in the Fab-ligase crystal form (Y. Koldobskaya *et al*., in preparation), moving to that crystal system seems an unlikely solution, and hence a search for new crystal forms would be required. Second, assuming that novel high-resolution crystals of a *trans*- construct could be found, the experiment's success would require that exogenous NTPs soaked into the crystal be specifically bound and well-ordered. Though one can never predict *a priori* if such will be the case, this approach has proven quite fruitful in analogous studies of yeast RNA polymerase II (21).

Another outstanding issue pertains to the catalytic roles played by the C47 N4 and the C30 2′-OH. While the biochemical data presented here provide strong evidence that they participate directly in catalysis, more corroborative evidence may come by explicitly probing their chemical properties biophysically. Though neither group is likely to act as a proton shuttle during catalysis, perturbation of either group's pK_a may enhance its ability to function electrostatically. NMR has been used to directly measure the pK_a of the HDV ribozyme C75 N3 (22), a group thought to function as a general acid catalyst. However, the technique proved inconsistent, yielding interpretable data only for RNA in its post-cleavage state, and may not be

applicable to the ligase. Moreover, for the ligase, the most powerful application of NMR would require site-specific labeling with ^{15}N at C47, and ^{13}C or ^{18}O at C30, a difficult and time-consuming process. An alternative technique that has recently proven quite powerful is Raman crystallography. This technique requires no specific labeling of the RNA, is blind to background signal from water and buffer components, and can be performed *in situ* within the crystallization experiment (23). Its shortcomings stem from the relatively weak signal it measures, and its requirement that the functional groups of interest exhibit spectroscopic signals different from those on all other residues. Still, the technique has been fruitfully applied in directly measuring the pK_a's of active site residues in the Hairpin (24) and HDV (25) ribozymes, as well as metal binding sites in the latter RNA (26-28). The ligase might serve as another excellent model system for Raman analysis. Since large (~100 μm), lower-resolution (~5Å) crystals with high symmetry seem best suited for the technique (23), an easily-obtained, 4.5Å $P2_12_12_1$ ligase crystal form discovered during the present experiments would appear to be an ideal candidate for further study.

Since some regions of interest are poorly ordered in the P1 crystal form, and as structure-function studies almost always benefit from higher resolution data, a final crystallographic experiment would seek to improve the diffraction quality of ligase crystals. This might be accomplished be redesigning crystal contacts *a posteriori*, using RNA constructs designed to bolster weak crystal contacts observed in the original structure (29, 30). The ligase P1 crystal form is an attractive target for this technique, since the sole contact near the 5′-end of the P1-P2 domain is a poorly-ordered stacking interaction between J1/3 nucleotide A21 and its symmetry mate. As a consequence, density is weakest, and the temperature factors highest at the 5′-ends of J1/3 the P1 duplex. Since adenosines are frequently observed to mediate higher-order RNA

contacts (31) and since this region of J1/3 easily accommodates poly-adenosine expansion (3) a systematic screen of adenosine inserts might bolster the symmetry-mate contact, further ordering helix P1, and yielding higher diffraction. In a similar vein, since 5′-end of the ligase is only ~4Å from the P6 minor groove of a symmetry mate, appending the 5′-terminus with one of more unpaired residues might foster a tertiary contact between these two regions.

Structure-function analyses of the polymerase ribozyme

An exciting route for further exploration lies in dissecting the general polymerase ribozymes that employ the Class I ligase as their catalytic engine (17, 32, 33). Crystallography might seem an ideal experimental approach towards this end, but in their current state the polymerase ribozymes are poor crystallization targets. Since these enzymes are expected to properly fold only when bound to their primer-template (PT) duplexes, and as they exhibit immeasurably high K_Ds for these duplexes (34), it is anticipated that a vanishingly small fraction of molecules in a crystallization experiment would be properly folded. One might circumvent this problem by covalently attaching the PT duplex to the ribozyme core, thus eliminating the bottleneck of bimolecular binding. Indeed, restoration of the polymerase-primer 5′-5′ linkage that was used in the original selection (Figure 1A)(32) results in a ~300-fold boost in activity (34), though a less exotic means toward the same end might involve coupling the 5′-end of the polymerase to its template (Figure 1B). This would restore the enzyme•PT linkage back to its context within the ligase ribozyme, an interaction now well understood given the ligase crystal structure. Due to steric constraints, it is unlikely that such a construct would be able to extend its primer by more than a few additions, but for the purposes of crystallography robust folding is

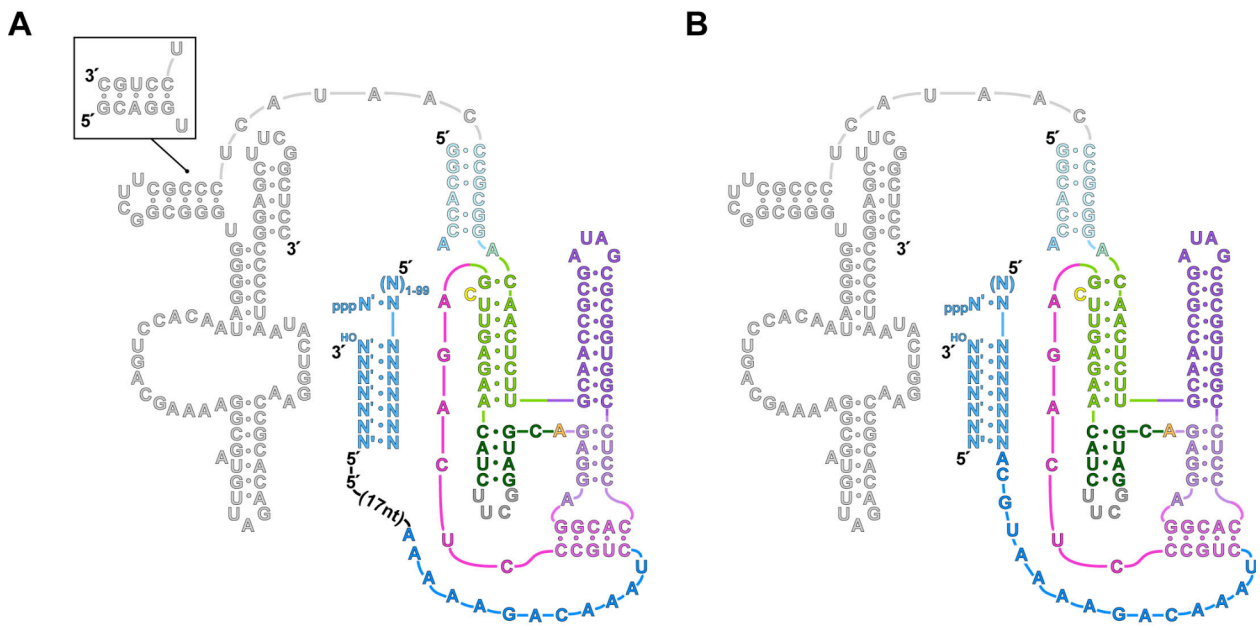


Figure 1. Round-18 Polymerase constructs for biochemical assays and crystallography. N indicates any nucleotide; N' is its Watson-Crick complement. (**A**) Proposed constructs for structural probing assays. The primer and enzyme core are joined by a 5′-5′ linkage, as in (32). The accessory domain (*gray*) can be provided in *cis* or in *trans* (*inset*). (**B**) Proposed construct for crystallography. The ligase core is linked to its template as in the primer-extending ligase variant in (1).

---

more important than processive activity. Moreover, this construct could serve as a model system for studying short polymerization reactions without the confounding effects of limiting PT–binding. In particular, the dependencies on Mg^{2+} concentration and pH might prove key in determining whether or not the polymerase could be crystallized in its current state, since these parameters were critical in the crystallization of the Class I ligase core. In theory, one might seek to improve suboptimal dependencies through further *in vitro* selection, explicitly targeting metal dependence with limiting Mg^{2+}, or targeting folding by isolating compact species from temperature gradient gels (35). While similar selections have previously proved successful, in the current case *in vitro* selection might be better applied toward the general aim of isolating more efficient ribozymes (*see below*), rather than toward the specific aim of improving crystallographic targets.

Even if the current generation polymerases cannot be crystallized, they can still be extensively characterized using the wide range of available RNA biochemical probes. In essence, there are two basic questions one might hope to address in this manner. First, how does the polymerase bind its PT duplex? Second, in those contexts where it catalyzes multiple turnovers (34), how does the polymerase translocate its PT duplex? These might be addressed using variants of the construct depicted in Figure 1A, wherein the ligase core is covalently tethered to its primer, facilitating the isolation of active species from inactive. Conversely, the accessory domain can also be severed from the ligase core and bound in *trans* via base pairing. Parallel experiments in the presence or absence of the accessory domain could be used to discriminate the influence each domain has on the system.

The issue of how the polymerase binds its PT duplex could be addressed by any number of means. Protection studies can provide particularly powerful structural insight: Fe•EDTA mapping would allow broad-scale interrogation of regions that are protected from solvent (36); chemical protection from myriad alkylating agents can assay base- and functional group-specific interactions (37); SHAPE analysis surveys nucleotide flexibility (38). Parallel application of these techniques in the presence or absence of the PT duplex or accessory domain will reveal the structural changes that occur concomitant with substrate binding and enzyme folding.

A number of powerful techniques employ the use of modified nucleotides to acquire detailed structural information, and could easily be applied to the polymerase ribozyme. Since the oligonucleotides of the PT duplex are small enough to be generated by solid phase synthesis, the substrate could be easily modified with unnatural nucleotides at any position. In particular, specific enzyme•PT contacts could be probed via 5-bromo- or 5-iodo-uridine incorporation and photochemical crosslinking, or by using thiolated nucleotides to attach free radical-generating

Iron(II) complexes at specific locations (39). Fluorescent analogs could be used to assay substrate binding, enzyme folding, or conformational dynamics in the steady-state (40, 41). Nucleotide analog interference mapping (NAIM) (6) could be used to obtain broad-scale, detailed structural information, potentially identifying the specific nucleotides and functional groups therein that participate in PT-binding and catalysis. To date, more than 30 modified nucleotides have been synthesized for use in NAIM studies, affording experimenters a truly diverse catalog of structural probes (7). Though many of these compounds are not commercially available, hence making NAIM initially somewhat more technically challenging than crosslinking or protection studies, the technique has the advantage of directly probing for functionally relevant interactions, since it explicitly compares the modification profiles of active species versus inactive.

By exploiting a technology used in the original polymerase selection, many of the above techniques can also be co-opted to study substrate translocation. On mercury-containing poylacrylamide gels, the mobilities of unmodified, singly- and doubly-thiolated RNA species are known to differ dramtically (32). Hence, if the polymerase is provided with a template coding for the addition of two thiolated nucleotides (4-thiouridine, 6-thioguanosine, phosphorothioates, etc…), species that processively extend a primer by two nucleotides can easily be purified from those that add none, or only one. For example, the points of photochemical crosslinking between a 5-iodouridine residue on the primer and the polymerase could be mapped prior to primer extension, and as the PT duplex is translocated through one or two rounds of addition. NAIM could be used to systematically map functional groups that are required for a single addition, but not multiple turnovers, or vice-versa. Since the polymerase in general exhibits variable processivity, these initial experiments would have to exploit those sequence contexts that are

known to be most efficient for multiple turnovers (34). However, the residues identified as having the most profound effect on processivity in this context could be directly targeted through mutagenesis, modification or *in vitro* selection. Monitoring and optimizing the effects of these modifications in the context of less efficient PT duplexes might prove a productive route in enhancing polymerase processivity overall.

In vitro *selection of superior polymerase ribozymes*

While extensive characterization of the Round-18 polymerase would invariably provide a fascinating glimpse into the catalytic capabilities of RNA, it might prove an unproductive route for generating a replicase ribozyme. Indeed, other polymerase *in vitro* selection experiments–starting from the original polymerase selection pool (33), or using one derived from the Round-18 polymerase (17)–have produced only modest results. Thus, it is possible that the current isolates represent the optimal catalysts that can be obtained given the stringency of our methods and the design of our pools. Isolating a true replicase ribozyme might require a redesign of the original polymerase selection, improving on both of these factors. Since, for the current isolates, the limiting activity appears to be PT-binding, explicitly targeting this activity would likely be the most fruitful approach in future attempts.

The series of selections that ultimately produced the Round-18 polymerase followed an incremental design philosophy. A relatively simple activity was initially targeted for selection, yielding a ligase ribozyme that could perform a single round of the desired chemistry. This was subsequently optimized for catalytic efficiency, and then used as the starting point for selecting a more complicated activity, general RNA polymerization. While this incremental approach was only applied in regard to catalysis, and did not explicitly select for substrate binding activity in

isolation, a similar design philosophy could be used for generating novel, more highly processive accessory domains. First, an aptamer would be selected that generally binds A-form RNA duplexes; selection of this aptamer would need to stringently avoid species that recognized their ligands in a sequence-dependent fashion. Design of a new *in vitro* selection pool would exploit known structure of the ligase catalytic core (*see below*), but would begin by appending this aptamer onto the ligase core in a strategically advantageous point. As has proven useful in the selection of allosteric ribozymes (42, 43), short regions of random sequence could bracket the aptamer insertion point, allowing the selection of an efficient "communication module" between the catalytic and accessory domains. Rounds of selection and optimization would ensue.

Given a different pool design (*see below*) either the incremental strategy outlined above, or a revised version of the original strategy could prove successful. In either case, however, the selection itself would need to apply more stringent pressure for *trans* activity and processivity. The *in vitro* compartmentalization (IVC) method developed by Zaher and Unrau (17) seems extremely promising in this regard. Briefly, the RNA PT-duplex substrate is ligated to the end of the pool DNA, and this pool is emulsified in oil droplets along with the components needed for T7 RNA polymerization and ribozyme activity. Active polymerases are selected owing to their ability to modify their templates *in trans*, not themselves. While the resulting B6.61 polymerase borne from this proof-of-principle experiment was only moderately improved compared to its parent, future selections might couple this compartmentalization technique with more explicit demands on processivity. For example, templates could be designed that require polymerases to extend their primers through longer sequences before incorporating the selective 4-thiouridine residues (32, 33), though this presents the possible disadvantage of selecting sequences that nonspecifically decorate themselves with 4-thiouridine. To circumvent this, the traditional

method of selection from a mercury gel could be supplemented by an oligo-capture step, in which extended primers are targeted with tagged DNA oligonucleotides complementary to progressively longer and longer extension products.

The pools used in this selection should be radically redesigned as well, which can be aided by knowledge of the ligase crystal structure. Given that we now know the interactions through which the P1-P2 domain–analgous to the polymerase PT duplex–is presented into the ligase core, future pools should be designed to place PT-binding modules or random sequences in positions where they would mimic these interactions. A few such pool designs are presented in Figure 2.

A binding module for the double-stranded region of PT duplex could optimally be inserted into or evolved from the 5′-end of J1/3 (Figure 2A). This region decorates the P1 duplex in the ligase, though the observed interactions appear relatively weak. Moreover, the identity of these residues has already been shown to be relatively malleable in previous selections (3, 17, 32, 44), making it an ideal spot for the insertion of a selected A-form RNA aptamer, or stretches of random sequence. A replicase ribozyme would also need to bind the single-stranded template and incoming NTPs, though selection for sequence-independent apatmers for either of these activities would be difficult. Hence, the best route in this case would be to evolve a functional domain within the context of the ligase, as was previously done for the Round-18 polymerase accessory domain. For this purpose, the most obvious insertion point would be P2 (Figure 2A), which, in cohort with J1/2, essentially serves the same purposes in the ligase core.

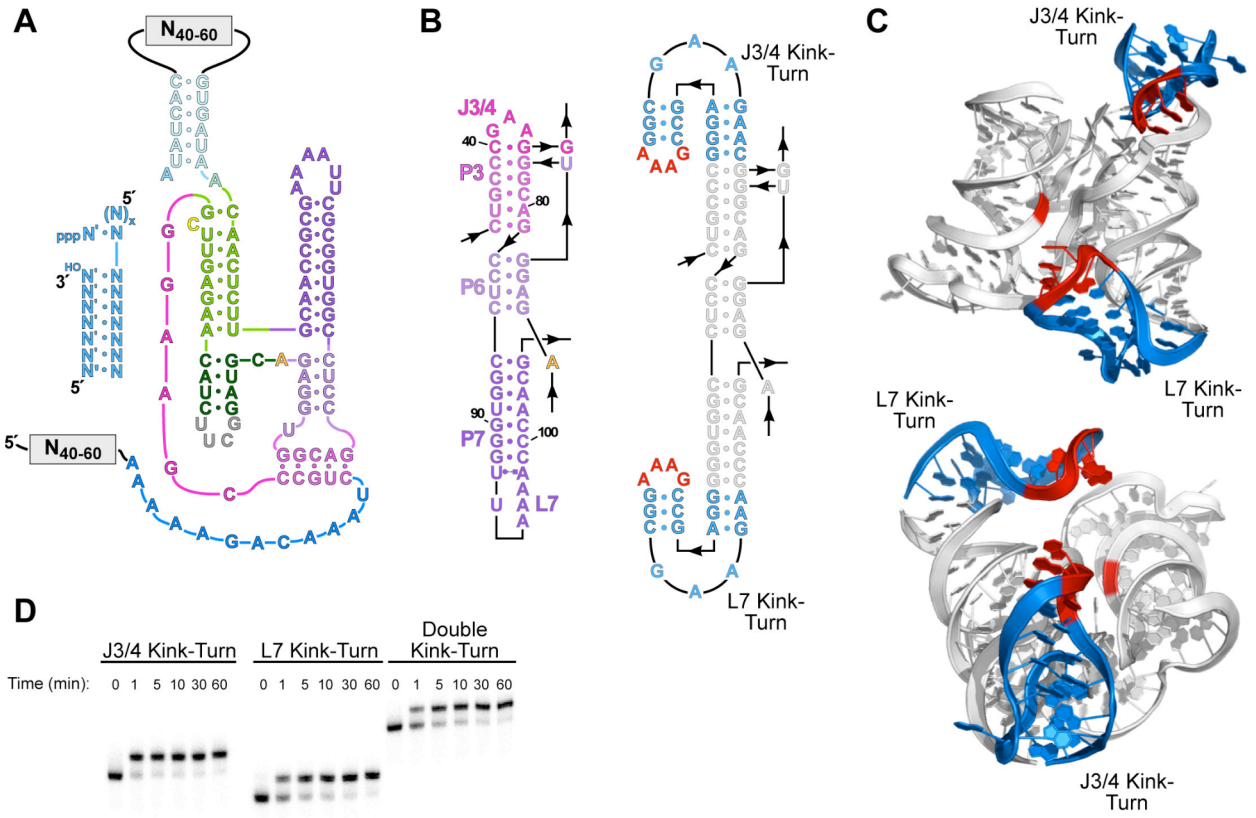


Figure 2. Pool designs for future polymerase selections. (**A**) Appending J1/3 (*bottom*) and P2 (*top*) with random sequences, or pre-selected modules (*not shown*). $N_{40\text{-}60}$ refers to random sequences 40–60 nt in length. (**B**) Structure-aided pool design. (*Left*) Secondary structure of the P3-P6-P7 domain, redrawn reflecting its conformation in the crystal structure (20). Coloring is as in (A). The remainder of the ligase core is not shown. (*Right*) The same domain appended with kink-turn motifs (*blue*, (45)), capped with GNRA tetraloops (*red*). For *in vitro* selection pools, these tetraloops would be replaced with stretches of random sequence, as in (A). (**C**) Rationale for the design in (B). Two views of a model of the double kink-turn ligase construct, modeled using (45). Coloring is the same as in (B); the ligation junction is highlighted in red. Note that the J3/4 kink-turn projects its tetraloop near the P2 helix, which would be the template in polymerase selections; the L7 kink turn is proximal to P1, analogous to the PT duplex. Figures made with PyMol (46). (**D**) Activity of ^{32}P body-labeled kink-turn ligases. In each experiment, the lower band corresponds to the unreacted ligase; the upper band corresponds to the ligated product. At pH 7.0, the observed rates for the J3/4, L7 and double kink-turn constructs are 1.9, 0.9 and 0.7 min^{-1}, respectively.

Since self-replication by a long polymerase species would require it to synthesize long products, an advantage to pool described above is that it is relatively short. A potential downside to this design, however, is that the random sequence insertions may disrupt the folding of the

ligase core domain. At the expense of sequence length, then, one might hope to leave the ligase core intact–or even buttress it with stably folding elements–and use structure aided design to position potential PT-binding domains near their desired points of action. One such design is shown in Figure 2B–C, wherein L7 and the J3/4 triloop are replaced with kink-turn motifs (45) that might be used to place random sequences near the PT-duplex and template reading frame, respectively. Appending the *cis*-ligase with either or both of these kink-turns decreases activity by only ~3-fold (Figure 2D), hence demonstrating that neither domain dramatically disrupts folding of the enzyme core. New polymerase pools based on this design, or the one outlined above, selected under stringent conditions that explicitly require *trans*-substrate binding, might ultimately produce a true self-replicase ribozyme.

References

1. M. E. Glasner, C. C. Yen, E. H. Ekland, D. P. Bartel. (2000) Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. *Biochemistry.* **39**, 15556.
2. M. E. Glasner, N. H. Bergman, D. P. Bartel. (2002) Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry.* **41**, 8103.
3. S. C. Bagby, N. H. Bergman, D. M. Shechner, C. Yen, D. P. Bartel. (2009) A class I ligase ribozyme with reduced Mg2+ dependence: Selection, sequence analysis, and identification of functional tertiary interactions. *RNA.* **15**, 2129.
4. J. A. Piccirilli, J. S. Vyle, M. H. Caruthers, T. R. Cech. (1993) Metal ion catalysis in the Tetrahymena ribozyme reaction. *Nature.* **361**, 85.
5. J. K. Frederiksen, J. A. Piccirilli. (2009) Identification of catalytic metal ion ligands in ribozymes. *Methods.* **49**, 148.
6. S. A. Strobel, K. Shetty. (1997) Defining the chemical groups essential for Tetrahymena group I intron function by nucleotide analog interference mapping. *Proc Natl Acad Sci U S A.* **94**, 2903.
7. J. C. Cochrane, S. A. Strobel. (2004) Probing RNA structure and function by nucleotide analog interference mapping. *Curr Protoc Nucleic Acid Chem.* **Chapter 6**, Unit 6 9.
8. S. H. Kim, W. C. Shin, R. W. Warrant. (1985) Heavy metal ion-nucleic acid interaction. *Methods Enzymol.* **114**, 156.
9. J. H. Cate, J. A. Doudna. (1996) Metal-binding sites in the major groove of a large ribozyme domain. *Structure.* **4**, 1221.
10. P. L. Adams, M. R. Stahley, A. B. Kosek, J. Wang, S. A. Strobel. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature.* **430**, 45.
11. M. R. Stahley, S. A. Strobel. (2005) Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science.* **309**, 1587.
12. N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle. (2008) Crystal structure of a self-spliced group II intron. *Science.* **320**, 77.
13. D. A. Harris, N. G. Walter. (2003) Probing RNA structure and metal-binding sites using terbium(III) footprinting. *Curr Protoc Nucleic Acid Chem.* **Chapter 6**, Unit 6 8.

14. E. H. Ekland, D. P. Bartel. (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature.* **382**, 373.
15. S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature.* **391**, 251.
16. E. T. Kool. (2001) Hydrogen bonding, base stacking, and steric effects in dna replication. *Annu Rev Biophys Biomol Struct.* **30**, 1.
17. H. S. Zaher, P. J. Unrau. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA.* **13**, 1017.
18. J. J. Dunn, F. W. Studier. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol.* **166**, 477.
19. F. Huang, C. W. Bugg, M. Yarus. (2000) RNA-Catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry.* **39**, 15548.
20. D. M. Shechner *et al.* (2009) Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science.* **326**, 1271.
21. J. F. Sydow *et al.* (2009) Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Mol Cell.* **34**, 710.
22. A. Luptak, A. R. Ferre-D'Amare, K. Zhou, K. W. Zilm, J. A. Doudna. (2001) Direct pK(a) measurement of the active-site cytosine in a genomic hepatitis delta virus ribozyme. *J Am Chem Soc.* **123**, 8447.
23. B. Gong *et al.* (2009) Raman crystallography of RNA. *Methods.* **49**, 101.
24. M. Guo *et al.* (2009) Direct Raman measurement of an elevated base pKa in the active site of a small ribozyme in a precatalytic conformation. *J Am Chem Soc.* **131**, 12908.
25. B. Gong *et al.* (2007) Direct measurement of a pK(a) near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography. *J Am Chem Soc.* **129**, 13335.
26. B. Gong *et al.* (2008) Detection of innersphere interactions between magnesium hydrate and the phosphate backbone of the HDV ribozyme using Raman crystallography. *J Am Chem Soc.* **130**, 9670.

27. J. H. Chen, B. Gong, P. C. Bevilacqua, P. R. Carey, B. L. Golden. (2009) A catalytic metal ion interacts with the cleavage Site G.U wobble in the HDV ribozyme. *Biochemistry.* **48**, 1498.

28. B. Gong, J. H. Chen, P. C. Bevilacqua, B. L. Golden, P. R. Carey. (2009) Competition between Co(NH(3))(6)(3+) and inner sphere Mg(2+) ions in the HDV ribozyme. *Biochemistry.* **48**, 11961.

29. C. MacElrevey, R. C. Spitale, J. Krucinska, J. E. Wedekind. (2007) A posteriori design of crystal contacts to improve the X-ray diffraction properties of a small RNA enzyme. *Acta Crystallogr D Biol Crystallogr.* **63**, 812.

30. R. K. Montange *et al.* (2009) Discrimination between Closely Related Cellular Metabolites by the SAM-I Riboswitch. *J Mol Biol.*

31. P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, T. A. Steitz. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci U S A.* **98**, 4899.

32. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science.* **292**, 1319.

33. M. S. Lawrence, D. P. Bartel. (2005) New ligase-derived RNA polymerase ribozymes. *RNA*. **11**, 1173.

34. M. S. Lawrence, D. P. Bartel. (2003) Processivity of ribozyme-catalyzed RNA polymerization. *Biochemistry.* **42**, 8748.

35. K. Juneau, T. R. Cech. (1999) In vitro selection of RNAs with increased tertiary structure stability. *RNA*. **5**, 1119.

36. D. W. Celander. (2001) Probing RNA structures with hydroxyl radicals. *Curr Protoc Nucleic Acid Chem.* **Chapter 6**, Unit 6 5.

37. W. A. Ziehler, D. R. Engelke. (2001) Probing RNA structure with chemical reagents and enzymes. *Curr Protoc Nucleic Acid Chem.* **Chapter 6**, Unit 6 1.

38. K. A. Wilkinson, E. J. Merino, K. M. Weeks. (2006) Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc.* **1**, 1610.

39. S. Joseph, H. F. Noller. (2000) Directed hydroxyl radical probing using iron(II) tethered to RNA. *Methods Enzymol.* **318**, 175.
40. N. G. Walter. (2001) Structural dynamics of catalytic RNA highlighted by fluorescence resonance energy transfer. *Methods.* **25**, 19.
41. S. Liu, G. Bokinsky, N. G. Walter, X. Zhuang. (2007) Dissecting the multistep reaction pathway of an RNA enzyme by single-molecule kinetic "fingerprinting". *Proc Natl Acad Sci U S A.* **104**, 12634.
42. G. A. Soukup, R. R. Breaker. (1999) Engineering precision RNA molecular switches. *Proc Natl Acad Sci U S A.* **96**, 3584.
43. A. Roth, R. R. Breaker. (2004) Selection in vitro of allosteric ribozymes. *Methods Mol Biol.* **252**, 145.
44. E. H. Ekland, D. P. Bartel. (1995) The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231.
45. D. J. Klein, T. M. Schmeing, P. B. Moore, T. A. Steitz. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214.
46. W. L. Delano. (2002) The PyMOL Molecular Graphics System. *http://pymol.sourceforge.net.* DeLano Scientific.

A class I ligase ribozyme with reduced Mg^{2+} dependence: Selection, sequence analysis, and identification of functional tertiary interactions

SARAH C. BAGBY,[1,2,3,4] NICHOLAS H. BERGMAN,[1,2,3,5] DAVID M. SHECHNER,[1,2] CATHERINE YEN,[1,2] and DAVID P. BARTEL[1,2]

[1]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA
[2]Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

ABSTRACT

The class I ligase was among the first ribozymes to have been isolated from random sequences and represents the catalytic core of several RNA-directed RNA polymerase ribozymes. The ligase is also notable for its catalytic efficiency and structural complexity. Here, we report an improved version of this ribozyme, arising from selection that targeted the kinetics of the chemical step. Compared with the parent ribozyme, the improved ligase achieves a modest increase in rate enhancement under the selective conditions and shows a sharp reduction in [Mg^{2+}] dependence. Analysis of the sequences and kinetics of successful clones suggests which mutations play the greatest part in these improvements. Moreover, backbone and nucleobase interference maps of the parent and improved ligase ribozymes complement the newly solved crystal structure of the improved ligase to identify the functionally significant interactions underlying the catalytic ability and structural complexity of the ligase ribozyme.

Keywords: in vitro selection; class I ligase; ribozyme; RNA structure; NAIM; interference mapping

INTRODUCTION

The first successful use of in vitro evolution to produce a ribozyme from completely random sequence yielded the class I ligase, a 119-nucleotide (nt) sequence that promotes formation of a phosphodiester bond between the 3′-hydroxyl group of an oligonucleotide substrate and its own 5′ α-phosphate (Bartel and Szostak 1993; Ekland et al. 1995). Since its isolation, the class I ligase ribozyme sequence has been the springboard for a number of other studies of non-natural ribozymes and evolution, being subjected to continuous evolution under approximately constant conditions (Wright and Joyce 1997), at changing pH (Kühne and Joyce 2003), at decreasing Mg^{2+} concentration (Schmitt and Lehman 1999), and in the presence of a ''predator'' DNA enzyme (Ordoukhanian and Joyce 1999), among other conditions. It was the class I ligase, too, that formed the basis of the in vitro compartmentalization work of Levy et al. (2005) that yielded the first ribozyme variants directly selected for multiple-turnover activity.

Another reason for continued interest in the class I ligase is its connection to the RNA world hypothesis, the simplifying idea that life in its earliest stages relied not on a DNA genome and protein enzymes, but on an RNA genome with self-replicating ribozyme activity (Joyce and Orgel 1999). An RNA replicase ribozyme is thus the sine qua non of the RNA world, but this complex activity has proved elusive. The ligase catalyzes a simpler reaction that models this activity. Relying on Watson–Crick base-pairing, as an RNA replicase would, to orient its substrate, the ligase catalyzes the attack of a ribonucleotide 3′-OH on a 5′-triphosphate, with expulsion of a pyrophosphate leaving group—precisely the chemistry an RNA replicase needs for each NTP addition. These features have made the ligase a useful scaffold on which to build template-directed RNA polymerase ribozymes that represent progress toward generating true replicase activity (Ekland and Bartel 1996; Johnston et al. 2001; McGinness et al. 2002; Lawrence and Bartel 2005; Zaher and Unrau 2007).

[3]These authors contributed equally to this work.
Present addresses: [4]Marine Science Institute, University of California, Santa Barbara, CA 93106, USA; [5]Battelle National Biodefense Institute, National Biodefense Analysis and Countermeasures Center, Frederick, MD 21702, USA.

Reprint requests to: David P. Bartel, Whitehead Institute for Biomedical Research, Massachusetts Institute of Biology, 9 Cambridge Center, Cambridge, MA 02142, USA; e-mail: dbartel@wi.mit.edu; fax: (617) 258-6768.

Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1912509.

Replicative potential aside, the class I ligase ribozyme is studied because it is fast. At 60 mM Mg^{2+} and pH 8.0, the *cis*-acting construct b1-207 reacts with k_c = 300 min^{-1} (Glasner et al. 2002), while the *trans*-acting construct b1-207t reaches k_c = 375 min^{-1} for the multiple turnover reaction (Bergman et al. 2000). For the latter construct, k_{cat}/K_M is 7 × 10^7 M^{-1} min^{-1}, just over an order of magnitude shy of the diffusion-controlled limit. As fast as class I ligase catalysis is, however, the chemical step remains slower than the folding reaction at pH ≤7 (Glasner et al. 2002). Thus, we decided to target the chemical step for improvement by using a rapid-quench flow apparatus to isolate variants that react rapidly at moderate-to-low pH. This targeted selection yielded a family of successful clones, from which we chose one, termed the improved ligase, for further investigation.

Under optimal conditions (60 mM Mg^{2+} at pH 9), the improved ligase ribozyme had a k_c exceeding 1300 min^{-1}, among the fastest ribozyme rate constants yet observed. The most successful isolates also showed a shallower dependence on [Mg^{2+}] than did the parent ligase, suggesting that the mutations acquired either improve the affinity of one or more specific Mg^{2+}-binding sites or buttress ribozyme structure by some metal-independent means. Statistical analysis of the population of successful isolates allowed us to identify mutations that are likely to have contributed to this improved activity and reduced metal dependence. To probe the tertiary interactions that contribute to ribozyme activity across the generations, we performed nucleotide analog interference mapping (NAIM) (Conrad et al. 1995; Strobel and Shetty 1997; Ryder and Strobel 1999) and dimethyl sulfate (DMS) interference mapping (Peattie and Gilbert 1980; Peattie and Herr 1981; Moazed et al. 1986; Stern et al. 1988) on both the parent ligase and the improved ligase. Our results complement the recently solved crystal structure of the improved ligase (Shechner et al. 2009) to reveal key functional groups in the network of interactions that promote activity in the class I ligase ribozyme.

RESULTS AND DISCUSSION

Selection of improved class I ligase variants

Selection experiments face a trade-off between the number of positions varied and the completeness with which sequence space can be covered at those positions. Previous selection experiments indicated that many of the unpaired residues within the ligase were important for activity, and suggested that single-stranded regions may play a role in defining the tertiary structure and active site of the ribozyme (Ekland et al. 1995). Thus, we sought to explore more fully the possible sequences and lengths of those regions, while holding the known stem regions largely constant (Fig. 1A). Where conservation in previous selections suggested that the optimal residue in a joining or loop region was already in place, we biased the pool toward that residue but mutagenized at the level of 10% (i.e., 90% of the pool had the parental base, while 3.3% had each of the nonparental bases). At the remaining positions in these regions, we randomized the pool completely in both length and sequence. Only one base pair in the seven stems was allowed to vary: both bases in the noncanonical G88:A103 pair in stem P7 were mutagenized at the 10% level. Finally, four engineered changes were made in stem regions of all pool molecules. Two of these changes sought to improve ribozyme function: first, G73:C84 was converted to an A:U pair, a change that improves class I ligase-catalyzed primer extension (EH Ekland and DP Bartel, unpubl.); second, helix P5 was extended by one base pair to improve its stability, so that L5 optimization could take place in a more constant environment. The remaining two changes facilitated the selection itself by making the substrate approximate the sequence of the T7 RNA polymerase promoter: the 3′ end of the substrate, forming the 5′ strand of P1, was changed from 5′-CCAGUC-3′ to 5′-CCAGUA-3′, and G13, in the 3′ strand of P1, was changed to U in order to maintain Watson–Crick pairing with this new substrate.

The engineered and mutagenized pool was subjected to successive rounds of selection in a scheme that differed from previous ligase selections in two key respects (Supplemental Fig. S1). First, the substrate used resembled the T7 RNA polymerase promoter, allowing amplification to proceed from reverse transcription (RT) and selective PCR directly to forward transcription without the intervening steps previously required to append the T7 promoter (Bartel and Szostak 1993; Ekland et al. 1995). Second, we sought to select specifically for rapid catalysis, by performing the selective step of later rounds in a rapid-quench flow apparatus. This apparatus enabled us to select for ribozymes capable of performing ligation in as little as 0.2 sec.

The sequences isolated after seven rounds of selection (Fig. 1B) included 35 unique clones, which were then used as templates for transcription of ribozyme RNAs. These ribozymes remained active in the absence of the reverse-transcription primer and its binding site; indeed, removal of the binding site typically improved activity slightly (data not shown). Similarly, activity was unaffected by 5′-terminal truncation of the 16-nt RNA–DNA chimeric substrate. All further characterization of the successful clones was performed using a 7-nt RNA substrate and ribozymes trimmed at the 3′ end to remove the RT primer-binding site, matching the length of the parent ligase.

We first measured the ligation activity of the 35 isolates under the conditions of the final round of selection (pH 6.0, 10 mM Mg^{2+}, 200 mM KCl). Ligation rate constants ranged from 0.15 min^{-1} to >2 min^{-1}, with a cluster of seven isolates (Table 1, clones 22, 23, 80, 91, 96, 101, 141) ligating roughly twice as fast as the parent ligase. We next asked whether having performed the selection at Mg^{2+}

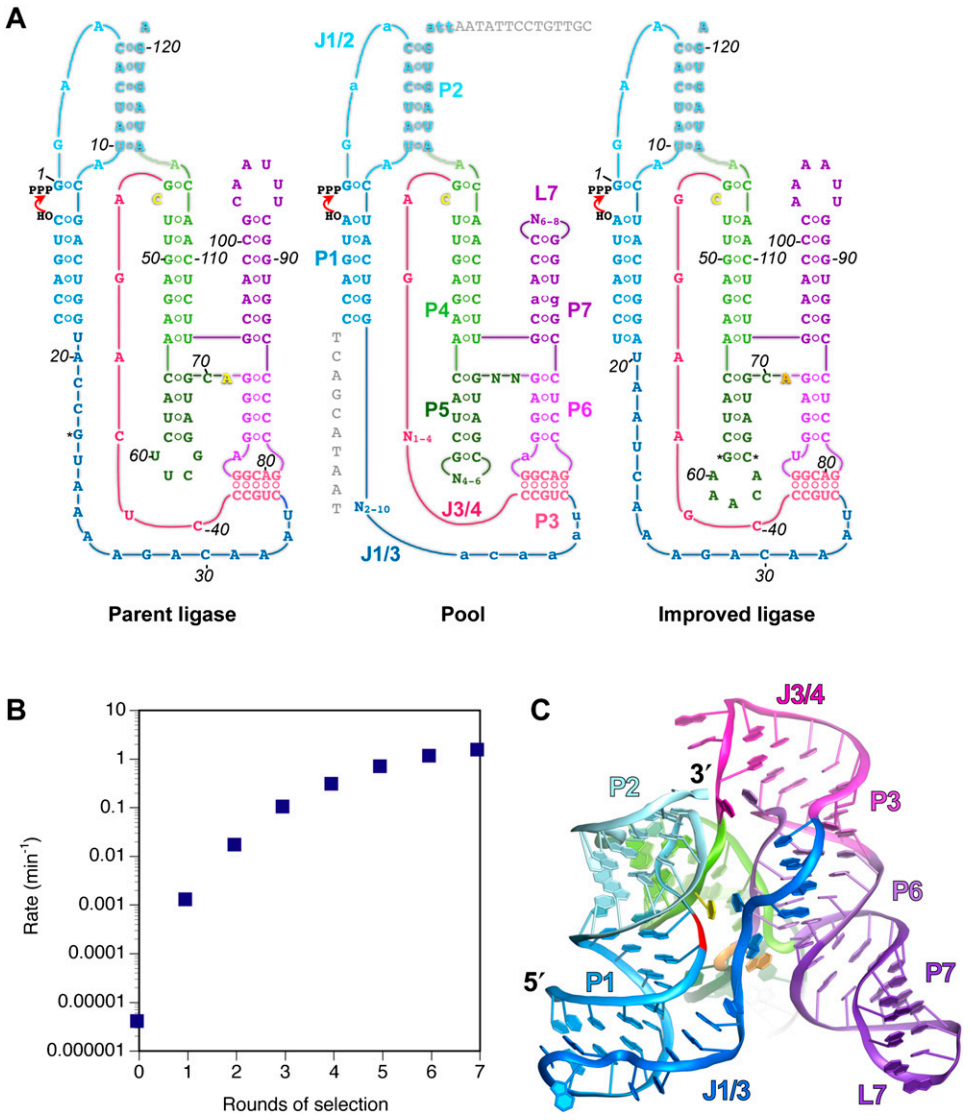


FIGURE 1. (*A*) Secondary structures of the parent ligase, the pool of sequences used for selection, and the improved ligase (clone 23). Secondary structural elements (common to all three) are labeled on the pool. Joining regions (J) are named for the two paired regions (P) they connect. Residue numbering is with respect to the ligation junction, with the 5′-terminal residue of ribozyme assigned +1, and the 3′-terminal residue of the substrate assigned –1. All numbering is with reference to the improved ligase; insertions are marked with asterisks. Positions that were randomized in the pool are labeled N on the pool secondary structure; positions mutagenized at the 10% level are shown in lowercase. Sequences in gray on the pool secondary structure are, at *left*, the DNA portion of the substrate used for selection and, at *top*, the RT primer-binding site. (*B*) Improvement of pool activity over the course of the seven rounds of selection. Rates were measured at pH 6.0 in 10 mM Mg^{2+}. (*C*) Crystal structure of the product of the improved ligase (Shechner et al. 2009). Elements of secondary structure are colored as in *A*.

levels known to be below saturation for the parent ligase had produced any change in Mg^{2+} dependence among the successful isolates. The parent ligase has a steep dependence on [Mg^{2+}] in this range, its rate constant dropping 40-fold when the Mg^{2+} concentration is reduced from 10 mM to 1 mM. Among the selected clones, five showed as steep a dependence as that of the parent; the remainder showed a shallower dependence, with three clones (Table 1, clones 187, 66, 23) dropping as little as 3.5- to 5.5-fold. Of these three, clones 187 and 66 were no faster than was the parent ligase at 10 mM Mg^{2+}, whereas clone 23 was twofold faster than the parent at 10 mM Mg^{2+}, and 15-fold faster at 1 mM. We chose clone 23 (Fig. 1C), hereafter referred to as the improved ligase, for further biochemical and structural characterization (and number all nucleotide positions, including those of the parent, with respect to this clone).

Sequence analysis of improved class I ligase variants

Before undertaking further biochemical experiments, we subjected the set of 35 isolates to statistical analysis to see whether we could identify candidate mutations underlying

TABLE 1. Rate constants for the 35 isolated ligase variants, measured at pH 6.0 in 10 mM and 1 mM Mg^{2+}

Clone	$k_{10\ mM}$ (min^{-1})	$k_{1\ mM}$ (min^{-1})
23	2.15	0.40
101	2.50	0.28
96	2.40	0.25
22	2.25	0.26
91	2.20	0.13
80	2.15	0.21
141	2.05	0.11
159	1.70	0.057
84	1.65	0.13
173	1.55	0.13
178	1.50	0.06
50	1.50	0.11
162	1.45	0.074
69	1.45	0.096
55	1.40	0.069
89	1.40	0.108
70	1.35	0.12
71	1.30	0.13
175	1.20	0.029
68	1.20	0.067
66	1.10	0.32
2	1.10	0.034
106	1.05	0.032
18	1.00	0.023
172	0.80	0.031
77	0.80	0.020
180	0.65	0.029
1	0.60	0.016
186	0.55	0.036
124	0.55	0.036
153	0.50	0.058
35	0.40	0.010
61	0.40	0.013
187	0.35	0.10
158	0.15	0.005

the observed improvements in catalysis. We aligned the sequences of all isolates (Fig. 2) and compared the nucleotide distribution at each mutagenized position to that of the starting pool (Fig. 3). The first fully randomized region of the ribozyme, positions 19–28, lies in a long joining region, J1/3. Although this segment was varied randomly in the pool from 2 to 10 nt in length, the successful isolates were tightly clustered at 10 nt (Fig. 3C), with five clones even acquiring additional nucleotides during the course of evolution. Notably, this is one nucleotide shorter than the parent J1/3; it is not clear whether this change is connected to the 1-nt extension of P1 made possible by the modified ligase substrate. A strong nucleotide composition bias was also evident in this region of the isolates, with adenosine constituting 74% of the nucleotides in this region, compared with 50% in this region of the parent ligase. Remarkably, the least significant nucleotide enrichment in this region was the preference for G at position 19 with $P = 0.011$ (Fisher's exact test); positions 20–27 were all preferentially an A with P values ranging from 4×10^{-5} to 4×10^{-12}, and position 28 was preferentially a G with $P = 9 \times 10^{-11}$.

The 5′ end of the joining segment J3/4 showed somewhat less sequence bias, but a clear length preference. The starting pool varied from 1 to 4 nt here, but 30 out of 35 successful clones maintained the parental length, 4 nt; and in the remaining five clones J3/4 was shortened only to 3 nt. Nucleotides 42 and 43 showed no significant trends, but nucleotide 41 was largely conserved as the parental U ($P = 0.004$), and the 5′ nucleotide of J3/4, position 40, was strongly conserved as the parental C ($P = 3 \times 10^{-9}$). Explaining this conservation, C40 pairs with G44, as shown by the crystal structure and accompanying experiments (Shechner et al. 2009). Indeed, both isolates that deviated from the parental C40 had a U at this position, which retained the potential to form either a U:G wobble or U:A base pair with nucleotide 44 (a position not intentionally mutagenized in our pool). The final joining segment to be randomized, J5/6, hewed to the parent ribozyme in both sequence and length. J5/6 is 2-nt long in the parent ligase and in 34 of 35 successful isolates, and 1 nt in the other. Even the sole shortened isolate maintains A71, which was absolutely conserved among successful isolates ($P = 4 \times 10^{-12}$). The 5′ position of J5/6, though not absolutely conserved, showed significant bias toward the parental C ($P = 0.003$).

In contrast with the joining regions, the two terminal loops, L5 and L7, showed little conservation in either size or sequence, consistent with previous observations that perturbing these loops has little impact on ligase activity (Ekland and Bartel 1995; Ordoukhanian and Joyce 1999; Schmitt and Lehman 1999). Only three positions in these two loops showed evidence of selection: nucleotides 62 and 64 in L5, and nucleotide 92 in L7. (Note that the improved ligase, clone 23, has a 5-nt L5, whereas 20 other clones have 6 nt here, with the inserted nucleotide falling between improved ligase nucleotides 62 and 63.) Adenosine was favored at both L5 sites, and guanosine at the L7 site; in none of these cases was the favored nucleotide selected to the exclusion of any other. In loops L5 and L7 alike, the first and last nucleotides of loops of the largest allowable size were often (in 12 of the 20 6-nt L5 sequences and 14 of the 18 8-nt L7 sequences) Watson–Crick or wobble complements, such that the paired stem could be extended by one base pair and the loop decreased by 2 nt.

Among the 19 positions mutagenized at the 10% level, two showed significant movement away from the parental identity. The first of these, nucleotide 76, is the sole nucleotide of J6/3; an A in the parent ligase and in 90% of the pool, it remained an A in just 17 isolates, becoming a C in one and a U in the other 17 ($P = 2 \times 10^{-6}$). The second such position, nucleotide 103, is part of the non-canonical G88:A103 base pair in helix P7. Remaining a G:A pair in nine isolates, mutation of nucleotide 103 gave a G:U

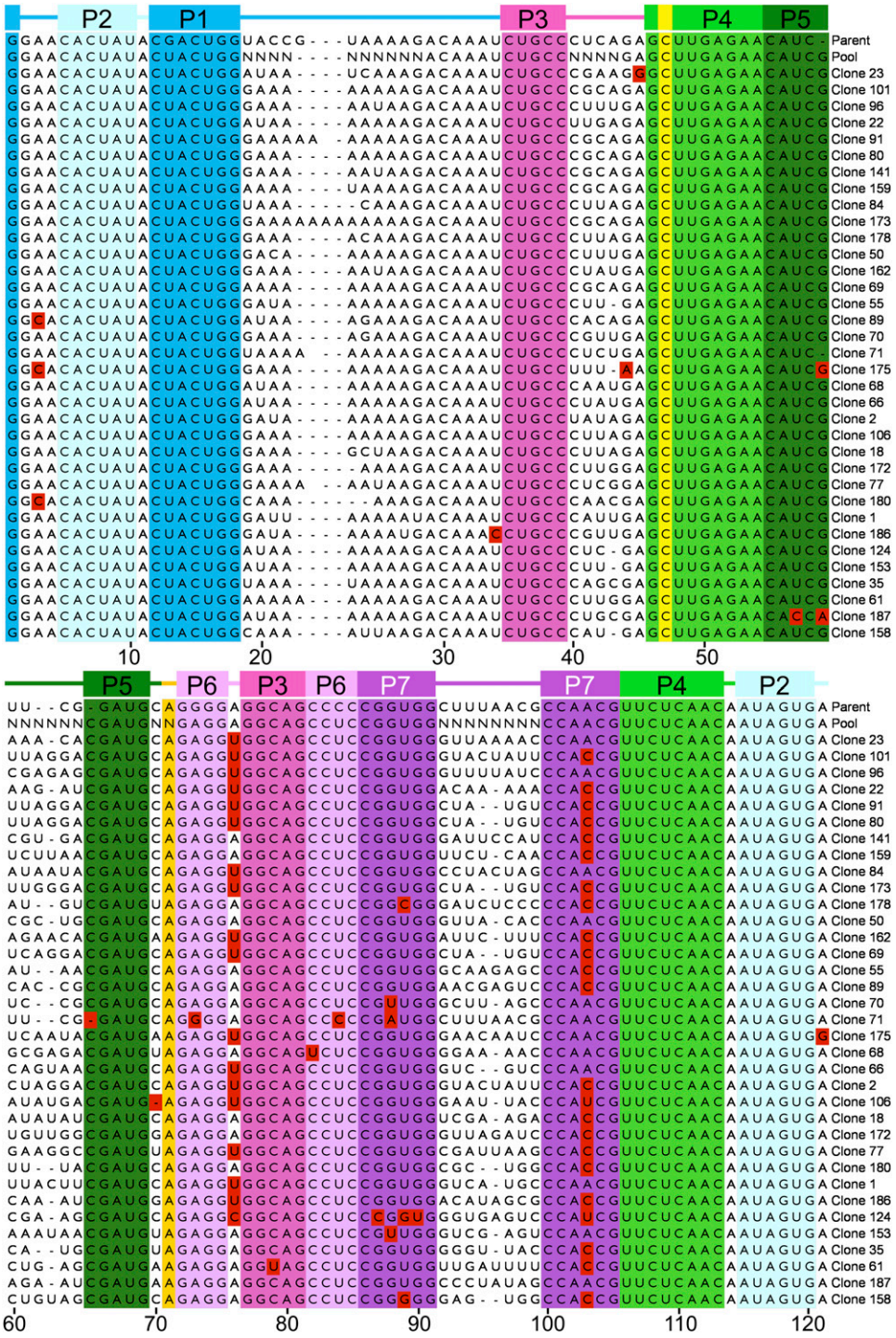


FIGURE 2. Alignment of the 35 isolated variants. Helix P1 is formed when the substrate oligonucleotide hybridizes to nucleotides 13–18. Red blocks highlight nonparental nucleotides in regions that were not fully randomized in the starting pool. Colors are as in Figure 1.

wobble in two isolates and a canonical G:C pair in 21 ($P = 1 \times 10^{-8}$). One clone lost the G:A pair only to gain an A:A mismatch by mutation of G88. Although G88, like A103, was also mutagenized at the 10% level, conversion of the G88:A103 mismatch to a U:A pair was observed only twice. This asymmetry suggests that if canonical pairing at this site confers a selective advantage, the benefit may be tied to the stability of the base pair, or may be secondary to a stronger bias for a purine at position 88.

Although Fisher's exact test can readily detect bias toward nonparental nucleotides at lightly mutagenized positions, the test lacks the statistical power to detect

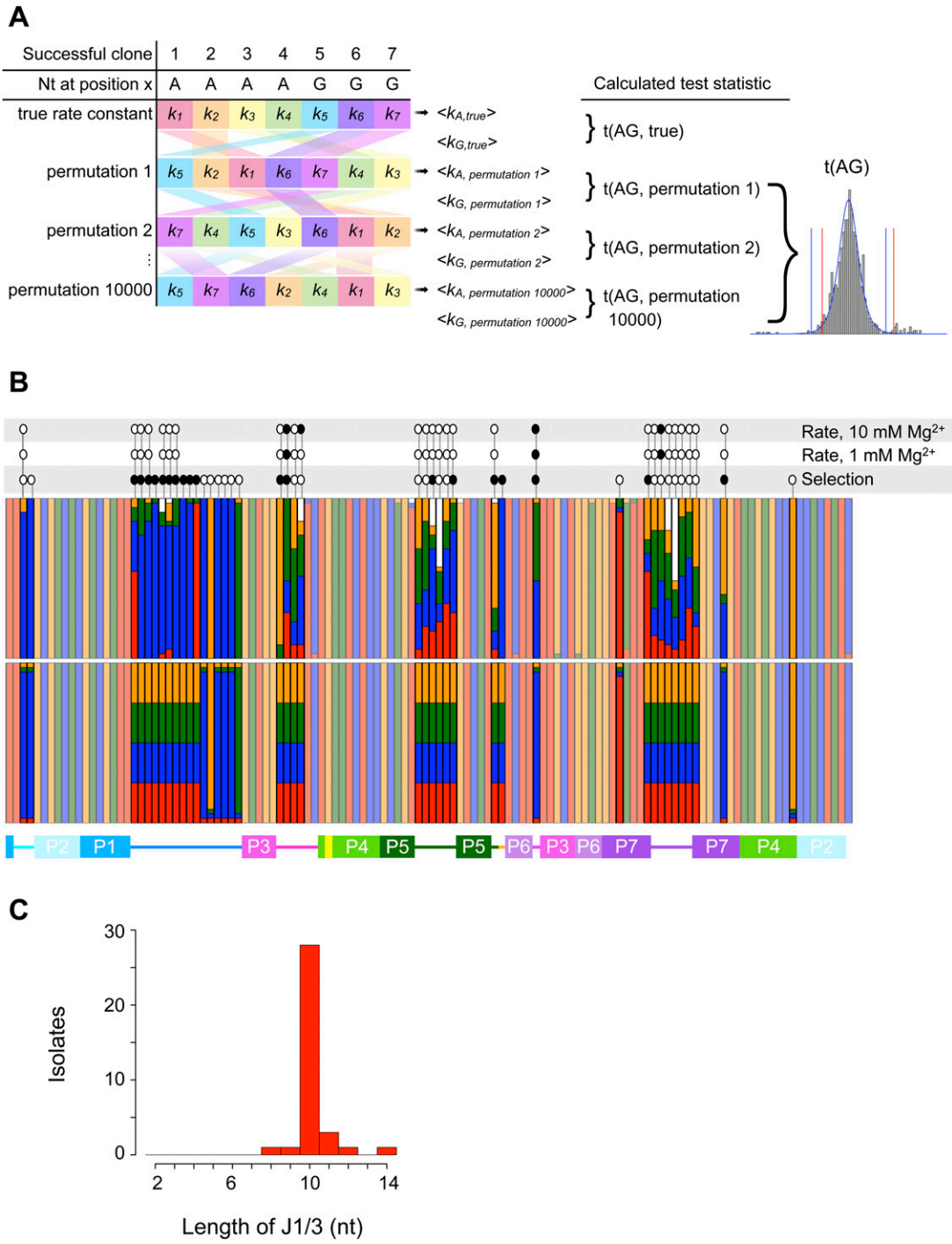


FIGURE 3. (*A*) Outline of Monte Carlo analysis of the kinetic effects of different nucleotides at each position in the ribozyme. At each position subjected to analysis, the pairings of nucleotide identity and rate constant were shuffled randomly 10,000 times, and the mean rate constants newly associated with each nucleotide were calculated. The *t*-statistic describing the difference in mean rate constants of ribozymes bearing, e.g., A and G residues was calculated for each permutation, revealing the underlying *t*-distribution and the critical values to which the true *t*-statistic was compared. Note that, whereas the canonical *t*-distribution (blue curve) has symmetric tails and thus symmetric critical values (blue vertical lines), the Monte Carlo simulation (gray bars) can reveal a *t*-distribution with markedly asymmetric tails and critical values (red lines). (*B*) Observed (*top*) and expected (*bottom*) nucleotide frequencies in the ligase selection. Red, G; blue, A; green, U; orange, C; white, gap. Less-saturated colors mark positions that were not deliberately varied in the pool. Above the colored bars are the results of Monte Carlo analyses of nucleotide identity effects on ribozyme kinetics at the indicated Mg^{2+} concentrations and of Fisher's exact test to detect significant deviation of observed from expected nucleotide frequencies. Open ovals indicate that a test was performed but revealed no significant effect; filled ovals indicate significant effects. The secondary-structure schematic below is colored as in Figure 1. (*C*) Histogram of the observed lengths of J1/3 sequences among successful ligase variants. J1/3 was varied from 2 to 10 nt in the starting pool; how some variants acquired longer J1/3 sequences is unknown.

significant conservation of the parental nucleotides at such sites; given expected frequencies of 32 parental and one each nonparental nucleotide at a given position, the probability of observing the parental nucleotide in all 35 isolates is 0.239. The two positions discussed above, nucleotides 76 and 103, account for nearly all of the substitutions and gaps observed among the 19 lightly mutagenized sites; setting these two positions aside and considering the other 17 positions in the aggregate, we observed only 10 nonparental features in all 35 isolates. We performed one million simulated selections at 19 lightly mutagenized sites, discarding from each simulation the two positions with the most nonparental features. On average, the 17 sites remaining in our simulations contained 53 ± 7 nonparental features, significantly more than the 10 observed in our in vitro isolates ($P < 10^{-6}$). Thus, the parental sequence appears to have been optimal at most of these 17 positions.

Connecting kinetic measurements to sequence variation

We next examined whether any mutations were specifically associated with rapid catalysis at 1 mM or 10 mM Mg^{2+}: at a given position, despite the changing background of variation at different positions, do ribozymes bearing an A (for example) have, on average, a higher mean rate constant than those bearing a G? Although a *t*-statistic could be calculated for any two such means, the group sizes and variances in our data set are frequently unequal; we could not safely assume that the underlying distribution of *t*-statistics would closely resemble the canonical *t*-distribution, and thus did not a priori know the critical values for our test statistic. We therefore took a Monte Carlo approach to determine the true distribution of *t*-statistics for our data (Fig. 3A; Materials and Methods). Starting with the measured rate constants and known group sizes (e.g., at position 64, 12 isolates carried a G, 12 carried an A, six carried a U, and five carried a C), we randomly reassigned rate constants to different groups. We could then calculate the mean rate constant in the "G" group and the mean in the "A" group, and from these mean values the *t*-statistic describing the difference between them. By repeating this process 10,000 times, we generated *t*-statistics with a distribution approaching that characteristic of our data. This distribution allowed us to determine the probability that the means of two random subsets of ligase rate constants would give a *t*-statistic as extreme as that obtained from the true groupings.

With this approach, we found several positions in the ligase at which nucleotide identity had a significant effect on rate. Two of these positions, 41 and 43, were in J3/4. Isolates bearing a G at position 41 had significantly faster rate constants at 1 mM Mg^{2+} than did isolates bearing an A (Bonferroni-adjusted $P < 0.018$), and at 10 mM Mg^{2+} than isolates bearing either an A or a U (Bonferroni-adjusted $P <$ 0.0066 and 0.041, respectively). Surprisingly, although we did observe significant selection at nucleotide 41, the favored residue was U (8.75 U41 clones and 8.75 G41 clones expected; 18 U41 and 11 G41 clones observed). Because the pool was not prefolded during selection, but was for rapid kinetic assays, one possible interpretation is that G41 might promote catalysis, but U41 might promote more optimal folding. At nucleotide 43, isolates bearing an A had significantly higher rate constants at 10 mM Mg^{2+} than did isolates with either a C (Bonferroni-adjusted $P <$ 0.002) or a G (Bonferroni-adjusted $P < 0.05$), but the effects were not significant at 1 mM Mg^{2+}. The crystal structure reveals that P3 is capped by a base pair between positions 40 and 44, with nucleotides 41–43 adopting a GNRA tetraloop-like configuration (Shechner et al. 2009); the base-specific effects at positions 41 and 43 likely reflect the energetics of this very short loop.

At two positions, 76 and 94, the Monte Carlo analysis detected a significant effect at 10 mM Mg^{2+} that grew stronger at 1 mM Mg^{2+}. At position 76, we had already seen that the parental identity was selected against strongly. Isolates bearing the parental adenosine had a mean ligation rate constant of 1.04 ± 0.55 min^{-1} at 10 mM Mg^{2+}, compared with 1.54 ± 0.64 min^{-1} among U76 clones ($P <$ 0.019); at 1 mM Mg^{2+}, the relative gap widened to 0.06 ± 0.04 min^{-1} versus 0.14 ± 0.12 min^{-1}, respectively ($P <$ 0.013). A kinetic difference that widens with decreasing metal concentrations in this way could reflect either of two situations: a pre-existing, but weak, metal-binding site might have its affinity improved by the favored substitution, allowing it to remain saturated at Mg^{2+} concentrations that would strip the site in the parent ligase; or the favored substitution might promote a non-metal-mediated interaction that compensates at low [Mg^{2+}] for a missing metal elsewhere in the ribozyme. Examination of the crystal structure of the improved ligase (Shechner et al. 2009) suggests that U76 may fall into the former class.

Several structural elements come together in the neighborhood of nucleotide 76. Moving up helix P6 from the 3′ end, the transition to the pseudoknot helix P3 is seamless; P3 simply continues the P7–P6 coaxial stack. But base pairing to form P3 occupies five of the six residues in what would otherwise be L6; the lone unpaired residue must negotiate the strand's return to P6 in register. That unpaired residue is nucleotide 76. In the improved ligase crystal structure, the U76 sugar and base are flipped out, allowing the base to stack favorably against G45. With the sugar flipped, the phosphate groups flanking U76 are brought closer together; the pro-R_p nonbridging phosphate oxygens at nucleotides 75 and 77 are separated by just 6.22 ± 0.19 Å (averaged across the two ribozyme chains in the crystallographic asymmetric unit), significantly smaller than the mean distance at other positions in the P3/P6 junction (9.84 ± 1.54 Å; $P < 5 \times 10^{-5}$, unpaired *t*-test assuming unequal variances). As discussed below, biochemical evidence suggests that the pro-R_p oxygens of positions

75 and 77, together with that of nucleotide 46, jointly ligand a structural Mg^{2+} ion. Notably, the U76/G45 stacking interaction and this proposed metal-binding site are the sole points of structural communication between the P7–P6–P3 domain of the ribozyme and the P4–P5 domain (Shechner et al. 2009). The flipped-out U76 base nearly fills the small cavity created by nucleotides 45–46; a larger sidechain at nucleotide 76 would likely force small realignments of the backbone, movement on a scale that could affect the affinity of the metal-binding site formed by nucleotides 75, 77, and 46. Thus, the parental adenosine could constrain the backbone at positions 75 and 77 to a conformation that more easily loses a bound Mg^{2+} ion, while mutation to U76 could improve the affinity of the backbone metal-binding site, such that low metal concentrations magnify the kinetic benefit of this mutation. Additional insights from our comparative sequence analysis are discussed in the supplemental materials.

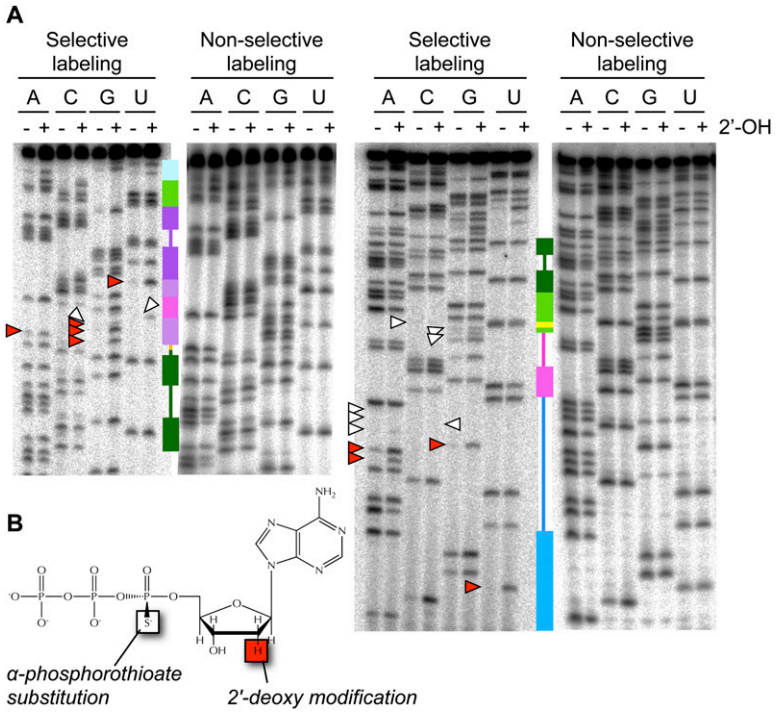


FIGURE 4. (*A*) Representative NAIM gels for quantification of phosphorothioate and 2′-deoxy effects in the improved ligase. Secondary-structure cartoons, colored as in Figure 1, provide landmarks. 6% gels were used to resolve the 3′ half of the ligase (*left*), and 15% gels were used to resolve the 5′ half of the ligase (*right*). White arrowheads mark positions of particularly strong phosphorothioate interference; red arrowheads mark positions of particularly strong 2′-deoxy interference. (*B*) [α-Phosphorothioate]-2′-deoxyadenosine triphosphate (dATPαS), one of the eight nucleotide analogs used for NAIM. Use of the α-phosphorothioate-bearing ribonucleotides ATPαS, CTPαS, GTPαS, and UTPαS permits quantification of phosphorothioate interference effects and establishes a baseline for comparison with the α-phosphorothioate-bearing deoxyribonucleotides dATPαS, dCTPαS, dGTPαS, and dUTPαS to determine 2′-deoxy interference effects (Conrad et al. 1995; Strobel and Shetty 1997; Ryder and Strobel 1999). The stereoisomer shown bears a pro-S_p sulfur substitution; this isomer is the only isomer recognized by T7 RNA polymerase, but because polymerization proceeds with inversion of stereochemistry, all sulfur substitutions in the resulting RNA are in the pro-R_p position (Verma and Eckstein 1998).

Interference mapping

Functional contacts in the class I ligase might have changed in either of two ways over the course of selection of the improved ligase. The change of a base might directly alter key contacts between that base and other base or backbone atoms; it might also have indirect effects if the altered base causes a shift in the positions of its backbone atoms and thus in their hydrogen-bonding or metal-binding potential. Two mapping techniques gave us access to the large majority of positions at which changes of either type might have occurred. First, nucleotide analog interference mapping (NAIM) (Conrad et al. 1995; Strobel and Shetty 1997; Ryder and Strobel 1999) of the parent and improved ligases allowed us to identify functionally significant nonbridging pro-R_p oxygens and 2′-hydroxyl groups (Fig. 4). Following this analysis of backbone contacts, interactions made by the bases themselves were investigated by interference mapping with dimethyl sulfate (DMS) (Fig. 5; Peattie and Gilbert 1980; Peattie and Herr 1981; Moazed et al. 1986; Stern et al. 1988). DMS methylates the Watson–Crick face of adenosine residues (at N1) and cytosine residues (at N3), the favored nucleotides at 15 of the 18 sites that showed significant bias in the ligase selection (Fig. 3), allowing us to investigate regions of the ribozyme where selective pressure for a given nucleotide may signal a tertiary interaction of note.

Independent interpretation of these results and of the crystal structure of the improved ligase reaction product (Shechner et al. 2009) led to a number of concordant conclusions; in other cases, patterns hidden in the interference data and the crystal structure considered separately came into focus when the observations were combined. (In some of these latter cases, the maps and the structure together predict interactions that will require additional mapping to confirm; these positions are discussed in the supplemental material.) Finally, at some positions the interference maps cannot be explained by the crystal structure, perhaps indicating contacts involved in ribozyme folding. We discuss the interference maps and their relation to the crystal structure below, as well as several telltale signs that some contacts particularly important to the polymerase are also central to ligase activity, supporting the notion that the ligase remains a good model for the polymerase core.

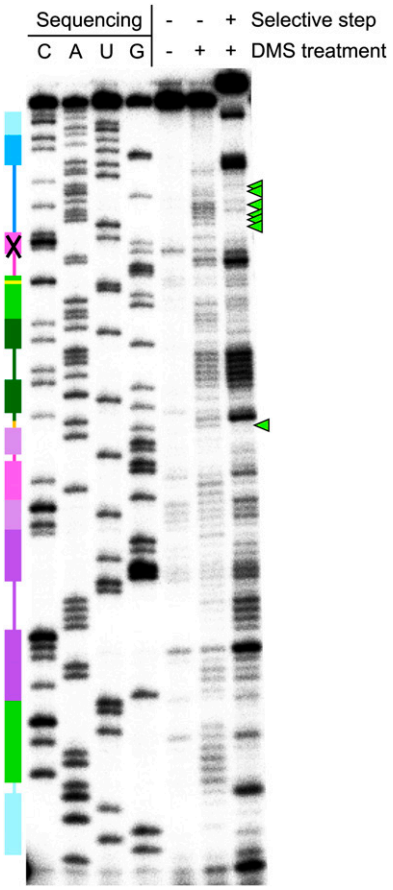


FIGURE 5. A representative primer-extension gel for quantification of DMS interference in the improved ligase, with sites of strong interference in single-stranded regions marked by green arrowheads. The secondary structure cartoon is colored as in Figure 1. Band identities were assigned using dideoxy sequencing ladders; due to extension pausing before modified nucleotides in experimental lanes, there is a one-base offset between experimental lanes and sequencing ladders. The gel compression that prevented analysis of nucleotides 34–41 is indicated by an X on the secondary-structure cartoon.

Pro-R_p phosphorothioate interference

On the whole, the phosphorothioate profiles of the parent and improved ligases were similar, although the magnitude of interference effects was typically greater in the parent ligase than in the improved ligase (Figs. 4, 6, 7). Each of the 20 sites of significant interference we detected has the potential to be an inner-sphere ligand of a functionally important Mg^{2+} ion. In principle, a metal rescue experiment could lend support to the role of some of these oxygens as inner-sphere ligands (Piccirilli et al. 1993). However, the ligase is strongly inhibited by the transition metals that are the best candidates for metal rescue (Glasner et al. 2002); pilot experiments confirmed that this line of inquiry was unlikely to be productive (data not shown). Instead, the improved ligase crystal structure (Shechner et al. 2009) provided a framework for interpreting the interference maps. Although some phosphorothioate interference effects (and most enhancement effects) remain unexplained, the crystal structure clearly suggests that nearly half of the phosphorothioate-interference effects we observed were related to metal-binding activity.

Candidate metal-binding site in the ribozyme core

The phosphorothioate interference effects for which the crystal structure offers a clear interpretation largely cluster in and around the ligase active site, proposed by Shechner et al. (2009) to center on phosphates from nucleotides 29 to 30 and the C47 nucleobase (Fig. 8A). As reported in that work, the strongest set of phosphorothioate interference effects lie at the 3′ end of J1/3, at positions 29–32. At each of these positions, the phosphorothioate substitution reduced ligase activity by a factor of ≥6, the limit of the assay's reliable range. Independent of the crystal structure, we hypothesized that this series of contiguous interference sites might indicate a bend in the backbone sharp enough that pro-R_p oxygens at adjacent residues would be positioned to ligate the same metal ion. This prediction was borne out by the backbone geometry at positions 30–32, where J1/3 abruptly makes a right-angled turn, positioning the pro-R_p oxygens of positions 31–32 to coordinate a magnesium ion (Shechner et al. 2009). That the refined $|F_{obs}|$-$|F_{calc}|$ difference map of the ligated product does not give evidence of a second metal bound at positions 29–30, despite the extremely strong interference effects observed, is consistent with a metal ion bound by these two phosphates that is important early in catalysis, but released later in the reaction trajectory (Shechner et al. 2009).

On the other side of the proposed active site, the J3/4 backbone closely abuts the backbone of the 5′ strand of helix P6, with phosphate–phosphate distances between adjacent strands of just ~7 Å. A chain of well-ordered metal ions is strung between the J3/4 and P6 backbones, providing the electrostatic screening necessary to permit such close strand–strand packing (Fig. 8A; Shechner et al. 2009). Loss of any of these metal-binding sites by phosphorothioate substitution should lead to repulsion between J3/4 and P6, presumably disrupting the positioning of the proposed catalytic nucleobase C47 in the active site. Indeed, as at nucleotides 29–32, we observed a ≥6-fold phosphorothioate interference effect at nucleotide 47; in the crystal structure, the pro-R_p oxygen of nucleotide 47 is well-positioned to be an inner-sphere ligand of one Mg^{2+} in the backbone-screening chain of ions. One position upstream, nucleotide 46 directs its pro-R_p oxygen toward the pro-R_p oxygens of P6 residues G75 and G77 (Fig. 8A); as discussed above, the extrusion of U76 and the close approach of J3/4 to P6 produce a spacing of ~6 Å between each pair of oxygens. This favorable geometry, observed in an early model of the crystal structure, was the first indication of a connection between the strong and, strikingly, nearly equal phosphorothioate interference effects at

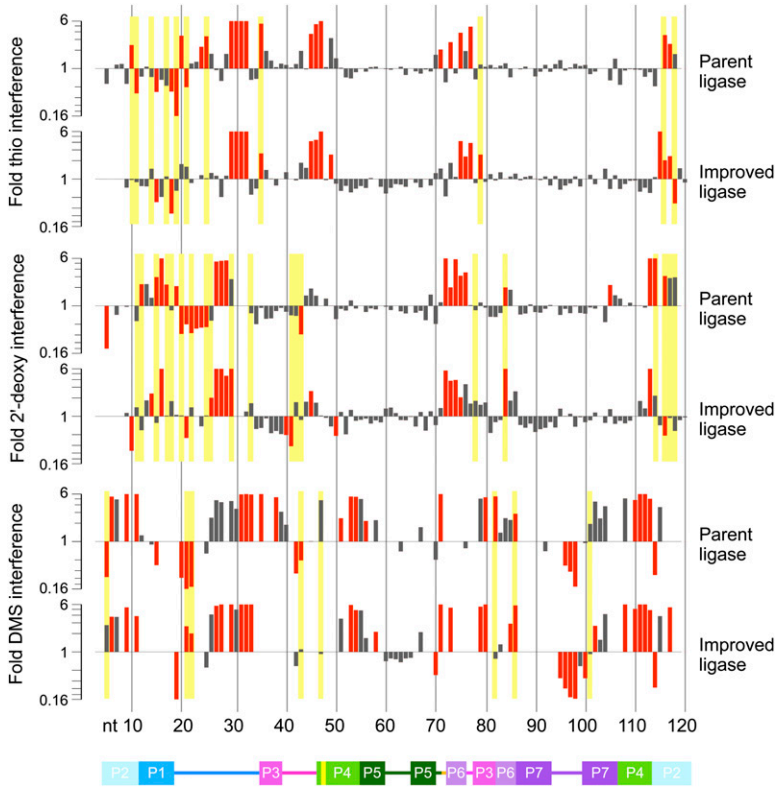


FIGURE 6. Log-scale maps of mean phosphorothioate (thio), 2′-deoxy, and DMS interference and enhancement effects in the parent and improved ligases. Positions of significant (95% confidence interval excludes 1.0) and strong (mean greater than or equal to twofold) effects are shown in red. Yellow bars highlight positions at which the mean fold effect differs both significantly ($P < 0.05$, *t*-test) and substantially (by a factor of ≥2) between the two ribozymes. Phosphorothioate interferences obscure possible 2′-deoxy effects at several positions in both ligases. DMS mapping by primer extension yields data only for A and C residues. Ribozyme positions are numbered below, with the secondary structure cartoon colored as in Figure 1.

positions 46, 75, and 77. Subsequent refinement of the crystal structure confirmed the presence of a strong (4–5σ) peak in the $|F_{obs}| - |F_{calc}|$ difference map in between the pro-R_p oxygens of positions 46, 75, and 77, the third apparent metal ion in the chain that links J3/4 with P6.

Hints of evolving interactions in J1/3

As discussed below, several lines of evidence indicate that some key structural difference in the 5′ end of J1/3 separates the improved ligase from the parent. A clear instance of evolving interference effects arose at position 20. Here, the parent ligase showed 3.5-fold phosphorothioate interference, twice as strong as the significant but mild 1.8-fold effect in the improved ligase. But the crystal structure gives no indication of a metal ion; instead, the pro-R_p oxygen is positioned well to serve as a hydrogen-bond acceptor for the 2′-hydroxyl group of position 19. Such an interaction could be disrupted by the electrostatic perturbation of phosphorothioate substitution. Looking to the 2′-deoxy interference map, we find that removing the 2′-hydroxyl group of position 19 produced twice as strong an interference in the parent (2.1-fold) as in the improved ligase (1.1-fold, not significant). The concordant relative impacts of the phosphorothioate effect at position 20 and the 2′-deoxy substitution at position 19 in the two ligases strongly suggest that both effects arose from disrupting the hydrogen bond observed in the crystal structure, and further, that the improved ligase tolerates the loss of this bond more readily than does the parent. The observation in both ribozymes that disrupting the pro-R_p oxygen partner of this hydrogen bond appears to come at a greater cost than disrupting the 2′-hydroxyl partner might be due to additional steric penalties incurred by the phosphorothioate substitution.

Lingering questions in the phosophorothioate map

At most of the remaining sites of significant phosphorothioate effects, the crystal structure offers no suggestion of the source of the effect. At two positions, moreover, the structure suggests an interpretation that is belied by other aspects of the interference maps. First, one might connect the parental phosphorothioate interference at position 71 to an interaction with the 2′-hydroxyl at position 70, but the geometry is suboptimal, and a 2′-deoxy substitution at the latter gives significant enhancement of ligase activity. Second, the structure suggests that the approximate twofold phosphorothioate interference at position 73 could reflect an interaction with the pro-R_p oxygen of nucleotide 107, when, in fact, the latter shows mild but significant phosphorothioate enhancement in both ligases. Such instances of seeming contradiction between crystallographic and biochemical data might mark steric or electrostatic side effects of the phosphorothioate substitution. Alternatively, they could indicate changes that take place along the reaction coordinate, with the interference maps reflecting the early part of the reaction, from folding through the transition state, and the crystal structure reflecting the product.

A mixture of mutability and evolutionary stability in the 2′-deoxy interference map

Although the 2′-deoxy interference maps of the parent and improved ligases generally resembled each other, at a few

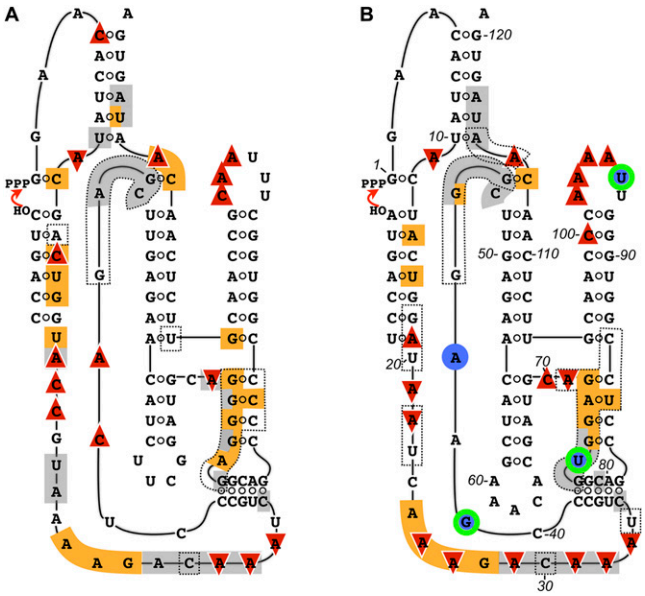


FIGURE 7. Secondary-structure projections of the (*A*) parent and (*B*) improved class I ligase ribozymes, superimposed with the results of structural and sequence mapping. Dotted lines in *A* indicate the parent residues at which the C4′ atom was protected from Fe-EDTA-generated hydroxyl radicals (Bergman et al. 2004). Dotted lines in *B* indicate sites at which the C4′ atom was calculated to be solvent inaccessible in the crystal structure (Shechner et al. 2009). Phosphorothioate interference is shown in gray; 2′-deoxy interference is shown in orange; DMS interference is shown by downward-pointing red triangles, and DMS enhancement by upward triangles. In *B*, positions at which the Monte Carlo analysis revealed a significant association between nucleotide identity and activity are highlighted by blue circles if the effect appeared at 10 mM Mg^{2+} and by green and blue rings if the effect appeared at both 10 mM and 1 mM Mg^{2+}.

positions the mean interference effects in the parent and improved ligases were not only significantly different according to a *t*-test, but also differed by at least a factor of two (Figs. 6, 9), sometimes with stronger interference in the parent (e.g., positions 12, 15, and 17) and other times in the improved ligase (e.g., positions 29 and 84). Such changes are not surprising at positions with weak or moderate interference effects; if a mildly favorable 2′-OH interaction is disrupted during selection, a return to the energetic starting point could be as easy as acquisition of a single hydrogen bond elsewhere in the ribozyme. However, the pattern at positions 29 and 114 reveals that even some of the backbone interactions on which ligase activity relies most heavily were in flux during selection. Even more strikingly, the 2′-deoxy interference maps at position 116 showed a more than sixfold change, from more than threefold interference in the parent to more than twofold enhancement in the improved ligase; it is as though in our seven rounds of selection we had taken the ligase apart, put it back together, and then discovered that, for once, leaving out the proverbial spare parts made the system run more smoothly.

Among the 2′-hydroxyl-interference effects consistent across the parent ligase and improved ligase, the strongest effect was at position 16 (Figs. 4, 6), in the 5′ strand of helix P1. This strand is the counterpart to the template oligonucleotide added in *trans* to the polymerase reaction; in both the ligase and the polymerase, this strand pairs both with the oligonucleotide bearing the nucleophilic 3′-OH and with the 5′-triphosphorylated nucleotide or residue, thereby aligning the reactive groups. In the polymerase ribozyme, a 2′-deoxynucleotide at template position –3, the polymerase analog of ligase position 16, is more detrimental than is 2′-deoxy substitution at any other template nucleotide (Müller and Bartel 2003). The robustness of this effect across the ligase family of ribozymes strongly suggests that the phosphoryl transfer reactions performed by all of these ribozymes require this hydroxyl group to form a specific contact with the ribozyme core, a contact likely to contribute substantially to proper alignment of the substrate helix in the active site. The crystal structure suggests that the nucleotide 16 2′-hydroxyl group may be involved in two hydrogen bonds, one to the 4′-oxygen of position 17 and one to the imine nitrogen (N1) of A25. The strong selection for an adenosine at position 25 suggested that the latter base-specific contact could make an important contact, but the DMS interference map does not strongly support this proposition, showing moderate and nonsignificant interference effects at A25 in both the parent and the improved ligase. This potential disparity between the interference maps is unusual; when considered side by side with the crystal structure and the DMS interference map, the 2′-deoxy interference map fully supported the identification and functional significance of a suite of other tertiary interactions that join and brace key elements of secondary structure and explain the longstanding observation of a strong preference for adenosine residues in the 3′ region of J1/3. We discuss these interactions in the context of the DMS map in the sections that follow.

Base-pairing interactions in the dimethyl sulfate interference map

The primer extension assay (Fig. 5) gave rise to a somewhat noisier map than did the NAIM assay, including a number of positions, noted with an asterisk (*) below, at which the two ribozymes had comparably high mean interference factors, but only one achieved significance (Fig. 6). Sequence differences and gel artifacts also precluded comparison between the ribozymes at several positions (see Materials and Methods).

A number of strong interference effects in the DMS map are readily explained as the result of disrupting Watson–Crick base pairing within helices; similarly, DMS interference at C113 supports the predicted participation of this nucleotide in a base pair flanking the catalytic nucleotide C47 (Bergman et al. 2004). However, not all

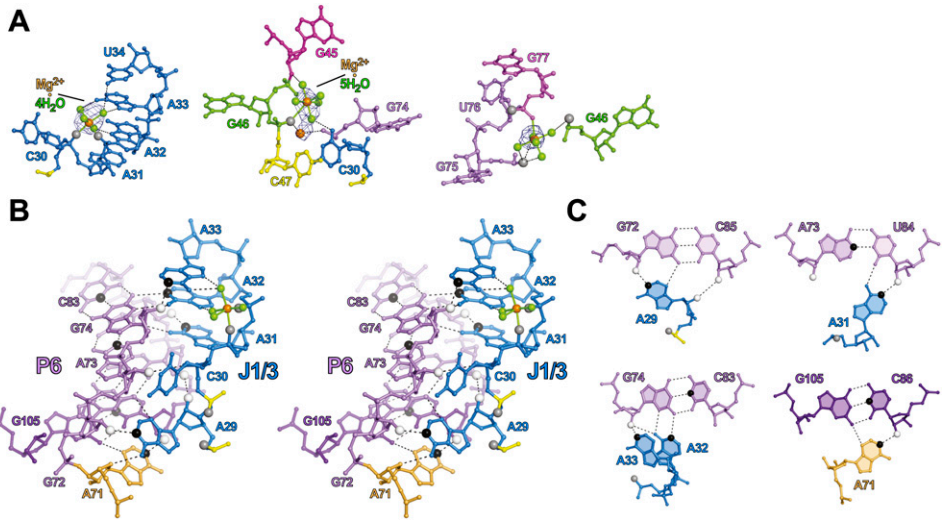


FIGURE 8. Structural interactions suggested by the crystal structure (Shechner et al. 2009) and confirmed as functionally significant by NAIM and DMS interference experiments. (*A*) Contacts with structural metals near (*left*) A31–A32, (*middle*) G45–C47, and (*right*) G75–G77. Pro-R_p phosphate oxygens for which phosphorothioate NAIM interference was significant are shown as larger gray spheres; pro-S_p oxygens are not accessible to NAIM (Verma and Eckstein 1998). Meshes represent simulated-annealing $|F_{obs}|$-$|F_{calc}|$ OMIT maps in which the hydrated metal clusters were excluded from map calculations, contoured at (*left*) 4σ, and (*middle, right*) 2.5σ. Black dashed lines indicate hydrogen bonds. Solid lines bound to Mg^{2+} ions indicate proposed inner-sphere contacts. (*B*) Wall-eyed stereograph highlighting interactions between J1/3 and the P3–P6–P7 domain. Functional groups showing significant interference in biochemical experiments are shown as larger spheres, colored white for 2′-deoxy interference, gray for phosphorothioate interference, and black for DMS interference. (*C*) Individual base triples or quadruples involved in this interaction, rendered as in *B*.

helical regions are equally important for ligase activity; it is known, for instance, that L5 and parts of P5 can be manipulated with little effect on activity, so it is reasonable that the DMS interference effect at A56 in this stem should be relatively weak (just 2.2-fold interference in the parent ligase, and no significant effect in the improved ligase). At the other end of the spectrum, the DMS interference map independently predicted that the crystal structure would reveal an error in our previous secondary-structure model of the ligase. That model (Ekland and Bartel 1995) predicted a noncanonical base pair between A11 and A114. A *trans* Watson–Crick/Watson–Crick base pair (Leontis and Westhof 2001) would almost certainly give rise to strong DMS interference at both positions, and a *trans* Hoogsteen/Hoogsteen or *trans* Watson–Crick/Hoogsteen base pair might do so as well, depending on the modulation of N6 hydrogen bonding by the N1 modification. We observe very strong interference at A11 in both ribozymes, and equally strong enhancement at A114, inconsistent with the formation of this base pair. The crystal structure bears this argument out: A11 forms a *cis* Hoogsteen/Sugar Edge pair with G2, while A114 is splayed out to stack between A4 and G46. Methylation of A11 N1 would certainly interfere with the A11:G2 pair, and the A114 enhancement effect may indicate that, in the context of the A114 stacking interaction, the additional surface area for van der Waals interactions that is introduced by N1 methylation is beneficial.

Extensive concordance between dimethyl sulfate and 2′-deoxy interference mapping

The DMS interference effects that arose in single-stranded regions were of particular interest, especially those in J1/3, where the DMS interference pattern echoed both the strong 2′-deoxy interference effects at positions 26–29 and the strong phosphorothioate interference effects at positions 29–32 (Figs. 6, 7, 9). The crystal structure suggests that the DMS interference effects at positions *26–*27 arose from interactions with the substrate oligo in helix P1. The N1 groups of these two nucleotides converge on substrate residue –3, with hydrogen bonds between A26 N1 and G–3 N2 and between A27 N1 and G–3 2′-OH. Although our 2′-deoxy interference maps do not encompass the ligase substrate, the proposal that a 2′-OH in this position contributes to ribozyme function draws support from the strong 2′-deoxy interference at the analogous position of the polymerase ribozyme primer strand (Müller and Bartel 2003).

Very strong interference was also observed at nucleotides *29 and 31–33 (Figs. 6, 8B,C). The crystal structure suggests a stack of A-minor interactions (Nissen et al. 2001), with each of these J1/3 adenosine residues packing into the minor groove of P6 in a slightly different geometry (Shechner et al. 2009). A29 docks into the G72:C85 pair at the base of P6, with a hydrogen bond forming between the A29 N1 and the G72 2′-OH, on the 5′ strand of P6; as expected, we see very strong DMS interference at A29 and

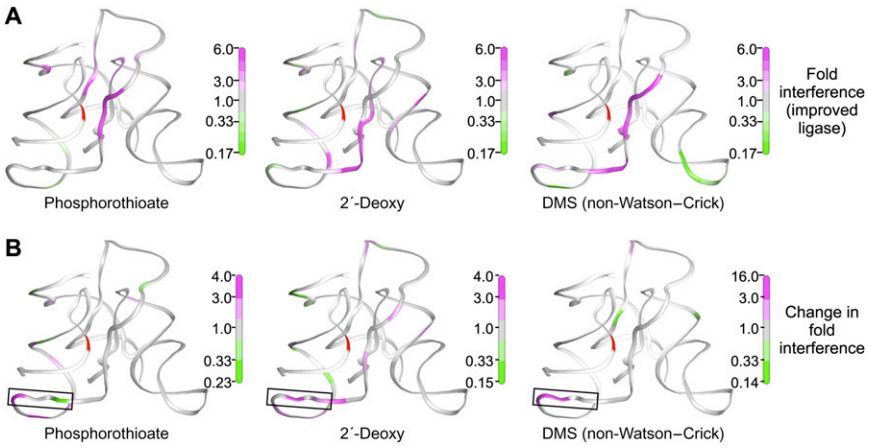


FIGURE 9. Results of interference mapping projected onto the improved class I ligase crystal structure (Shechner et al. 2009). (*A*) Projections of (*left*) phosphorothioate, (*middle*) 2′-deoxyribonucleotide, and (*right*) dimethyl sulfate interference mapping results in the improved ligase onto the view of the ligase shown in Figure 1A. The ligation junction is highlighted in red. Residue colors are scaled from green (strong enhancement) to magenta (strong interference). Positions at which interference could not be quantified are colored white. For clarity, positions at which DMS interference arises from Watson–Crick pairing in known helices are not colored. (*B*) Projections of the extent of evolutionary change seen in the (*left*) phosphorothioate, (*middle*) 2′-deoxyribonucleotide, and (*right*) dimethyl sulfate interference maps of the parent and improved ligase ribozymes. The mean effect in the improved ligase was divided by the mean effect in the parent ribozyme and the results scaled from green (stronger enhancement or weaker interference—i.e., less reliance on the unperturbed residue—in the improved ligase) to magenta (weaker enhancement or stronger interference in the improved ligase). The cluster of changes in the boxed region at the 5′ end of J1/3 is discussed in the text.

equally strong 2′-deoxy interference at G72. C30 is flipped out of the stack of J1/3–P6 interactions to form its own stacking interaction with the proposed catalytic nucleobase C47, and J1/3-P6 docking resumes with A31, the first position of the 3′ "A-minor triad" motif (Shechner et al. 2009). This residue docks into the A73:U84 pair, this time with a hydrogen bond forming between the A31 N1 and the position 84 2′-OH, on the 3′ strand of P6; again we observe very strong DMS interference and, in the improved ligase, equally strong 2′-deoxy interference. A32 forms an N3-amino, amino-N1 A:G pair with G74 (which also forms a Watson–Crick pair in P6 with C83), forming hydrogen bonds between A32 N1 and G74 N2 and between A32 N6 and G74 N2, N3, and possibly 2′-OH. And A33 caps this stack of J1/3–P6 interactions with a hydrogen bond from its N1 to the 2′-OH of G74; disrupting either end of this bond produces very strong interference in both ribozymes. A33 may further contribute to ligase activity by forming hydrogen bonds between its N6 and as many as three of the ordered water molecules that coordinate the metal ion bound by the pro-R_p oxygens of nucleotides 31–32. The near-perfect agreement of the interference maps with each other and the crystal structure in this crucial region of the ligase serves to confirm that, whatever structural changes the ligase may undergo in the wake of reaction, the state of P6 and J1/3 captured by the crystal still bears a strong resemblance to the active conformation.

These J1/3 and P6 results also agreed well with the hydroxyl radical footprinting data previously reported for the parent ligase (Bergman et al. 2004). Those experiments revealed robust solvent protection of the ligase backbone on both strands of P6, as expected if the 5′ strand is buried against other helices in the catalytic core and the 3′ strand is protected by J1/3. Notably, the solvent protection calculated from the improved ligase crystal structure corresponds with that measured in the parent ligase at nearly all positions (Shechner et al. 2009), with a handful of intriguing differences (Fig. 7). One of these exceptions is nucleotide 84. In the parent, measured C4′ solvent accessibility at this position was well below average (Bergman et al. 2004); in the improved ligase, calculated C4′ solvent accessibility is above average (Shechner et al. 2009). In the parent, 2′-deoxy interference here was a meager twofold effect; in the improved ligase, it was greater than or equal to sixfold (Fig. 6). As noted above, this base pair was changed from the parental G:C to an A:U during pool construction because an A:U pair slightly improved function (EH Ekland and DP Bartel, unpubl.); perhaps that improvement is tied to a slight realignment of the helix backbone, permitting the improved ligase to further stabilize contact(s) that already play a significant role in the parent.

The idea that a slight but important realignment has taken place here was echoed in the differences in the relative importance of the other 2′-OH groups in the P6 helices of the two ribozymes. For instance, the improved ligase showed nearly twofold higher 2′-deoxy interference than did the parent ligase at A73, the complement to U84, whereas contacts made by the 2′-hydroxyl groups of G74 and G75 appeared less important in the improved ligase than in the parent. Continuing this trend, DMS-based disruptions of helix P6 and the proximal base pair of P7 did not affect the two ribozymes in quite the same way. At *C85, DMS modification gave only a threefold interference effect in the improved ligase, despite presumably disrupting the Watson–Crick base pair that caps P6. At C82, DMS modification had the expected very strong (5.5-fold) effect in the parent ligase, while tending, if anything, toward enhancement in the improved ligase. The tables were turned at C86, the proximal base pair of P7, where the 5.8-fold effect in the improved ligase was just over twice the interference seen in the parent. If indeed both ribozymes do make dense but distinct networks of tertiary contacts in this

region of the ribozyme, it is possible that these networks provide an energetic buffer against disruption of the helices, but that the different networks can best accommodate disruption at different positions.

A hotspot of ribozyme evolution?

Selection favored nonparental residues at positions 19 and 21–23, and that change in sequence entailed a change in backbone conformation: C4′ atoms at positions 19–20 and 22–23 are slightly more exposed than average in the parent and far more protected than average in the improved ligase (Shechner et al. 2009). Moreover, the differences between the NAIM and DMS maps of the two constructs cluster in this region (Figs. 6, 7, 9). In particular, positions 21–22, at which our selection favored substitution of adenosines for the parental cytosines, show DMS enhancement in the parent, but interference in the improved ligase (Figs. 6, 7, 9). Whatever the contacts made possible by acquisition of adenosines in this region, the benefits are not confined to the ligase ribozyme. A recent continuation of polymerase ribozyme evolution produced a variant with improved primer extension, as well as improved fidelity (Zaher and Unrau 2007). The key difference—indeed, almost the only difference—between this variant and the original polymerase (Johnston et al. 2001) is the acquisition of three additional adenosine residues at positions analogous to ligase nucleotides 20, 21, and 23. Thus, multiple lines of evidence indicate that the contacts made by these few residues, continuing to improve through evolution, have a substantial impact on ribozyme function despite their distance from the active site.

Conclusions

Continued selection on the class I ligase ribozyme demonstrated that even this efficient catalytic machine had room for optimization, particularly with regard to metal dependence. Statistical analysis of the population of ribozymes that survived selection allowed us to identify residues under selection and candidate residues at which nucleotide identity may buffer the ribozyme against changing Mg^{2+} concentrations. Targeted structural probing allowed us to interrogate defined subsets of these possible interactions, often revealing congruent results when we separately modified each partner of a proposed interaction. In general, the probe results we report here are in good agreement not only with each other but also with the recently solved crystal structure of the improved class I ligase ribozyme. Together, these methods offered evidence for catalytic metals bound at the ligase active site (Shechner et al. 2009) and revealed a diversity of adenosine-mediated interactions that belies the sequence-level homogeneity of the long single-stranded region J1/3. Our improved understanding of ligase structure and function, and our insights into the regions of the ligase ribozyme that have responded to selective pressure, will inform future efforts to convert the RNA polymerase ribozyme into a general RNA replicase, the proof of principle the RNA world hypothesis requires.

MATERIALS AND METHODS

Pool construction and selection

The DNA encoding a partially randomized Class I ligase pool was synthesized as a single, variable length oligonucleotide [5′-acgact cactataggAAcactatactactgg(N_{2-10})-ACAAATctgcc(N_{1-4})gagcttgag aacatcg(N_{4-6})cgatg(N_{1-2})gaggAggcagcctccgGtgg(N_{6-8})-ccaAcgtt ctcaaCaatagtgATTaatattcctgttgc-3′]. Positions in lowercase were constant, and positions in uppercase were mutagenized at 10% (Fig. 1). Thus, at these positions 90% of the molecules had the parental sequence, and the remaining 10% had one of the other three nucleotides. Positions noted as (N_{x-y}) were completely randomized, and varied in length within the range noted. Randomized residues were synthesized using phosphoramidite mixtures that compensated for the different efficiencies with which each monomer adds to a growing DNA chain (relative efficiency 0.70:0.73:0.85:1.0, dA:dC:dG:dT). Length variation was introduced at the points indicated using a split-and-pool protocol (Giver et al. 1993). This procedure resulted in a template pool, varying from 134 to 150 nt, made up of a randomized ligase domain ranging from 107 to 123 nt that was flanked by constant primer-binding sites on the 5′- and 3′-ends. The synthetic DNA was purified by denaturing gel electrophoresis and used as a template in a large-scale PCR reaction (150 mL). Six cycles of PCR were done, wherein each cycle was 4 min at 96°C, 5 min at 42°C, and 7 min at 72°C. Primers for PCR were 5′-AAAGCAACAG GAATATT-3′ and 5′-TTCTAATACGACTCACTATAGG-3′. Sequencing of clones from this PCR confirmed that randomization and length variation in the pool were as designed. Amplified DNA was phenol extracted, precipitated, and used as a template for an in vitro transcription reaction that generated pool RNAs.

The first round of selection was performed as diagrammed in Supplemental Figure S1. Pool RNA (580 μg, 1 μM final concentration) was heated in H_2O (80°C, 10 min) with an RT primer (5′-AAAGCAACAGGAATATT-3′) that hybridized to the 3′ terminus of the pool. After the RNA was cooled slowly to room temperature, biotinylated substrate (Dharmacon Research, 5′-biotin-taatacgactCCAGUA-3′, DNA bases lowercase) was added to a final concentration of 1.2 μM along with buffer and salts (final concentrations, 10 mM $MgCl_2$, 200 mM KCl, 50 mM EPPS at pH 8.0). The pool and substrate were incubated at 22°C for 5 min, and the reaction was stopped by addition of EDTA to 15 mM. The RNA was then precipitated and purified in a 10% polyacrylamide gel using radiolabeled RNA markers that were 138- and 160-nt long (ligated pool ranged from 141 to 157 nt). After elution and precipitation, ligated product was heated (65°C, 10 min) and cooled (22°C, 10 min), then incubated with streptavidin paramagnetic beads (Promega) for 10 min according to the manufacturer's recommendations. After washes with 0.1X SSC (three times), water (once), and RT buffer (once), additional RT primer was added, and the remaining RNA was reverse transcribed using an RNase H-deficient reverse transcriptase

(Superscript II, Gibco-BRL) while rotating in an incubator kept at 48°C. A control reaction was run in parallel without reverse transcriptase to detect any contaminating DNA in subsequent PCR steps. After 1 h, EDTA was added to the RT mixture (final concentration, 3 mM), and the beads were washed three times with 0.1X SSC. Following the final wash, both the RT primer and a 5′-PCR primer (5′-TTCTAATACGACTCCAGTAGG-3′) were added to 10 μM final concentration, and cDNA was removed from beads by alkaline denaturation (100 mM KOH, 20 mM Tris-Base). After removal from the beads, the cDNA solution was heated at 90°C for 10 min to hydrolyze any remaining RNA. It was then adjusted to pH 8.5 with HCl and used as a template for PCR.

Amplification of cDNA was performed using the RT primer and 5′-PCR primer described above, using a standard hot start protocol. After amplification, pool DNA was purified by Qiaquick column (Qiagen) and used as a template in a second PCR in which the 5′ primer was changed to 5′-TTCTAATACGACT CACTGTAGG-3′. The resulting DNA was again purified by Qiaquick column and used as a template in a final PCR reaction using a 5′ primer containing the T7 RNA polymerase promoter sequence (5′-TTCTAATACGACTCACTATAGG-3′). Amplified DNA was extracted with phenol and chloroform, then precipitated and used as template for an in vitro transcription, thereby generating the pool RNA for the next round.

Further rounds of selection were performed as in round 1, with the following changes: Beginning in round 2, RNA was treated with DNase (DNase RQ1, Promega) to eliminate template DNA prior to the selective ligation step. Also, in rounds 2–4 the gel purification following the ligation step was omitted in favor of a simple precipitation. In rounds 5–7, gel purification was again used to eliminate molecules that might have remained on the streptavidin beads because they were able to bind tightly to substrate. Furthermore, in rounds 5–7 RT preceded streptavidin-biotin affinity purification so that the RNAs were hybridized with their cDNAs and thus unable to form structures that might have high affinity for the beads. The stringency of the ligation step was increased in each successive round by decreasing pH or reaction time (round 1, 5 min at pH 8.0; round 2, 20 sec at pH 7.0; round 3, 1 sec at pH 7.0; round 4, 2 sec at pH 6.0; rounds 5–7, 0.2 sec at pH 6.0). Buffers were at a final concentration of 50 mM, and were EPPS (pH 8.0), BES (pH 7.0), or MES (pH 6.0). For ligation steps requiring incubations of ≤2 sec, a KinTek RQF-3 rapid-quench flow apparatus (KinTek Corp.) was used. Finally, we observed that after round 6 the pool activity was identical with and without added RT primer, and thereafter it was omitted from the ligation step. After seven rounds the pool cDNA was cloned (TOPO-XL kit, Novagen), and individual isolates were sequenced.

Ribozyme and substrate RNAs

The parent Class I ligase ribozyme (GenBank no. U26413) was transcribed in vitro from a plasmid template linearized with EarI (Bergman et al. 2000). Transcripts were purified on 6% polyacrylamide/8 M urea gels and stored in water at –20°C. Individual clone RNAs were transcribed from PCR templates and purified in the same way. PCR templates were generated by amplifying specific clone cDNAs using a 5′-primer containing a T7 RNA polymerase promoter and either the RT primer used throughout selection or a primer (5′-AAAGCAACAGGAAATATT-3′) that truncated ribozymes to the same length as that of the parent ligase. Ribozyme constructs in which individual changes were made to a specific clone were prepared in the same way using PCR primers to generate transcription templates with the desired changes. All substrates used in this study (5′-aaaCCAGUC-3′, parent ligase; 5′-taatacgactCCAGUA-3′, selection; and 5′-UCCA GUA-3′, ribozyme clones; DNA bases lowercase) were purchased (Dharmacon Research) and purified by anion-exchange chromatography (Nucleopac 9 × 250 column, Dionex). Substrates were radiolabeled using T4 polynucleotide kinase and [γ^{32}P]-ATP.

Kinetic analyses

In manual ribozyme assays, the ribozyme (1 μM final concentration) was heated (80°C, 2 min, in H_2O) and then cooled (22°C, 2 min), and the reaction was initiated by simultaneous addition of 50 mM buffer, the indicated concentration of Mg^{2+}, and trace ^{32}P-labeled substrate. Buffers were the same as described for selection. KCl (200 mM) was included in initial experiments, but omitted in experiments measuring the Mg^{2+}-dependence of ligation, and in all reactions in which the concentration of Mg^{2+} was <10 mM, because at low Mg^{2+} concentration it inhibits the parent ligase (Glasner et al. 2002). Omitting KCl did not affect rates of either parent or clones at ≥10 mM Mg^{2+}. Aliquots were taken at appropriate time points and were added to 2 vol of stop solution containing 120 mM EDTA and 8 M urea. Product and substrate were separated in 20% polyacrylamide gels and quantified by phosphorimaging. Ligation rates were calculated as described previously (Bergman et al. 2000).

Statistical analysis of selection results

Fisher's exact test was used to detect positions at which the nucleotide composition of the selected pool deviated significantly from that expected by chance. Analysis was performed in R version 2.7.1 using a 2 × 4 contingency table in the function fisher.test in the package stats (R Development Core Team 2008). For pooled analysis of lightly mutagenized sites, 19 sets of 35 random numbers uniformly distributed between 0 and 1 were generated using the Mersenne-Twister random-number generator as implemented in R; those numbers <0.1 were considered "nonparental." Simulated nonparental features were counted in each of the 19 sets, and the two sets with the most nonparental features were discarded. The remaining 17 sets' nonparental features were summed to give an aggregate count of in silico "observed" nonparental features. This simulation was repeated 10^6 times, generating a normally distributed set of in silico observations. The in vitro observation (10 nonparental features) was far smaller than the smallest of our 10^6 in silico observations (24 nonparental features, from a distribution with mean 53 ± 7), allowing us to conclude that the in vitro observation would arise by chance with $P < 10^{-6}$.

For Monte Carlo analysis, each ligase position x under scrutiny gave rise to a vector of length 35, whose ith element was the nucleotide identity of successful clone i at position x. Some ordered subset {A} of these elements will be adenosines, some subset {G} will be guanosines, some subset {C} cytosines, some subset {U} uracils, and some subset {E} gaps. The rate constants measured for the successful clones at each Mg^{2+} concentration were treated as another vector of length 35. R was used to re-order the components of the rate constant vector randomly 10,000 times, producing a matrix of 35 rows by 10,000 columns. Simulated mean rate

constants for ribozymes bearing A, G, C, U, or a gap at nucleotide x were computed for each iteration of the Monte Carlo by averaging the elements in, respectively, the set of positions {A}, {G}, {C}, {U}, or {E} of column j of the matrix. Simulated standard deviations were computed in similar fashion. The t-statistic for each accessible pairwise comparison of means was then calculated in each column, where "accessible" comparisons were those for which both nucleotide identities had at least three representatives. Thus, if all five nucleotide identities were represented in at least three clones each at position x, all 10 pairwise combinations—A/C, A/G, A/U, A/gap, C/G, C/U, C/gap, G/U, G/gap, and U/gap—were performed; if four identities were represented, six pairwise comparisons were performed; if three, three; if two, one; and if only one, the position was entirely inaccessible to this analysis.

At each accessible position x in the ribozyme, this procedure yielded a histogram of 10,000 t-statistics for each accessible pairwise nucleotide comparison. The true t-statistic was then calculated from the original vector of rate constants for each accessible comparison and compared with the appropriate histogram. Simulated t-statistics more extreme than the true t-statistic were counted, and the reported P value was determined as

$$P \leq \left(\frac{m}{10,000}\right)2B,$$

where m is the count of more extreme simulated t-statistics, the factor of 2 is a correction to give the two-tailed probability, and B is the appropriate Bonferroni penalty factor (10, 6, 3, or 1) for the number of comparisons performed.

Nucleotide analog interference mapping

To incorporate nucleotide analogs into the ligase RNAs used for NAIM, linearized DNA encoding ribozyme RNAs at a final concentration of 0.5 mg/mL was mixed with 1 mM ATP, 1 mM GTP, 1 mM CTP, 1 mM UTP; the α-phosphorothioate nucleotide analog of interest (Glen Research) at the concentration recommended by the supplier; and 1X T7 Y639F buffer (40 mM Tris-HCl at pH 8, 4 mM spermidine, 10 mM DTT, 15 mM $MgCl_2$, 0.05% Triton X-100). Transcriptions of the parent ligase also included 2.5 μM T7 promoter oligonucleotide. The reaction was mixed by vortexing prior to addition of the Y639F mutant of T7 RNA polymerase (Sousa and Padilla 1995; Ortoleva-Donnelly et al. 1998) and incubated for 60–75 min at 37°C, when a cloudy precipitate had appeared in all reactions, then gel purified.

Levels of nucleotide analog incorporation at each position were measured by nonselectively labeling the transcribed RNAs and subjecting them to I_2 cleavage. Briefly, 20 pmol of each transcript was treated with alkaline phosphatase (Roche) according to the manufacturer's instructions, purified by phenol-chloroform extraction and ethanol precipitation, labeled with [γ^{32}P]-ATP and polynucleotide kinase, and then gel purified. Following ethanol precipitation, pellets were resuspended in a 1:2 mixture of water and 2X denaturing gel loading buffer (8 M urea, 25 mM EDTA). Solutions were split into two aliquots; to the first was added 0.1 vol of freshly prepared 100 mM I_2 in EtOH, and to the second, 0.1 vol EtOH. These solutions were heated to 50°C for 10 min and loaded onto two sequencing gels, a 15% and a 6% polyacrylamide TBE-urea gel, to permit resolution of the ribozyme 5′ and 3′ regions, respectively. Banding patterns were visualized by phosphorimaging (Fujifilm BAS-2500) and quantified with the Fujifilm program ImageGauge. Two calibration replicates were performed for each nucleotide-analog transcription.

For interference reactions, nucleotide analog-bearing ribozyme transcripts were mixed in a 2:1 molar ratio with [γ^{32}P]-labeled substrate (5′-dAdAdACCAGUC-3′, parent-ligase substrate; 5′-UCC AGUA-3′, improved-ligase substrate), heated to 80°C for 2 min and then cooled to 22°C for 5 min. Reactions were initiated by addition of stringent ligase buffer (final concentration 50 mM MES at pH 6, 1 mM $MgCl_2$). Parent ligase reactions were incubated at 22°C for 100 min; improved ligase reactions, for 1 min. Reactions were stopped by the addition of 2 vol of 2X denaturing gel-loading buffer. Reactions were then split into two aliquots for treatment with freshly prepared dilute I_2 or mock-treatment with EtOH and subjected to gel electrophoresis and quantification as described above. The results reported are the average of three replicates of this interference experiment for each nucleotide analog.

2′-deoxy interference effects were calculated from background-subtracted peak intensities as follows. First, for each experimental replicate, the raw interference effect at each modified residue was calculated as

$$I_{\text{raw}} = \frac{\left(\frac{N}{dN}\right)_{\text{selective}}}{\left(\frac{\sum N}{\sum dN}\right)_{\text{selective}}},$$

where N is the background-subtracted peak intensity for the α-phosphorothioate ribonucleotide analog, dN is the background-subtracted peak intensity for the α-phosphorothioate deoxyribonucleotide analog, and the denominator is a gel loading and exposure normalization factor (total background-subtracted intensity of the α-phosphorothioate ribonucleotide analog lane divided by total background-subtracted intensity of the α-phosphorothioate deoxyribonucleotide analog lane) to permit comparison of replicates run on different gels. Similarly, the relative levels of nucleotide-analog incorporation for each ribo- and deoxyribonucleotide pair were calculated at each position as

$$R = \frac{\left(\frac{N}{dN}\right)_{\text{nonselective}}}{\left(\frac{\sum N}{\sum dN}\right)_{\text{nonselective}}},$$

from background-subtracted band intensities on the nonselectively labeled calibration gels. The raw interference values from each replicate were then corrected for any differences in α-phosphorothioate ribonucleotide and deoxyribonucleotide analog incorporation at each position:

$$I_{\text{cal}} = \frac{I_{\text{raw}}}{\langle R \rangle},$$

where $\langle R \rangle$ is the average value obtained from two calibration replicates. Calculation of phosphorothioate effects was similar but used only the α-phosphorothioate ribonucleotide analogs:

$$\Theta_{\text{cal}} = \frac{\left(\frac{\langle N_{\text{nonselective}} \rangle}{N_{\text{selective}}}\right)}{\left(\frac{\langle \sum N_{\text{nonselective}} \rangle}{\sum N_{\text{selective}}}\right)}.$$

Because quantification of band intensities becomes less accurate as bands approach background levels, measured interference values become unreliable at very strong effect levels. We applied a cutoff at the sixfold effect level, truncating I_{cal} values at 6 and 0.167. After application of this cutoff, the I_{cal} values measured at each position were averaged across the three interference replicates run for each nucleotide analog pair. At a given position, interference effects in the two ligase ribozymes were compared using the two-tailed *t*-test for two populations of equal sizes assuming equal variance. Positions at which average interference effects differed at the $P < 0.05$ level and at which the average parent and improved ligase interference effects were separated by at least a factor of 2 are highlighted in Figure 7.

Dimethyl sulfate interference mapping

For DMS-interference mapping experiments, we used ligase constructs bearing the 18-nt 3′ tail that had been used during selection, which served as a primer-binding site for primer-extension readout. For DMS modification, 44 pmol body-labeled transcript together with 44 μg carrier tRNA in magnesium-free methylation buffer (50 mM Tris-HCl at pH 7.5, 1 mM EDTA) was mixed with 0.05 vol of freshly prepared 5% DMS (in 95% EtOH). After 4 min at 37°C, modification was stopped with 1 vol of freshly prepared DMS-stop solution (1.5 M NaOAc at pH 8.5, 7% β-mercaptoethanol, 0.1 mM EDTA), ethanol precipitated twice, and resuspended in water. This stock of modified RNA was split into two, with one aliquot to serve as the unselected control and the other to be put through the ligase reaction and APM gel-purification. In parallel, mock modifications were performed, using 95% EtOH alone in place of the 5% DMS in 95% EtOH. Because the tRNA required in the DMS reaction inhibited the ligase reaction, the aliquot of modified RNA that was to undergo selection was gel purified. The RNA was then mixed in a 2:1 molar ratio with substrate bearing a 5′-thiophosphate and a 1:1 molar ratio with unlabeled primer (5′-GCGCTGGCGTCTGGCCGG-3′, added to limit misfolding in the presence of the 3′ primer binding site), and incubated at 80°C for 2 min and then at 22°C for 5 min. Ligation was initiated by addition of stringent ligase buffer. Because ligation with the modified substrate and the 3′ extension was slow, reaction times were extended to 11 h and 22 min for the parent and improved ligases, respectively. Reactions were stopped with denaturing gel-loading buffer and resolved on a urea denaturing gel prepared with 6% polyacrylamide and 40 μM APM in the bottom portion of the gel and no APM in the top portion of the gel. Reacted ribozyme, which did not migrate into the bottom portion of the gel, was excised, eluted, and ethanol precipitated.

This reacted RNA was compared with the initial DMS-modified RNA and to mock-DMS-modified RNA using primer extension largely as described in Moazed et al. (1986). Briefly, each sample was mixed with labeled primer and hybridization buffer (final concentrations, 50 mM K-HEPES at pH 7.0, 100 mM KCl) and subjected to denaturation and slow cooling (85°C for 48 sec, followed by a decrease to 40°C at –2.5°C/min) in an MJ Research PTC-100 thermocycler. Primer-extension master mix (final concentrations, 125 mM Tris-HCl at pH 8.4, 200 μM dATP, 200 μM dTTP, 200 μM dGTP, 300 μM dCTP, 6.4 mM $MgCl_2$, 9 mM DTT, 5% glycerol, and 0.027 U/μL AMV-RT [Seikagaku]) was added and extension run at 42°C for 30 min. Extension was stopped with addition of 16 vol of a 1:3 mixture of salt (a solution of one part 300 mM NaOAc pH 8.5 and 1 mM EDTA) and 95% EtOH. After ethanol precipitation, pellets were resuspended in denaturing gel-loading buffer and treated with 0.33 M NaOH for 5 min at 90°C to degrade the RNA from the RT product. Primer extension products were then separated on a sequencing-thickness 8% polyacrylamide TBE-urea gel with a buffer gradient running from 0.5X TBE in the upper buffer chamber to 1X TBE with 0.75 M NaOAc in the lower chamber. To permit identification of bands, RT products were run in parallel with dideoxy sequencing ladders.

We quantified interference with respect to the extent of methylation at a given position as

$$I_{Me} = \frac{\left(\frac{N_{Me,\text{nonselective}}}{N_{Me,\text{selective}}}\right)}{\left(\frac{\sum N_{Me,\text{nonselective}}}{\sum N_{Me,\text{selective}}}\right)}.$$

Reported interference effects are the average of three replicates, each with its own nonselective control. Significance was determined as for the phosphorothioate and 2′-deoxy interference maps. Not all positions in the ligase ribozymes could be resolved. The cDNAs of the improved ligase consistently showed a gel compression spanning nucleotides 34–41, preventing measurement of interference effects at C35 and C38–C40; all primer-extension reactions performed on the parent ligase, regardless of the presence or absence of DMS modification, showed intense bands at the base of P6, where the reverse transcriptase encounters four stacked G:C pairs; and position 117 could not be resolved on parent ligase gels. Sequence differences also prevented some comparisons between the ribozymes: at positions 19, 20, 45, 73, 76, 84, 95, and 99, only one construct bears a C or A residue; the other is inaccessible to the DMS interference mapping by primer extension.

Figures 1C, 8, and 9 were prepared using the software package PyMOL (Delano 2002).

SUPPLEMENTAL MATERIAL

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

We thank Steven Wang for statistical advice; Piyali Chatterjee, I-hung Shih, and Edward Curtis for useful discussions regarding interference mapping methods; and members of the Bartel laboratory for support and comments. This work was supported by NIH grant GM61835. S.C.B. was an HHMI Pre-doctoral Fellow, D.M.S. was an NSF Pre-doctoral Fellow, and D.P.B. is an HHMI Investigator.

Received September 3, 2009; accepted September 15, 2009.

REFERENCES

Bartel DP, Szostak JW. 1993. Isolation of new ribozymes from a large pool of random sequences. *Science* **261:** 1411–1418.

Bergman NH, Johnston WK, Bartel DP. 2000. Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry* **39:** 3115–3123.

Bergman NH, Lau NC, Lehnert V, Westhof E, Bartel DP. 2004. The three-dimensional architecture of the class I ligase ribozyme. *RNA* **10:** 176–184.

Conrad F, Hanne A, Gaur RK, Krupp G. 1995. Enzymatic synthesis of 2′-modified nucleic acids: Identification of important phosphate and ribose moieties in RNase P substrates. *Nucleic Acids Res* **23:** 1845–1853.

Delano WL. 2002. PyMOL. DeLano Scientific. <http://pymol.sourceforge.net>.

Ekland EH, Bartel DP. 1995. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res* **23:** 3231–3238.

Ekland EH, Bartel DP. 1996. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382:** 373–376.

Ekland EH, Szostak JW, Bartel DP. 1995. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269:** 364–370.

Giver L, Bartel D, Zapp M, Pawul A, Green M, Ellington AD. 1993. Selective optimization of the Rev-binding element of HIV-1. *Nucleic Acids Res* **21:** 5509–5516.

Glasner ME, Bergman NH, Bartel DP. 2002. Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry* **41:** 8103–8112.

Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP. 2001. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* **292:** 1319–1325.

Joyce GF, Orgel LE. 1999. Prospects for understanding the origin of the RNA world. In *The RNA world* (eds. Gesteland RF et al.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Kühne H, Joyce GF. 2003. Continuous in vitro evolution of ribozymes that operate under conditions of extreme pH. *J Mol Evol* **57:** 292–298.

Lawrence MS, Bartel DP. 2005. New ligase-derived RNA polymerase ribozymes. *RNA* **11:** 1173–1180.

Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7:** 499–512.

Levy M, Griswold KE, Ellington AD. 2005. Direct selection of *trans*-acting ligase ribozymes by in vitro compartmentalization. *RNA* **11:** 1555–1562.

McGinness KE, Wright MC, Joyce GF. 2002. Continuous in vitro evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. *Chem Biol* **9:** 585–596.

Moazed D, Stern S, Noller HF. 1986. Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J Mol Biol* **187:** 399–416.

Müller UF, Bartel DP. 2003. Substrate 2′-hydroxyl groups required for ribozyme-catalyzed polymerization. *Chem Biol* **10:** 799–806.

Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. 2001. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proc Natl Acad Sci* **98:** 4899–4903.

Ordoukhanian P, Joyce GF. 1999. A molecular description of the evolution of resistance. *Chem Biol* **6:** 881–889.

Ortoleva-Donnelly L, Szewczak AA, Gutell RR, Strobel SA. 1998. The chemical basis of adenosine conservation throughout the *Tetrahymena* ribozyme. *RNA* **4:** 498–519.

Peattie DA, Gilbert W. 1980. Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci* **77:** 4679–4682.

Peattie DA, Herr W. 1981. Chemical probing of the tRNA–ribosome complex. *Proc Natl Acad Sci* **78:** 2273–2277.

Piccirilli JA, Vyle JS, Caruthers MH, Cech TR. 1993. Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature* **361:** 85–88.

R Development Core Team. 2008. R: A language and environment for statistical computing. <http://www.R-project.org>.

Ryder SP, Strobel SA. 1999. Nucleotide analog interference mapping. *Methods* **18:** 38–50.

Schmitt T, Lehman N. 1999. Nonunity molecular heritability demonstrated by continuous evolution in vitro. *Chem Biol* **6:** 857–869.

Shechner DM, Grant RA, Bagby SC, Koldobskaya Y, Piccirilli J, Bartel DP. 2009. Crystal structure of the catalytic core of an RNA polymerase ribozyme. *Science* (in press).

Sousa R, Padilla R. 1995. A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J* **14:** 4609–4621.

Stern S, Moazed D, Noller HF. 1988. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol* **164:** 481–489.

Strobel SA, Shetty K. 1997. Defining the chemical groups essential for *Tetrahymena* group I intron function by nucleotide analog interference mapping. *Proc Natl Acad Sci* **94:** 2903–2908.

Verma S, Eckstein F. 1998. Modified oligonucleotides: Synthesis and strategy for users. *Annu Rev Biochem* **67:** 99–134.

Wright MC, Joyce GF. 1997. Continuous in vitro evolution of catalytic function. *Science* **276:** 614–617.

Zaher HS, Unrau PJ. 2007. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13:** 1017–1026.

SUPPLEMENTARY TEXT

An additional base-specific effect revealed by Monte Carlo analysis

In addition to position 76, nucleotide 94 shows a nucleotide-specific rate effect at 10 mM Mg^{2+} that widens at 1 mM Mg^{2+}. Nucleotide 94 seems to be not so much a case where one nucleotide succeeded as a case where another failed; notably, though, the disfavored nucleotide was not the parental one. G94 was slightly but not significantly under-represented among successful isolates; its disadvantage only became clear in kinetic assays, where in which G94 isolates performed somewhat more poorly than did U94 isolates at 10 mM Mg^{2+} (0.49 ± 0.36 min^{-1} vs. 1.52 ± 0.55 min^{-1}; Bonferroni-adjusted $p < 0.031$) and much more poorly than did either U94 or A94 isolates at 1 mM Mg^{2+} (0.01 ± 0.01 min^{-1}, 0.13 ± 0.11 min^{-1}, 0.11 ± 0.09 min^{-1}, respectively; Bonferroni-adjusted $p < 0.028$ for the G–U comparison and 0.035 for the G–A comparison). Thus, U94 isolates slowed by just 12-fold on average, and A94 isolates by 14-fold, at 1 mM vs. 10 mM Mg^{2+}; by contrast, the nearly 40-fold drop in rate constant for G94 isolates is exactly on par with the drop seen in the parent ligase. Given that the parent ligase itself bears a uracil at this position, its steep metal dependence must be due to some other cause; indeed, for the G94 isolates to slow at low Mg^{2+} by only as much as the parent, they must have acquired some other mutation(s) that can compensate for the cost of G94. With only a subset of the 35 clones to examine, however, we lacked the statistical power to identify that beneficial change.

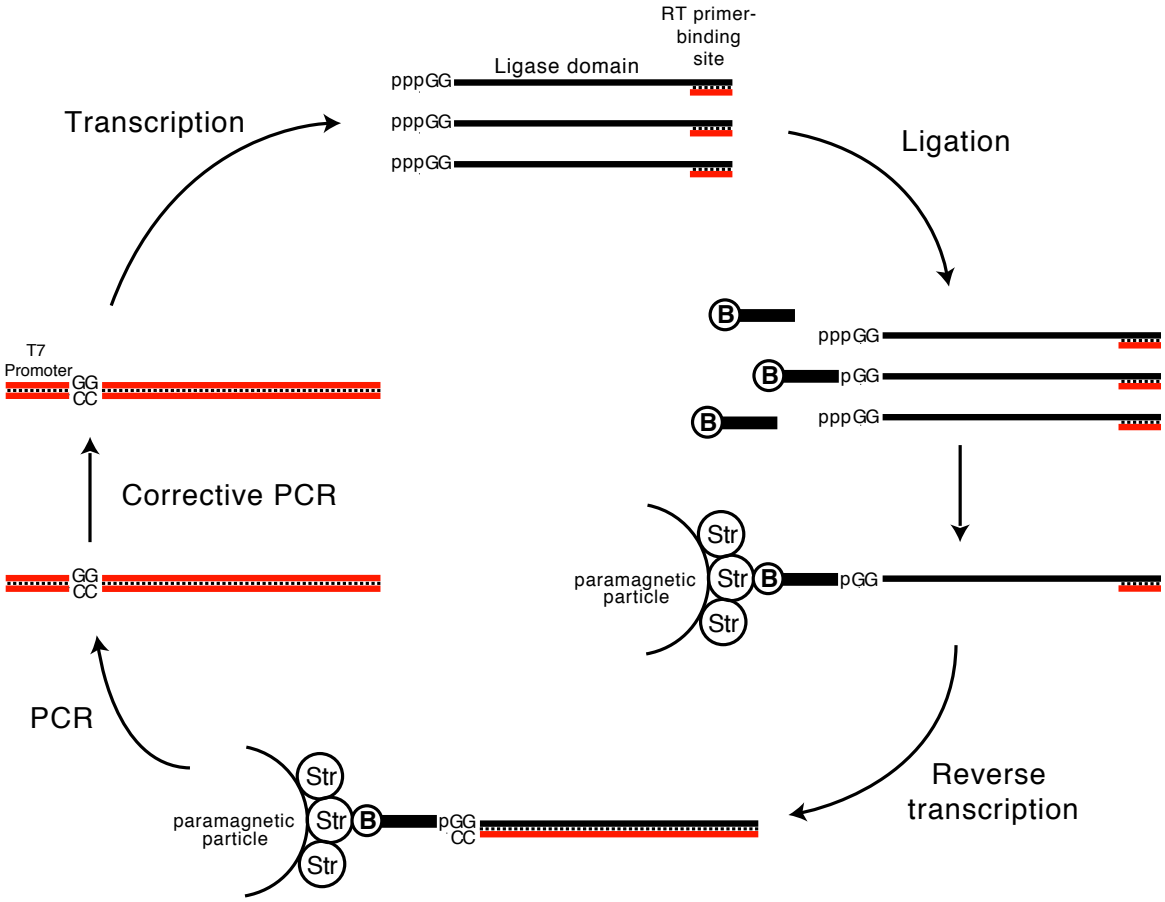
Interactions predicted by interference mapping and structural analysis

At a number of positions showing significant interference effects in both ligases, the crystal structure predicts interactions with other atoms that are not probed by the interference maps presented. High backgrounds prevented the quantification of interference effects at the very 5'

and 3' ends of either ligase; within those bounds, we have probed only the pro-R_p oxygens, the 2'-hydroxyl groups, and the N1 atoms of adenosines and N3 atoms of cytosines. Thus, we predict but cannot confirm that the 2'-deoxy interference at position 12 reflects disruption of a hydrogen bond to an ordered water molecule that coordinates a bound metal; the 2'-deoxy interference at position 75, disruption of a hydrogen bond to the 3' oxygen of position 34; the 2'-deoxy interference at position 113, disruption of hydrogen bonds to the imine (N1) and exocyclic amine (N6) groups of A3; and the 2'-deoxy interference at position 114, disruption of hydrogen bonds to the pro-R_p oxygen of A4 and to another metal-bound water.

Supplementary Figure S1. *In vitro* selection scheme. Ligase variants were incubated with a biotinylated substrate (**B**), and those ribozymes that ligated themselves to substrate were separated using streptavidin-coated paramagnetic beads. Reverse transcription and PCR regenerated dsDNA, and further PCR steps (corrective PCR) changed the substrate sequence into a functional T7 RNA polymerase promoter. The resulting DNA was transcribed to generate the RNA pool for the next round of selection. The first round was performed as shown, and furhter rounds were modified slightly as detailed in Materials and Methods. The stringency of the ligation step was increased with each successive round by decreasing pH and/or reaction time.

Figure S1


RT primer-binding site
Ligase domain
pppGG
Transcription
Ligation
T7 Promoter
GG
CC
B
pGG
Corrective PCR
Str
paramagnetic particle
PCR
Reverse transcription