The structural basis of RNA-catalyzed RNA polymerization

David M Shechner[1–4] & David P Bartel[1–3]

Early life presumably required polymerase ribozymes capable of replicating RNA. Known polymerase ribozymes best approximating such replicases use as their catalytic engine an RNA-ligase ribozyme originally selected from random RNA sequences. Here we report 3.15-Å crystal structures of this ligase trapped in catalytically viable preligation states, with the 3′-hydroxyl nucleophile positioned for in-line attack on the 5′-triphosphate. Guided by metal- and solvent-mediated interactions, the 5′-triphosphate hooks into the major groove of the adjoining RNA duplex in an unanticipated conformation. Two phosphates and the nucleophile jointly coordinate an active-site metal ion. Atomic mutagenesis experiments demonstrate that active-site nucleobase and hydroxyl groups also participate directly in catalysis, collectively playing a role that in proteinaceous polymerases is performed by a second metal ion. Thus artificial ribozymes can use complex catalytic strategies that differ markedly from those of analogous biological enzymes.

The RNA world hypothesis proposes that primordial organisms lacked both DNA genes and protein enzymes, and instead used RNA as the sole biopolymer of information-storage and catalysis[1]. Support for this theory comes from the observation that many of life's most fundamental processes depend on the action of RNA, its precursors or its derivatives[1–4]. Additional support has come through the study of ribozymes. Although known natural ribozymes catalyze only the cleavage or rearrangement of phosphodiester bonds[5] and the formation of peptide bonds[4], the diversity of artificial ribozymes produced from *in vitro* selection experiments has revealed the catalytic potential of RNA to be quite broad[6,7]. Of particular relevance to the RNA world hypothesis would be ribozymes with the RNA-dependent RNA polymerization activity needed for general RNA replication[8,9].

Known catalytic RNAs with activities best approximating that required for general RNA replication derive from the class I ligase, an artificial ribozyme selected from random RNA sequences[10,11]. This ribozyme rapidly promotes a reaction with chemistry identical to that catalyzed by proteinaceous enzymes that replicate RNA (**Fig. 1a**): nucleophilic attack by a 3′-hydroxyl on a 5′-triphosphate in the context of Watson-Crick base pairing to form a normal phosphodiester RNA linkage with concomitant release of pyrophosphate[10,11]. Indeed, ligase variants have been engineered to catalyze primer extension using exogenous nucleoside triphosphates (NTPs) and an internal template (**Fig. 1a**, inset)[12]. Through further engineering and *in vitro* selection, polymerase ribozymes have been developed in which the ligase provides the catalytic engine for RNA polymerization using NTPs and the information of exogenous RNA templates[13–16]. Although more efficient with some primer templates than with others,

this primer-extension reaction is general in that all primers and templates tested support detectable extension (provided that the 3′ terminus of the primer pairs to the template)[13–16]. Optimized versions of these polymerase ribozymes can extend a primer by up to 22 nucleotides[15], or by up to 95 nucleotides when appending to the ribozyme a segment that pairs to a short region of the template[16].

We recently reported the crystal structure of the class I ligase self-ligation product[17,18]. This structure revealed details of the ribozyme architecture and the location of the active site. However, the active site visualized in this post-ligation structure lacked the pyrophosphate leaving group and had no detectable catalytic metal ions. To develop a deeper understanding of the structural basis for RNA-catalyzed RNA polymerization, we determined the crystal structures of the ligase trapped in catalytically viable preligation states and then biochemically interrogated new hypotheses for the functions of active-site residues. These results corrected our model of the ligase catalytic mechanism and showed that the ligase uses a catalytic strategy both more complex than those of other artificial ribozymes[19–23] and distinct from that of the proteinaceous polymerases.

RESULTS

Structures of class I ligase preligation complexes

Because the ligase normally undergoes extremely rapid self-ligation[18], to visualize an intact, catalytically viable active site, we used two modifications that each markedly reduce, but do not completely ablate, catalysis. First, the active-site nucleotide C47 was mutated to uridine, which reduces the catalytic rate by a factor of >10[4] (ref. 17, **Fig. 1a**). Second, the preligation complex was reconstituted in a mixture of

Figure 1 The class I ligase ribozyme preligation complex. (**a**) Secondary structure of the C47U crystallization construct[17], depicted undergoing ligation (curved arrows). The nucleophile, electrophile and leaving group (red), active-site backbone phosphates (yellow bars) and nucleotides added to facilitate crystallization (gray) are indicated. Residue numbering, base-pair geometries and tertiary interactions, as in reference 17. Inset, primer extension catalyzed by a ligase variant resembling the self-ligation construct, except that exogenous GTP replaces G1–A4. (**b**) Overview of the Mg^{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. Active-site metal ions are shown as spheres (orange, Mg^{2+}; green, metal-coordinated water). Meshes are simulated-annealing $|F_o| - |F_c|$ OMIT maps calculated without G1 (magenta, contoured at 5σ) or active-site solvent atoms (dark blue, contoured at 4.5σ). (**c**) Superposition of Mg^{2+}–Sr^{2+} preligation and product (PDB ID: 3HHN)[17] structures. (**d**) Superposition of Mg^{2+}–Sr^{2+} preligation and product structures (sticks and black lines, respectively) near the active site. A structural metal ion and water molecules are shown as orange and green spheres, respectively; active-site solvent atoms have been removed from view.
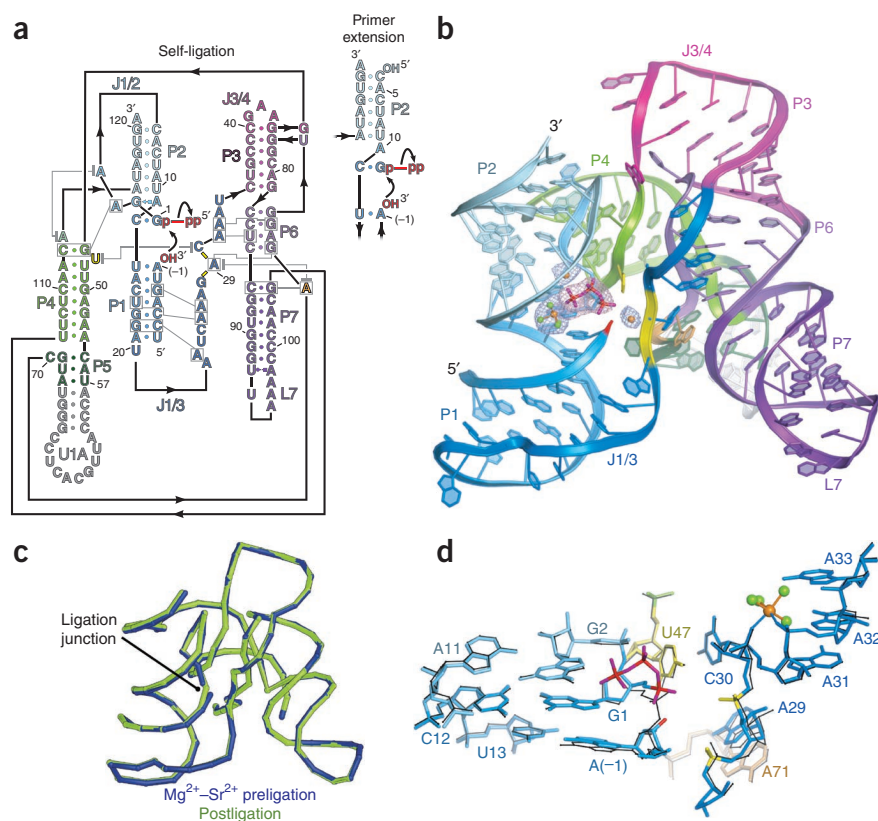
Ca^{2+} and Sr^{2+}, which diminishes the self-ligation rate by $>10^5$ but allows the enzyme to fold into a conformation indistinguishable (within the resolution limits of Fe–EDTA solvent-accessibility mapping) from that of its active state[24]. Using these modifications, the ribozyme–substrate preligation complex was crystallized together with a mutant fragment of U1A (a spliceosomal protein used to facilitate crystallization)[17]. Mg^{2+}-bound crystals were obtained by soaking these crystals (hereafter, 'Ca^{2+}–Sr^{2+} complex' or 'Ca^{2+}–Sr^{2+} preligation' crystals) into isotonic conditions containing Mg^{2+} (hereafter, the 'Mg^{2+}–Sr^{2+} complex' or 'Mg^{2+}–Sr^{2+} preligation' crystals; **Table 1**). Crystals soaked without Sr^{2+}, and those soaked with other Mg^{2+} isosteres (such as Mn^{2+} or Yb^{3+}), proved intractable. The structure of each of these crystals was solved to 3.15 Å, allowing visualization of the ligase poised before catalysis (**Fig. 1b**, **Supplementary Fig. 1a** and **Supplementary Methods**).

The global conformation of the ligase was markedly similar in the pre- and post-catalytic states, being more greatly altered by binding different metals than it was by undergoing ligation (**Fig. 1c,d** and **Supplementary Fig. 1b,c**). The Mg^{2+}–Sr^{2+} complex and product structures superposed with an all-atom r.m.s. deviation of 0.31 Å, smaller than the mean precision of either structure (ref. 17), whereas the r.m.s. deviation between the Ca^{2+}–Sr^{2+} complex and product structures was 0.76 Å. In the latter case, the majority of the structural perturbations fell either near the active site and its environs (**Supplementary Fig. 1b–d**) or at sites of inner-sphere metal-ion coordination and thus attributable to steric effects incurred by Ca^{2+}, which is ~0.3 Å larger than Mg^{2+} (ref. 25). All of these structural perturbations were alleviated in the active site of the Mg^{2+}–Sr^{2+} complex (**Fig. 1d**). The close resemblance between the active site of the Mg^{2+}–Sr^{2+} complex and that of the product implied that, unlike for the analogous variant of the hepatitis delta virus (HDV) ribozyme[26],

the diminished activity of the C47U mutant was not caused by a gross structural perturbation. Furthermore, as expected from its minor effect on catalysis[24], the presence of Sr^{2+} had no detectable effect on the ribozyme structure, and no peaks corresponding to Sr^{2+} density were observed in the electron-density maps.

Conformation of the 5′-triphosphate

Within the active site of each preligation structure, 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. 2a–c**). This conformation is markedly different from that adopted by deoxynucleoside triphosphates (dNTPs) or NTPs in the active sites of proteinaceous polymerases, wherein the triphosphate is pulled away from the plane of the adjoining base[27,28]. Moreover, the (d)NTPs in these enzymes are typically Mg^{2+} bound, with the metal ion coordinated by two or more phosphate groups[29]. Although the ligase 5′-triphosphate makes no such polydentate contacts, the 5′-γ-phosphate coordinates a Mg^{2+}–$4H_2O$ cluster positioned by groups on the Hoogsteen face of G1 and G2 (**Fig. 2b,d**). If protonated, a γ-phosphate oxygen would also hydrogen bond with the G2 N7, further stabilizing the hooked triphosphate conformation (**Fig. 2b,d**).

In Mg^{2+}, docking into the major groove places the 5′-triphosphate in a nearly ideal conformation for in-line nucleophilic attack: its α-phosphate is positioned 3 Å from the substrate 3′-hydroxyl, with a ~176° angle formed between these two atomic centers and the [α,β]-bridging oxygen leaving group. Moreover, pulling the β- and γ-phosphates proximal to the G1 base exposes the [α,β]-bridging oxygen, presenting it toward potentially stabilizing functional groups in the active site. In Ca^{2+}, which greatly slows catalysis, this favorable conformation is perturbed. Because the Ca^{2+}–hydrate cluster near G2 does not support close approach of the γ-phosphate and the

Table 1 Data collection and refinement statistics

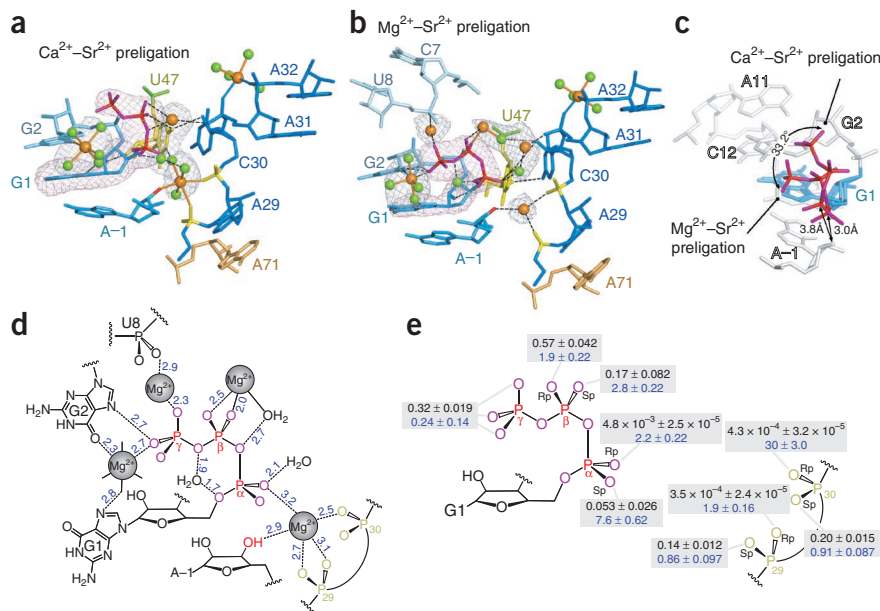|	Ca²⁺–Sr²⁺ preligation complex	Mg²⁺–Sr²⁺ preligation complex
Data collection		
Space group	P1	P1
Cell dimensions		
a, b, c (Å)	58.70, 69.99, 71.86	59.19, 70.24, 71.21
α, β, γ (°)	99.86, 99.73, 103.68	99.86, 99.34, 103.81
Resolution (Å)	30–3.15 (3.26–3.15)ᵃ	40–3.15 (3.26–3.15)
R_{merge} (%)	9.8 (39.8)	13.5 (42.9)
Mean $I/\sigma(I)$	13.5 (2.1)	17.5 (3.2)
Completeness (%)	98.4 (94.1)	99.1 (98.3)
Redundancy	3.3 (2.9)	7.2 (5.7)
Refinement		
Resolution (Å)	29.9–3.15	37.9–3.15
No. reflections	18,190	18,739
R_{work} / R_{free} (%)	21.2 / 25.5	19.7 / 24.0
No. atoms		
RNA and protein	7,231	7,254
Ligand / ion	26	46
Water	38	51
Mean B factors (Å²)		
RNA and protein	89.9	76.4
Ligand/ion	103.2	75.5
Water	88.0	76.0
r.m.s. deviations		
Bond distances (Å)	0.015	0.006
Bond angles (°)	1.133	1.310

Data from one crystal were used for each complex structure.
ᵃValues in parenthesis are for highest-resolution shell.

G2 base (**Fig. 2a**), the triphosphate is rotated ~33° into the major groove (**Fig. 2c**), which contorts the angle between nucleophile, electrophile and leaving group to ~154° (**Fig. 2a**). This conformation also pulls the α-phosphorus to displace it ~1 Å further away from the 3′-hydroxyl nucleophile.

Metal ions in the active site

Although biochemical data[17,18] support a model in which the active-site A29 and C30 backbone phosphates directly coordinate a catalytic metal ion, electron density for this metal is not observed in the product structure[17]. This metal was observed in the preligation complexes: during crystallographic refinement, $|F_o| - |F_c|$ difference Fourier maps prominently featured 7.8σ and 4.8σ positive peaks centered within the active sites of the Ca²⁺–Sr²⁺ and Mg²⁺–Sr²⁺ complexes, respectively (**Fig. 2a,b**). In the Ca²⁺–Sr²⁺ structure, this peak perfectly accommodated a partially hydrated Ca²⁺ ion, in which the metal was centered between the A29 pro-Sp oxygen, the C30 pro-Rp oxygen and the nucleophilic A(−1) 3′-hydroxyl (**Fig. 2a**). The corresponding peak in the Mg²⁺–Sr²⁺ maps represented a less ideal Mg²⁺ binding site[30], in which the active-site Mg²⁺ center was ~2.5–2.7 Å from the A(−1) 3′-hydroxyl, C30 pro-Rp oxygen and each of the A29 nonbridging phosphate oxygens (**Fig. 2b,d**). Moreover, in neither preligation structure was the active-site metal ion within inner-sphere distance of the 5′-α-phosphate nonbridging oxygens, even though the ribozyme is substantially inhibited by sulfur substitution of the 5′-α-phosphate pro-Rp oxygen, implying that ligase uses a direct inner-sphere contact to this group[31]. The discrepancy between the biochemical and crystallographic data might be explained if the phosphorothioate substitution perturbed an outer-sphere contact[32], which could be easily accommodated in either structure. Alternatively, because the formation of a more ideal active-site Mg²⁺-binding pocket would require only minor (0.5–1.0 Å) adjustments to nucleotides A(−1), G1, A29 and C30, such a site might have been perturbed by the C47U mutation or might be fully occupied only during the transition state. Furthermore, the model's mean coordinate precision (0.4 Å) and the maximum observed inner-sphere bonding distance for Mg²⁺ (2.45 Å)[30] leave open the possibility of direct coordination of the α-phosphate pro-Rp oxygen in the Mg²⁺–Sr²⁺ complex.

Natural RNA and DNA polymerases use a two-metal mechanism in which metal A activates the 3′-hydroxyl nucleophile for attack, metal B (aided by a general acid[28]) facilitates loss of the pyrophosphate leaving group, and both metals stabilize the geometry and

Figure 2 The class I ligase active site.
(**a**) The active site, as observed in the Ca²⁺–Sr²⁺ preligation structure. Calcium ions and waters (orange and green spheres, respectively), inner-sphere metal interactions (thin sticks) and outer-sphere interactions and hydrogen bonds (black dashes) are indicated. Meshes are simulated-annealing $|F_o| - |F_c|$ omit maps calculated without GTP (magenta, contoured at 5σ) or active-site solvent atoms (navy, contoured at 4.5σ). (**b**) The active site, as observed in the Mg²⁺–Sr²⁺ preligation structure. As in **a**, except that the orange spheres depict Mg²⁺ ions. (**c**) Reorientation of the 5′-GTP during metal-ion exchange. (**d**) Solvent interactions in active site of the Mg²⁺–Sr²⁺ preligation complex. Blue numbers indicate the distance, in angstroms, between atomic centers; the error of these measurements is approximated by the mean crystallographic coordinate precision, 0.4 Å. Waters hydrating the Mg²⁺ bound by G1 and G2 are represented by black lines. (**e**) Summary of phosphorothioate interference (black, $^{Mg}k_{thio}/^{Mg}k_{oxy}$) and metal rescue (blue, $(^{X}k_{thio}/^{X}k_{oxy})/(^{thioMg}k_{thio}/^{Mg}k_{oxy})$) values for oxygens at the active site. Functional groups on the 5′-GTP were interrogated in primer-extension assays, and those at positions 29 and 30 were interrogated in self-ligation assays. Error values are the s.d. from three independent experiments.

Figure 3 Biochemical interrogation of C47. (**a**) View of the active site observed in the Mg^{2+}–Sr^{2+} complex, roughly orthogonal to that in **Figure 2b**, highlighting interactions between C47, C30 and the 5′-triphosphate. Depiction and coloring are as in **Figure 2b**, with the thin red line indicating the assumed line of nucleophilic attack. (**b**) Self-ligation rates of C47 variants, relative to that of the parental ligase (red bar). The rates of the faster constructs (red and dark blue bars) were measured at pH 6.0. To allow time courses to approach completion within 24 h, the rates of slower constructs (light blue bars) were measured at pH 8.0 and re-normalized assuming log-linear pH dependencies[24]. Constructs marked with an asterisk were synthesized with a 2′-deoxyribose at position 47, and rates were normalized to account for this modification. The 4-thiouridine variant showed biphasic kinetics in which approximately 10% reacted in a burst phase with a relative rate of 0.23 ± 0.04 (s.d.); shown is the relative rate relevant to the majority of the molecules. Error bars, s.d. from three independent experiments.
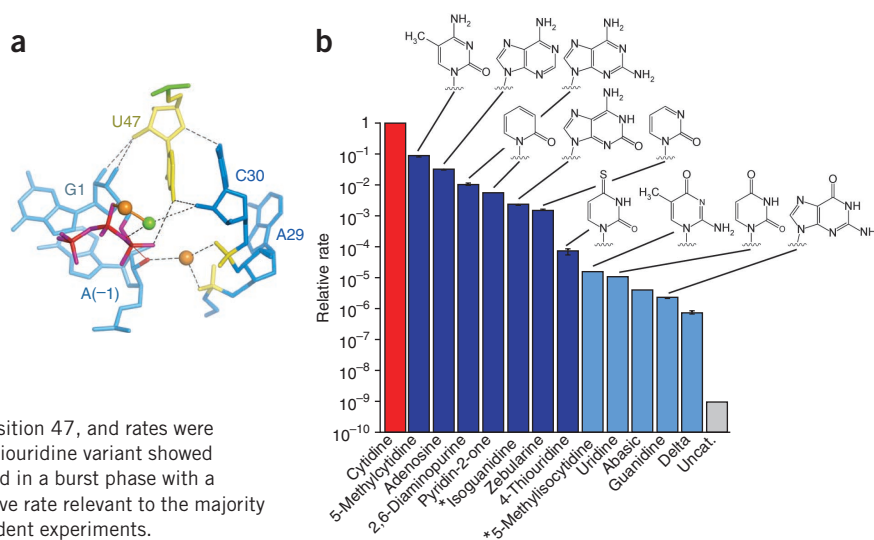
charge of α-phosphate during the transition state[27,33]. If the ligase transition state adopts a more ideal Mg^{2+}-binding site—and one that supports an inner-sphere contact with the α-phosphate—then the active-site metal ion would be analogous to metal A. However, we did not observe a metal ion analogous to metal B (**Fig. 2a,b**), which suggested that its role might be played by other groups. Nonetheless, we did observe an array of solvent and metal interactions that stabilize the hooked conformation of the triphosphate, some of which might also help stabilize a growing negative charge on the β phosphate during the transition state (**Fig. 2a,b,d**). These include the partially hydrated Mg^{2+} cluster bound to the Hoogsteen faces of GTP1 and G2, a Mg^{2+} ion jointly coordinated by the GTP1 γ-phosphate and the U8 Pro-*Sp* nonbridging oxygen, and a water molecule bound by the GTP1 5′ and [β,γ]-bridging oxygens (**Fig. 2b,d**). An additional partially hydrated Mg^{2+} ion near the A31 Pro-*Sp* nonbridging oxygen bridges the gap between the P6 A-minor triad[17] and the 5′-triphosphate, potentially placing a water molecule near the α-Pro-*Rp* nonbridging phosphate oxygens. Also of potential catalytic relevance, we observed evidence for a partially hydrated Mg^{2+} near the β-phosphate, which would place an ordered water molecule within hydrogen-bonding distance of the triphosphate [α,β]-bridging oxygen, the reaction leaving group (**Fig. 2b,d**).

To probe for functional metal-ion interactions at the active site, we carried out site-directed sulfur substitutions and thiophilic metal-rescue experiments[34]. We surveyed all nonbridging phosphate oxygens at positions 1, 29 and 30, using GTP analogs in primer-extension reactions (**Fig. 1a**, inset)[12] to survey the oxygens at position 1 and using splint-ligation products to survey oxygens within the ribozyme active site (**Fig. 2e** and **Supplementary Figs. 2** and **3**). Sulfur substitutions at the G1-α-Pro-*Rp* or the A29 and C30-pro-*Rp* oxygens inhibited catalysis by factors of 10^3–10^4, consistent with previous results[17,31]. Among these, the C30-pro-*Rp* substitution was substantially rescued by addition of a thiophilic metal (30-fold rescue in Cd^{2+}), a result that supported the proposed functional inner-sphere metal contact to this oxygen. Because lack of rescue is difficult to interpret, similar inner-sphere contacts might be important at the other strongly interfering positions. With the exception of the G1-α-pro-*Sp* site, where phosphorothioate inhibition and metal rescue were modest (inhibition to 0.053 that of the parent with 7.8-fold rescue), inhibition and metal rescue were weak at all other sites probed (inhibition to 0.17–0.6 with 1.9–2.2-fold rescue). These results generally corroborated the contacts

observed crystallographically (**Fig. 2d,e** and **Supplementary Fig. 2**) and implied that, if direct metal coordination at the β or γ oxygens have more than modest roles in substrate binding or catalysis, then these roles are redundant and could only be uncovered by simultaneously substituting more than one oxygen.

The C47 nucleobase participates in catalysis

On the basis of their proximity to the 5′-triphosphate, two additional functional groups seemed capable of direct catalytic roles (**Fig. 3a**). In the Mg^{2+}–Sr^{2+} complex, the U47 O4 group (mutated from the C47 N4 exocyclic amine to slow ligation) is 3.7 Å from the 5′-α-phosphate pro-*Rp* oxygen, a distance that decreases to 3.1 Å for the analogous interaction in the product structure (**Supplementary Movie 1**). In each structure, the C30 2′-hydroxyl (owing to the unusual 3′-exo ribose conformation[35] of C30, a consequence of the P6 A-minor triad[17]) is within hydrogen-bonding distance of U47 O4 (C47 N4 in the parent). In the Mg^{2+}–Sr^{2+} complex, the C30 2′-hydroxyl is also within hydrogen-bonding distance of a well-ordered water molecule near the leaving-group oxygen. Hypothesizing that the C47 exocyclic amine and the C30 2′-hydroxyl might perform roles played by metal B of proteinaceous polymerases, we biochemically interrogated their contributions to catalysis.

We first assessed 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. 3b** and **Supplementary Fig. 4a**). Variants that replaced this functional group with a carbonyl were among the slowest, with self-ligation rates 10^{-6}–10^{-4} times that of the unmodified ligase, irrespective of the protonation state of the imine (**Supplementary Fig. 4b**). Variants that preserved a major-groove exocyclic amine—irrespective of the protonation state of the proximal imine group—were not as strongly affected, with rates 0.0024–0.09 times of that of the unmodified ligase. Substitution of C47 with zebularine (which lacks the N4 amino group altogether but is otherwise identical to cytosine) and pyridin-2-one ('P2o', which lacks both the cytosine N4 and N3 groups) diminished the relative rate to 0.0016 and 0.006, respectively (**Fig. 3b** and **Supplementary Fig. 4a**). From these data, we conclude that it is the C47 N4 amine, and not the N3 imine, that participates in catalysis, and that replacing N4 with a carbonyl has an inhibitory effect beyond that of simple loss of the amine. Furthermore, on the basis of the loss of activity

Figure 4 Catalytic roles of active-site functional groups. (**a**) Proton inventories for self-ligation by wild type and the indicated variants. The solid black line indicates the best fit for a single-proton transfer model; the dashed blue line indicates the best fit for a two-proton transfer model[28,37,38]. Error bars, s.d. from three independent experiments. (**b**) Relative rates for self-ligation of constructs with the indicated substitutions at the 2′ position of C30, surveyed in combination with the indicated subsitions at position 47. For each C47 substitution, rates are normalized to that of the C30 2′-OH construct. Error bars, s.d. from three independent experiments. (**c**) Schemes summarizing functional interaction between the C30 2′-hydroxyl and the C47 N4, showing the effects of removing these groups (left) or replacing them with inhibitory groups (right). Substitutions, relative to the wild type,

are denoted in red. Fractional numbers indicate the decrease in self-ligation rate, k_{right}/k_{left} or k_{bottom}/k_{top}, resulting from the indicated substitution. At pH 6.0, the observed rate of the unmodified ligase (upper left corner of each scheme) was 0.93 ± 0.009 min^{-1} (s.d.).
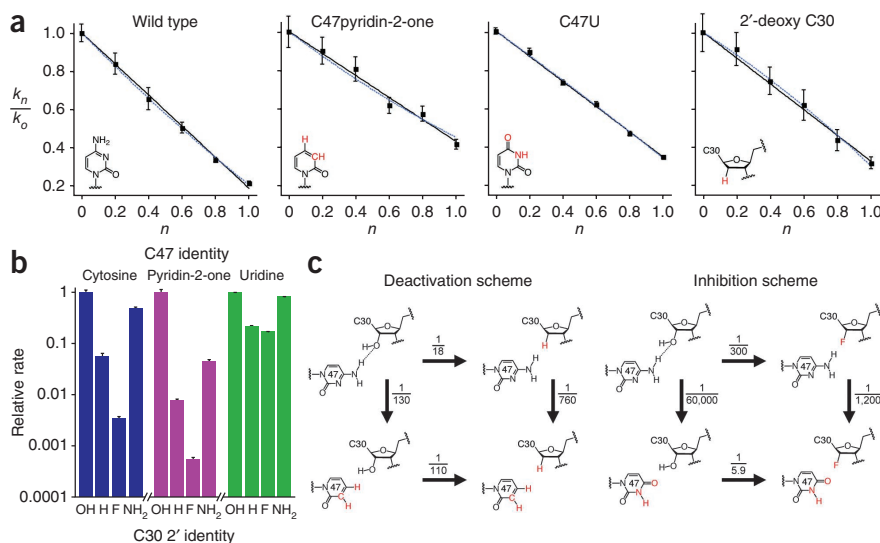
observed with the zebularine and pyridin-2-one substitutions, the energetic contribution of the C47 N4 was less than that expected from a general acid-base catalyst, but more than that anticipated from a single hydrogen bond donor[36]. Because the C47pyridin-2-one variant was comparably more active yet lacked any pertinent nucleobase functional groups, we focused on this variant in subsequent studies.

## Electrostatic catalysis in the ligase transition state

We next determined whether the C47 N4 amine functions electrostatically, through the donation of hydrogen bonds, or as a general acid catalyst, through the donation of protons. Between pH 5.8 and 8.5, self-ligation by both the C47pyridin-2-one and C47U variants showed log-linear pH dependence with a slope of 1.0 (**Supplementary Fig. 5a**), behavior identical to that observed for the parental ligase[24]. These results suggest that a single net deprotonation event determines the rate of ligase catalysis, regardless of the presence of the C47 N4 group. To more directly address the number of protons in transit during the transition state, we carried out proton inventory experiments[28,37] on the parental ligase and the two variants (**Fig. 4a**). For the parental ligase, the dependence of the relative self-ligation rate constant with molar fraction D$_2$O fit well to a line with a slope (Φ) of 0.19 ± 0.01 (**Fig. 4a**, left), which corresponded to a solvent deuterium kinetic isotope effect (SDKIE)[37] of 5.0 ± 0.3, comparable to those observed for some natural polymerases[28,37]. A higher-order polynomial fit (**Fig. 4a**, left, dotted line) failed to explain the data with greater significance (F tests: 0.99 and 0.98 for the single-proton and double-proton models, respectively; $R^2 > 0.99$ for each)[38] and yielded SDKIE values of 4.1 ± 0.4 and 1.15 ± 0.1 for the individual protons transferred, which implied that, were the ligase to shuttle two protons during its transition state, the second proton would contribute insignificantly to the reaction rate. Likewise, proton inventories for the C47P2o and C47U variants could be convincingly fit to only single-proton transfer models, yielding SDKIE values of 2.4 ± 0.1 and 2.9 ± 0.1, respectively (**Fig. 4a**, middle). That each of these variants had an SDKIE lower than that of the parent might imply that the transition states they stabilize depend less strongly on proton transfer[37]. Because a linear proton-inventory fit can have alternative interpretations[37,38],

these data were insufficient to explicitly quantify the number of protons in transit during the transition state (**Supplementary Discussion**). Nonetheless, because neither the removal of the C47 N4 amino group nor its replacement with an inhibitory carbonyl group changed the apparent number of protons in transit during ligation, these results suggest that C47 functions electrostatically during catalysis. This role differs from that of an active-site cytosine in the HDV ribozyme, for which the N3 group functions as a general acid catalyst[26,39,40].

Interactions between active-site functional groups

We next investigated the function and energetic contribution of the C30 2′-hydroxyl group. Self-ligation activity by the 2′-deoxy C30 variant was reduced to nearly 5% that of the parent and followed a log-linear pH dependence between pH 5.8 and 8.5 (**Supplementary Fig. 5a**, right). This construct also showed a linear proton inventory with a measured SDKIE of 3.1 ± 0.2 (**Fig. 4a**, right), implying that the C30 2′-hydroxyl is not involved in proton transfer. Substitution of the C30 2′-hydroxyl with a 2′-fluoro group (which can accept but not donate hydrogen bonds) reduced self-ligation activity to 0.3% of the parental activity, whereas substitution with a 2′-amino group (which can donate but not accept hydrogen bonds) restored activity to within 50% of the parental activity (**Fig. 4b**, left). These results indicate that the C30 2′-hydroxyl—much like the C47 N4 amine—functions electrostatically, acting as a catalytically important hydrogen bond donor.

If the C47 and C30 functional groups act independently, then the effect of altering one group would be unchanged by the status of the other. To test this hypothesis, we generated C47pyridin-2-one and C47U self-ligase variants bearing the same C30 2′-modifications described above. Although each series of variants showed the same qualitative trend in behavior, the magnitude of these effects depended on the identity of the C47 base (**Fig. 4b**, middle and right). In the C47P2o variants, which lacked the C47 N4 amine, C30 2′ substitutions were ~10–20 times more severe than in the parental C47 background. Conversely, in the C47U variants, in which the N4 amine was replaced by an inhibitory carbonyl group, the activities of all variants fell within one order of magnitude of one another, with the C30 2′-amino modification restoring activity to ~82% that of C47U.
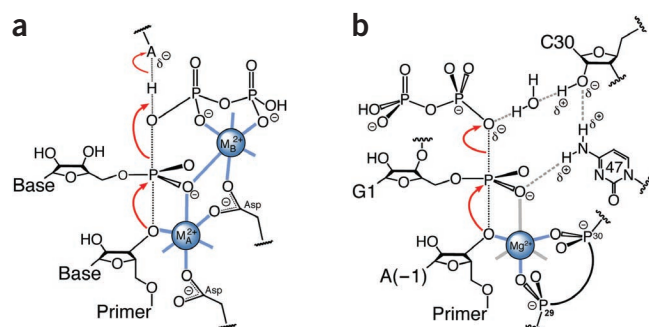
Figure 5 Transition-state stabilization by protein and RNA active sites. (**a**) Catalysis by proteinaceous polymerases[28]. Black dotted lines indicate bonds formed or broken during the transition state. Red arrows indicate the reaction direction during polymerization. (**b**) Catalysis by the class I ligase. Gray dashed lines denote hydrogen bonds. Thick lines denote inner-sphere contacts with a single catalytic metal ion; those in blue are observed crystallographically, those in gray are either inferred from biochemical results or are presumed to be water (not shown). Red arrows indicate the reaction direction during ligation or polymerization.

Schematic examination of the effects of single and double mutations indicated that in the absence of the C30 2′-hydroxyl, catalysis became six times more dependent on the presence of the C47 N4 amine, and *vice versa* (**Fig. 4c**, left). Conversely, when catalysis was quelled by an inhibitory group at either position, it became less dependent on the status of the other group (**Fig. 4c**, right). These results strongly support a model in which the C47 N4 exocyclic amine donates a hydrogen bond to the C30 2′-hydroxyl, which in turn donates a hydrogen bond during catalysis.

DISCUSSION

The class I ligase is the first artificial ribozyme to be crystallographically visualized in a catalytically viable precatalytic state and is one of three such ribozymes for which the active site has been biochemically interrogated in detail[19,20,23]. Our results indicate a model for transition-state stabilization in which the A29 and C30 pro-*Rp* phosphate oxygens coordinate a catalytic Mg^{2+} cofactor that activates the 3′-hydroxyl for nucleophilic attack, and the C47 N4 exocyclic amine and the C30 2′-hydroxyl participate in a network of hydrogen bonds that stabilizes both the leaving group and the transition-state geometry (**Fig. 5**). 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. The C47 amine also donates a hydrogen bond to the C30 2′-hydroxyl, which contributes a water-mediated partial positive charge to the [α,β]-bridging oxygen, thereby helping to stabilize the leaving group.

The C47 and C30 functional groups are proposed to act through hydrogen bonding rather than proton transfer because removing either group does not change the proton inventory and because removing the C47 N4 amine does not influence the pH dependence (**Fig. 4a** and **Supplementary Fig. 5a**). In contrast, removal of a general acid catalyst from the active sites of proteinaceous polymerases changes both of these properties[28]. Our model proposes that a network of hydrogen bonds orients a partial positive charge on a bound water molecule, allowing it to stabilize the developing negative charge on the leaving group during the transition state. Although this same water might eventually contribute a proton to the leaving group, our results suggest that if this transfer happens, it happens well after the transition state and therefore is not rate limiting.

In its use of an acid-bound Mg^{2+} cofactor to activate the 3′-hydroxyl for nucleophilic attack, the catalytic strategy of the ligase resembles those of proteinaceous polymerases and self-splicing introns (which use two-metal mechanisms; **Fig. 5a**)[27,28,33,41,42] and that of the ββα-me and HUH nucleases (which use a single active-site metal cofactor)[43]. However, the manner in which the ligase replaced metal B—a universal feature of the active sites of natural polymerases[27,33]—has no known precedent[44]. Moreover, although catalytic nucleobases function in active sites of some natural catalytic RNAs[45], the ligase C47 is the first known example of a catalytic nucleobase acting in an artificial ribozyme.

With regard to mechanisms of natural ribozymes (reviewed in ref. 44), the ligase mechanism most resembles that of the HDV ribozyme, which also uses a metal-ion cofactor to deprotonate its nucleophile and an active-site cytosine to stabilize the leaving group[26,39,40,45–47]. In HDV, however, this cytosine functions in proton transfer[26,39,44,45], using an N3 group with a near-neutral pK_a[40,47], whereas the ligase uses a cytosine N4 amine and does not perform general acid-base catalysis. Like the ligase mechanism, the GlmS ribozyme mechanism involves hydrogen bonds with both a nucleobase exocyclic amine (G57 N2) and a hydroxyl group (the glucosamine-6-phosphate C1-hydroxyl)[48,49]. However, both of these groups stabilize the scissile phosphate transition-state geometry rather than the leaving group. Although the ligase mechanism differs from that of the hairpin ribozyme in depending on an active-site metal ion[50], its extensive use of electrostatics is reminiscent of the hairpin ribozyme catalytic strategy, which accelerates self-cleavage by forming more hydrogen bonds with the scissile phosphate at the transition state than it does at either ground state[51]. A water molecule might also play a catalytic role for the hairpin, although it is proposed to function as a specific base, activating the reaction nucleophile[52].

The proposed catalytic strategy of the ligase, which makes extensive use of electrostatic catalysis in addition to a metal-ion cofactor and substrate approximation (**Fig. 5b**), is more complex than strategies proposed for less efficient artificial ribozymes that promote similar reactions[21,23], suggesting that the relatively complicated catalytic strategy of the ligase might be tied to its catalytic power[53]. Applying similar structural and biochemical analyses to other ribozymes borne from *in vitro* selection experiments will increase understanding of the differences between natural and artificial ribozymes and of RNA catalysis more generally. Furthermore, a detailed understanding of the structural contributions to substrate recognition and catalysis by the class I ligase might facilitate efforts to isolate an RNA replicase ribozyme.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Atomic coordinates and structure factors are deposited in the Protein Data Bank under accession codes 3R1H and 3R1L for the Ca^{2+}–Sr^{2+} and Mg^{2+}–Sr^{2+} preligation complexes, respectively.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

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

1. Joyce, G.F. & Orgel, L.E. Prospects for understanding the origin of the RNA world. in *The RNA World* 2nd edn. (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 49–77 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 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. III. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7**, 101–104 (1976).
4. Nissen, P., Hansen, J., Ban, N., Moore, P.B. & Steitz, T.A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
5. Fedor, M.J. & Williamson, J.R. The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* **6**, 399–412 (2005).
6. Chen, X., Li, N. & Ellington, A.D. Ribozyme catalysis of metabolism in the RNA world. *Chem. Biodivers.* **4**, 633–655 (2007).
7. Lincoln, T.A. & Joyce, G.F. Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–1232 (2009).
8. Bartel, D.P. Re-creating an RNA replicase. in *The RNA World* 2nd edn. (eds. Gesteland, R.F., Cech, T.R. & Atkins, J.F.) 143–162 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 1999).
9. McGinness, K.E. & Joyce, G.F. In search of an RNA replicase ribozyme. *Chem. Biol.* **10**, 5–14 (2003).
10. Bartel, D.P. & Szostak, J.W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418 (1993).
11. Ekland, E.H., Szostak, J.W. & Bartel, D.P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364–370 (1995).
12. Ekland, E.H. & Bartel, D.P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373–376 (1996).
13. 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**, 1319–1325 (2001).
14. Lawrence, M.S. & Bartel, D.P. New ligase-derived RNA polymerase ribozymes. *RNA* **11**, 1173–1180 (2005).
15. Zaher, H.S. & Unrau, P.J. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**, 1017–1026 (2007).
16. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209–212 (2011).
17. Shechner, D.M. *et al.* Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* **326**, 1271–1275 (2009).
18. Bagby, S.C., Bergman, N.H., Shechner, D.M., Yen, C. & Bartel, D.P. A class I ligase ribozyme with reduced Mg^{2+} dependence: selection, sequence analysis, and identification of functional tertiary interactions. *RNA* **15**, 2129–2146 (2009).
19. Stuhlmann, F. & Jaschke, A. Characterization of an RNA active site: interactions between a Diels-Alderase ribozyme and its substrates and products. *J. Am. Chem. Soc.* **124**, 3238–3244 (2002).
20. Serganov, A. *et al.* Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat. Struct. Mol. Biol.* **12**, 218–224 (2005).
21. Robertson, M.P. & Scott, W.G. The structural basis of ribozyme-catalyzed RNA assembly. *Science* **315**, 1549–1553 (2007).
22. 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**, 358–361 (2008).
23. Pitt, J.N. & Ferre-D'Amare, A.R. Structure-guided engineering of the regioselectivity of RNA ligase ribozymes. *J. Am. Chem. Soc.* **131**, 3532–3540 (2009).
24. Glasner, M.E., Bergman, N.H. & Bartel, D.P. Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry* **41**, 8103–8112 (2002).
25. Harding, M.M. Geometry of metal-ligand interactions in proteins. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 401–411 (2001).
26. Chen, J.H. *et al.* A 1.9 A crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry* **49**, 6508–6518 (2010).
27. Steitz, T.A. A mechanism for all polymerases. *Nature* **391**, 231–232 (1998).
28. Castro, C. *et al.* Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* **16**, 212–218 (2009).
29. 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).
30. 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**, 857–867 (2000).
31. Glasner, M.E., Yen, C.C., Ekland, E.H. & Bartel, D.P. Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. *Biochemistry* **39**, 15556–15562 (2000).
32. Basu, S. & Strobel, S.A. Thiophilic metal ion rescue of phosphorothioate interference within the Tetrahymena ribozyme P4–P6 domain. *RNA* **5**, 1399–1407 (1999).
33. Sträter, N., Lipscomb, W.N., Klabunde, T. & Krebs, B. Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. *Angew. Chem. Int. Edn Engl.* **35**, 2024–2055 (1996).
34. Frederiksen, J.K. & Piccirilli, J.A. Identification of catalytic metal ion ligands in ribozymes. *Methods* **49**, 148–166 (2009).
35. Saenger, W. *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
36. Jencks, W.P. *Catalysis in Chemistry and Enzymology* (McGraw-Hill, New York, 1969).
37. Venkatasubban, K.S. & Schowen, R.L. The proton inventory technique. *CRC Crit. Rev. Biochem.* **17**, 1–44 (1984).
38. Schowen, K.B. & Schowen, R.L. Solvent isotope effects of enzyme systems. *Methods Enzymol.* **87**, 551–606 (1982).
39. Das, S.R. & Piccirilli, J.A. General acid catalysis by the hepatitis delta virus ribozyme. *Nat. Chem. Biol.* **1**, 45–52 (2005).
40. 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**, 13335–13342 (2007).
41. Stahley, M.R. & Strobel, S.A. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309**, 1587–1590 (2005).
42. Toor, N., Keating, K.S., Taylor, S.D. & Pyle, A.M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008).
43. Yang, W. An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* **15**, 1228–1231 (2008).
44. Fedor, M.J. Comparative enzymology and structural biology of RNA self-cleavage. *Annu. Rev. Biophys.* **38**, 271–299 (2009).
45. Bevilacqua, P.C. & Yajima, R. Nucleobase catalysis in ribozyme mechanism. *Curr. Opin. Chem. Biol.* **10**, 455–464 (2006).
46. Ke, A., Zhou, K., Ding, F., Cate, J.H. & Doudna, J.A. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **429**, 201–205 (2004).
47. Lupták, A., Ferre-D'Amare, A.R., Zhou, K., Zilm, K.W. & Doudna, J.A. Direct pK_a measurement in the active-site cytosine in a genomic hepatitis delta virus ribozyme. *J. Am. Chem. Soc.* **123**, 8447–8452 (2001).
48. Klein, D.J. & Ferre-D'Amare, A.R. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **313**, 1752–1756 (2006).
49. 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).
50. Nesbitt, S., Hegg, L.A. & Fedor, M.J. An unusual pH-independent and metal-ion-independent mechanism for hairpin ribozyme catalysis. *Chem. Biol.* **4**, 619–630 (1997).
51. Rupert, P.B., Massey, A.P., Sigurdsson, S.T. & Ferre-D'Amare, A.R. Transition state stabilization by a catalytic RNA. *Science* **298**, 1421–1424 (2002).
52. 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).
53. Emilsson, G.M., Nakamura, S., Roth, A. & Breaker, R.R. Ribozyme speed limits. *RNA* **9**, 907–918 (2003).

ONLINE METHODS

Sample preparation for crystallography. The transcription template for the unreacted C47U mutant ligase, p307HU_C47U, was generated from p307HU (ref. 17) and linearized by EarI digestion before use. RNA was synthesized by T7 *in vitro* transcription, processed and purified as described previously[17], but without removal of the terminal 2′-3′ cyclic phosphate. The ligase substrate oligonucleotide was chemically synthesized (Dharmacon) and gel-purified. The U1A A1-98 Y31H Q36R double mutant[54] was expressed and purified as described[17].

Crystallization and soaking. Approximately 200 μM unreacted C47U ligase RNA was heated and cooled 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). Sample preparation then proceeded as described[17]. Crystal clusters were grown at 22 °C from 1-μl hanging drops containing 16–20% (v/v) 2-methyl-2,4-pentanediol (MPD), 50 mM sodium cacodylate (pH 6.0), 20 mM calcium acetate, 10 mM strontium acetate and 1 mM spermine. To generate microseed stocks, these clusters were stabilized by bringing the drop MPD concentration to 30% (v/v), crushed with Seed-Beads (Hampton Research) and serially diluted with stabilization buffer. Diffraction-quality crystals were obtained by mixing these seed stocks with ligase–substrate–U1A samples that had been annealed in the presence of 100 mM KCl and then equilibrating over wells containing 20–24% (v/v) MPD. Crystals were cryostabilized by bringing the drop MPD concentration to 30% (v/v) under otherwise isotonic conditions. Stabilized crystals were either mounted in nylon loops and plunged into liquid nitrogen or equilibrated with magnesium before freezing. To equilibrate them, stabilized drops were serially supplemented by carrying out eight two-fold additions of an isotonic solution in which Mg^{2+} replaced Ca^{2+}, and after each addition incubating for 10 min and then withdrawing liquid so as to bring the volume back to 4 μl. After a final 2-h incubation, crystals were frozen as above. Because in the presence of Mg^{2+} the ligation rate of the C47U mutant is ~1.1×10^{-5} min^{-1} at pH 6.0, between 0.1–0.3% presumably ligated during equilibration with Mg^{2+}.

Data collection, structure solution and refinement. Data were collected at Northeastern Collaborative Access Team (NE-CAT) beamlines 24-ID-C and 24-ID-E at the Advanced Photon Source (APS) and were indexed, scaled and integrated using the HKL2000 software suite[55]. Each structure was solved independently; each monomer was built separately. Initial phases were obtained by molecular replacement in PHASER[56], using as a search model the U1A-bound class I ligase product structure (PDB ID: 3HHN), in which residues abutting the active site had been removed. The optimal solutions for each search (z scores of 26–30) were used as the starting points for further refinement in PHENIX[57]. Model building was done in COOT[58]. Refinement incorporated rounds of rigid-body and individual ADP strategies, interspersed with rounds of simulated annealing. Active-site atoms were built iteratively, guided by inspection of $|F_o| - |F_c|$ difference maps and the reduction of R_{free} values. Residues abutting GTP1 were built first, followed by GTP1, its triphosphate, hydrated metal clusters and finally individual solvent atoms. The final refined Ca^{2+}–Sr^{2+} complex and Mg^{2+}–Sr^{2+} complex models have estimated coordinate errors (maximum likelihood[57]) of 0.47 Å and 0.40 Å, respectively. Structural figures were made in PyMol (http://www.pymol.org/, Schrödinger LLC). An additional description of model building and refinement is provided (**Supplementary Methods**).

Sample preparation for biochemistry. Ligase constructs bearing modified nucleotides were assembled using DNA-splinted RNA ligation[59]. To enable modular manipulation of all positions near the active site, each ligase (including the unmodified ribozyme) was constructed from five oligonucleotides: *in vitro*–transcribed 5′ and 3′ fragments spanning residues 1–23 and 52–121 (including the P5 U1A loop), respectively; a synthetic decamer spanning residues 24–33; and synthetic nonamers spanning residues 34–42 and 43–51. The synthetic oligonucleotides were synthesized with 5′-monophosphates; the 3′ fragment was processed with a hammerhead ribozyme to yield a 5′-hydroxyl and phosphorylated with T4 polynucleotide kinase (New England Biolabs). Each assembly was a single-pot ligation reaction using T4 RNA ligase 2 (New England Biolabs) guided by a DNA that paired to residues 13–61. More information on reactions and oligonucleotide synthesis is provided (**Supplementary Methods**).

Ribozyme kinetics. Self-ligation and primer-extension assays were done under single-turnover conditions at 22 °C, as described[24,31], with final concentrations of 1 μM ribozyme, 0.5 μM oligonucleotide substrate supplemented with trace [5′-^{32}P]radiolabeled substrate and 10 mM $MgCl_2$. In primer-extension assays, 1 mM GTP (or analog) was added with an additional 1 mM $MgCl_2$. Reaction time points were quenched in EDTA and separated on denaturing urea polyacrylamide gels. Gels were quantified by phosphorimaging (Fujifilm BAS-2500). Data were fit using the program KaleidaGraph (Synergy Software), to a single exponential equation

$$F_p(t) = F_M(1 - e^{-k_{obs}t})$$

where $F_p(t)$ is the fraction of product at time t, and F_M (the maximum fraction reacted) and k_{obs} (the observed rate constant) are treated as unknowns. All kinetic experiments were done in triplicate; reported errors are s.d. In experiments calculating relative rates, uncertainties were propagated as follows

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

where A and B are individual rate constants, R is their ratio, and σ_A, σ_B and σ_R are their associated uncertainties. An additional description of kinetic assays is provided (**Supplementary Methods**).

Phosphorothioate interference and metal rescue experiments. Interference values were calculated from the apparent first-order rate constants for modified and unmodified constructs in Mg^{2+} alone

$$\text{Interference} = \frac{^{Mg}k_{thio}}{^{Mg}k_{oxy}}$$

Rescue values were calculated from interference values in the presence and absence of rescuing metal X

$$\text{Rescue} = \frac{\dfrac{^{X}k_{thio}}{^{X}k_{oxy}}}{\dfrac{^{Mg}k_{thio}}{^{Mg}k_{oxy}}}$$

Metal-rescue experiments were attempted at several $MgCl_2$ concentrations, titrating $CdCl_2$, $CoCl_2$, $MnCl_2$ and $ZnCl_2$ as potential rescuing metals. The most substantial rescue was observed when 50 mM $MgCl_2$ was supplemented with 2 mM $CdCl_2$. All interference and rescue values (**Fig. 2e**) were determined under these conditions using the self-ligation assay, except effects of modifications at the GTP1 α- and β-phosphates, which were determined under these conditions using primer-extension assays (**Fig. 1a**, inset)[31]. Values for GTPγS were measured using both assays.

Proton inventories. All reaction components were reconstituted in both D_2O and H_2O, and self-ligation reactions were measured under varying molar ratios of the two solvents. Data were fit to the modified Gross-Butler equations[37]

$$\frac{k_n}{k_{H_2O}} = \prod_i (1 - n + n \cdot \Phi_i)$$

where k_n is the observed rate constant at n molar fraction of D_2O, k_{H_2O} is the rate constant in pure water, each Φ_i is the inverse of the SKIE for an individual ionizable group and i is 1 or 2.

54. 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**, 541–556 (2000).
55. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
56. McCoy, A.J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
57. Adams, P.D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
58. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
59. Bullard, D.R. & Bowater, R.P. Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Biochem. J.* **398**, 135–144 (2006).