



# HHS Public Access

Author manuscript

*Nat Struct Mol Biol.* Author manuscript; available in PMC 2012 March 01.

Published in final edited form as:

*Nat Struct Mol Biol.* ; 18(9): 1036–1042. doi:10.1038/nsmb.2107.

## The structural basis of RNA-catalyzed RNA polymerization

David M. Shechner<sup>1,2,3</sup> and David P. Bartel<sup>1,2</sup>

<sup>1</sup>Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA.

<sup>2</sup>Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

### Abstract

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 pre-ligation 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 employ complex catalytic strategies that differ dramatically 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<sup>1</sup>. 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<sup>1–4</sup>. Additional support has come through the study of ribozymes. Although known natural ribozymes catalyze only the cleavage or rearrangement of phosphodiester bonds<sup>5</sup> and the formation of peptide bonds<sup>4</sup>, the diversity of artificial ribozymes produced from *in vitro* selection experiments has revealed the catalytic potential of RNA to be quite broad<sup>6,7</sup>. Of

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

Correspondence and requests for materials should be addressed to D.P.B. (dbartel@wi.mit.edu).

<sup>3</sup>Present address: Department of Stem Cell and Regenerative Biology, Harvard University, Bauer Laboratories, 7 Divinity Avenue, Cambridge, MA 02138, USA and The Broad Institute, Cambridge, MA 02142, USA.

#### Accession Codes

Atomic coordinates and structure factors are deposited in the Protein Data Bank under accession codes 3R1H and 3R1L for the Ca<sup>2+</sup>-Sr<sup>2+</sup>- and Mg<sup>2+</sup>-Sr<sup>2+</sup>-preligation complexes, respectively.

#### Supplementary Information

Supplementary figures 1–5, Supplementary discussion, Supplementary methods, and Supplementary movie 1 are available.

#### Author Contributions

D.S. and D.B. designed the experiments and wrote the manuscript. D.S. performed the experiments.

#### Author Information

The authors declare no competing financial interests.

particular relevance to the RNA world hypothesis would be ribozymes with the RNA-dependent RNA polymerization activity needed for general RNA replication<sup>8,9</sup>.

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<sup>10,11</sup>. 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<sup>10,11</sup>. Indeed, ligase variants have been engineered to catalyze primer-extension using exogenous nucleoside triphosphates (NTPs) and an internal template (Fig. 1a, inset)<sup>12</sup>. 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<sup>13-16</sup>. 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)<sup>13,15,16</sup>. Optimized versions of these polymerase ribozymes can extend a primer by up to 22 nucleotides<sup>15</sup>, or up to 95 nucleotides when appending to the ribozyme a segment that pairs to a short region of the template<sup>16</sup>.

We recently reported the crystal structure of the class I ligase self-ligation product<sup>17,18</sup>. 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 employs a catalytic strategy both more complex than those of other artificial ribozymes<sup>19-23</sup> 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<sup>18</sup>, to visualize an intact, catalytically viable active site, we employed two modifications that each dramatically 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  $\text{Ca}^{2+}$  and  $\text{Sr}^{2+}$ , 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<sup>24</sup>.  $\text{Mg}^{2+}$ -bound crystals were obtained by soaking these crystals (hereafter, “ $\text{Ca}^{2+}$ - $\text{Sr}^{2+}$ -complex” or “ $\text{Ca}^{2+}$ - $\text{Sr}^{2+}$ -preligation” crystals) into isotonic conditions containing  $\text{Mg}^{2+}$  (hereafter, the “ $\text{Mg}^{2+}$ - $\text{Sr}^{2+}$ -complex” or “ $\text{Mg}^{2+}$ - $\text{Sr}^{2+}$ -preligation” crystals, Table 1). Crystals soaked without  $\text{Sr}^{2+}$ , and those soaked with other  $\text{Mg}^{2+}$  isosteres (such as  $\text{Mn}^{2+}$  or  $\text{Yb}^{3+}$ ) proved intractable. The structure of each of these

crystals was solved to 3.15 Å, allowing visualization of the ligase poised prior to catalysis (Fig. 1b, Supplementary Fig. 1a and Supplementary Methods).

The global conformation of the ligase was remarkably similar in the pre- and post-catalytic states, being more dramatically altered by binding different metals than it was by undergoing ligation (Fig. 1c,d and Supplementary Fig. 1b,c). The Mg<sup>2+</sup>-Sr<sup>2+</sup>-complex and product structures superposed with an all-atom root mean square deviation (RMSD) of 0.31 Å, less than the mean precision of either structure (ref. 17), whereas the RMSD between the Ca<sup>2+</sup>-Sr<sup>2+</sup>-complex and product structures was 0.76 Å. In the latter case, the majority of the structural perturbations fell 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<sup>2+</sup>, which is ~0.3 Å larger than Mg<sup>2+</sup> (ref. 25). All of these structural perturbations were alleviated in the active site of the Mg<sup>2+</sup>-Sr<sup>2+</sup> complex (Fig. 1d). The close resemblance between the active site of the Mg<sup>2+</sup>-Sr<sup>2+</sup> complex and that of the product implied that, unlike the analogous variant of the hepatitis delta virus (HDV) ribozyme<sup>26</sup>, the diminished activity of the C47U mutant was not caused by a gross structural perturbation. Furthermore, as expected from its minor effect on catalysis<sup>24</sup>, the presence of Sr<sup>2+</sup> had no detectable effect on the ribozyme structure, and no peaks corresponding to Sr<sup>2+</sup> 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 dramatically 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<sup>27,28</sup>. Moreover, the (d)NTPs in these enzymes are typically Mg<sup>2+</sup>-bound, with the metal ion coordinated by two or more phosphate groups<sup>29</sup>. Although the ligase 5'-triphosphate makes no such polydentate contacts, the 5'-γ-phosphate coordinates a Mg<sup>2+</sup>•4H<sub>2</sub>O 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<sup>2+</sup>, 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 towards potentially stabilizing functional groups in the active site. In Ca<sup>2+</sup>, which greatly slows catalysis, this favorable conformation is perturbed. Because the Ca<sup>2+</sup>•hydrate cluster near G2 doesn't support close approach of the γ-phosphate and the 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<sup>17,18</sup> 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<sup>17</sup>. This metal was observed in the preligation complexes: during crystallographic refinement,  $|F_o| - |F_c|$  difference Fourier maps prominently featured  $7.8\sigma$  and  $4.8\sigma$  positive peaks centered within the active sites of the  $\text{Ca}^{2+}\text{-Sr}^{2+}$  and  $\text{Mg}^{2+}\text{-Sr}^{2+}$  complexes, respectively (Fig. 2a,b). In the  $\text{Ca}^{2+}\text{-Sr}^{2+}$  structure, this peak perfectly accommodated a partially hydrated  $\text{Ca}^{2+}$  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  $\text{Mg}^{2+}\text{-Sr}^{2+}$  maps represented a less ideal  $\text{Mg}^{2+}$  binding site<sup>30</sup>, in which the active-site  $\text{Mg}^{2+}$  center was  $\sim 2.5\text{-}2.7$  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 to the 5'- $\alpha$ -phosphate nonbridging oxygens, even though the ribozyme is substantially inhibited by sulfur substitution of the 5'- $\alpha$ -phosphate pro-*Rp* oxygen, implying that ligase employs a direct inner-sphere contact to this group<sup>31</sup>. The discrepancy between the biochemical and crystallographic data might be explained if the phosphorothioate substitution perturbed an outer-sphere contact<sup>32</sup>, which could be easily accommodated in either structure. Alternatively, since formation of a more ideal active-site  $\text{Mg}^{2+}$ -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  $\text{Mg}^{2+}$  (2.45 Å)<sup>30</sup> leave open the possibility of direct coordination of the  $\alpha$ -phosphate pro-*Rp* oxygen in the  $\text{Mg}^{2+}\text{-Sr}^{2+}$  complex.

Natural RNA and DNA polymerases employ a two-metal mechanism in which Metal A activates the 3'-hydroxyl nucleophile for attack, Metal B (aided by a general acid<sup>28</sup>) facilitates loss of the pyrophosphate leaving group, and both metals stabilize the geometry and charge of  $\alpha$ -phosphate during the transition state<sup>27,33</sup>. If the ligase transition state adopts a more ideal  $\text{Mg}^{2+}$ -binding site—and one that supports an inner-sphere contact with the  $\alpha$ -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  $\beta$  phosphate during the transition state (Fig 2a,b,d). These include the partially hydrated  $\text{Mg}^{2+}$  cluster bound to the Hoogsteen faces of GTP1 and G2, a  $\text{Mg}^{2+}$  ion jointly coordinated by the GTP1  $\gamma$ -phosphate and the U8 Pro-*Sp*-nonbridging oxygen, and a water molecule bound by the GTP1 5' and  $[\beta,\gamma]$  bridging oxygens (Fig. 2b,d). An additional partially hydrated  $\text{Mg}^{2+}$  ion near the A31 Pro-*Sp*-nonbridging oxygen bridges the gap between the P6 A-minor triad<sup>17</sup> and the 5'-triphosphate, potentially placing a water molecule near the  $\alpha$ -Pro-*Rp*-nonbridging phosphate oxygens. Also of potential catalytic relevance, we observed evidence for a partially-hydrated  $\text{Mg}^{2+}$  near the  $\beta$ -phosphate, which would place an ordered water molecule within hydrogen-bonding distance of the triphosphate  $[\alpha,\beta]$ -bridging oxygen, the reaction leaving group (Fig. 2b,d).

To probe for functional metal-ion interactions at the active site, we performed site-directed sulfur substitutions and thiophilic metal-rescue experiments<sup>34</sup>. We surveyed all nonbridging phosphate oxygens at positions 1, 29 and 30, using GTP analogues in primer-extension reactions (Fig. 1a, inset)<sup>12</sup> to survey the oxygens at position 1 and using splint-ligation products to survey oxygens within the ribozyme active site (Fig. 2e and Supplementary Fig. 2). Sulfur substitutions at the G1- $\alpha$ -Pro-*Rp*, or the A29 and C30-pro-*Rp* oxygens inhibited catalysis by factors of  $10^3$ – $10^4$ , consistent with previous results<sup>17,31</sup>. Of these groups, the C30-pro-*Rp* substitution was substantially rescued by adding a thiophilic metal (30-fold rescue in  $\text{Cd}^{2+}$ ), which 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- $\alpha$ -pro-*Sp* site, where phosphorothioate inhibition and metal rescue were modest (inhibition to 0.053 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  $\beta$  or  $\gamma$  oxygens plays 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

Based on their proximity to the 5'-triphosphate, two additional functional groups seemed capable of direct catalytic roles (Fig. 3a). In the  $\text{Mg}^{2+}$ – $\text{Sr}^{2+}$  complex, the U47 O4 group (mutated from the C47 N4 exocyclic amine to slow ligation) is 3.7 Å from the 5'- $\alpha$ -phosphate pro-*Rp* oxygen, a distance that decreases to 3.1 Å between the C47 N4 and G1 pro-*Sp* oxygen in the product structure (Supplementary Movie 1). In each structure, the C30 2'-hydroxyl (owing to the unusual 3'-exo ribose conformation<sup>35</sup> of C30, a consequence of the P6 A-minor triad<sup>17</sup>) is within hydrogen-bonding distance of U47 O4 (C47 N4 in the parent). In the  $\text{Mg}^{2+}$ – $\text{Sr}^{2+}$  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  $10^{-6}$ – $10^{-4}$  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 dramatically affected, with rates 0.0024–0.09 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 simply losing the amine. Furthermore, based on the loss of activity 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<sup>36</sup>. 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 exhibited log-linear pH dependence with a slope of 1.0 (Supplementary Fig. 3a), identical to the behavior observed for the parental ligase<sup>24</sup>. 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 performed proton inventory experiments<sup>28,37</sup> 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<sub>2</sub>O fit well to a line with a slope ( $\Phi$ ) of  $0.19 \pm 0.01$  (Fig. 4a, left), which corresponded to a solvent deuterium kinetic isotope effect (SDKIE)<sup>37</sup> of  $5.0 \pm 0.3$ , comparable to those observed for some natural polymerases<sup>28,37</sup>. 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)<sup>38</sup> and yielded SDKIE values of  $4.1 \pm 0.4$  and  $1.15 \pm 0.1$  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  $2.4 \pm 0.1$  and  $2.9 \pm 0.1$ , 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<sup>37</sup>. Because a linear proton-inventory fit can have alternative interpretations<sup>37,38</sup>, 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<sup>26,39,40</sup>.

### 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. 3a, right). This construct also exhibited a linear proton inventory with a measured SDKIE of  $3.1 \pm 0.2$  (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 exhibited 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, which replaced the N4 amine with 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. 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 pre-catalytic state and is one of three such ribozymes for which the active site has been biochemically interrogated in detail<sup>19,20,23</sup>. Our results indicate a model for transition-state stabilization in which the A29 and C30 pro-*Rp* phosphate oxygens coordinate a catalytic Mg<sup>2+</sup> 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'- $\alpha$ -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  $[\alpha,\beta]$ -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. 3a). In contrast, removal of a general-acid catalyst from the active sites of proteinaceous polymerases changes both of these properties<sup>28</sup>. Our model proposes that a network of hydrogen bonds confers 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. 5, left)<sup>27,28,33,41,42</sup> and that of the  $\beta\beta\alpha$ -me and HUH nucleases (which use a single active-site metal cofactor)<sup>43</sup>. However, the manner in which the ligase replaced Metal B—a universal feature of the active sites of natural polymerases<sup>27,33</sup>—has no known precedent<sup>44</sup>. Moreover, although catalytic nucleobases function in active sites of some natural catalytic RNAs<sup>45</sup>, 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<sup>26,39,40,45–47</sup>. In HDV, however, this cytosine functions in proton transfer<sup>26,39,44,45</sup>, utilizing an N3 group with a near-neutral pKa<sup>40,47</sup>, whereas the ligase uses a cytosine N4 amine and does not perform general acid-base catalysis. Like the ligase mechanism, the GlnS ribozyme mechanism involves hydrogen bonds with both a nucleobase exocyclic amine (G57 N2) and a hydroxyl group (the glucosamine-6-phosphate C1-hydroxyl)<sup>48,49</sup>. 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<sup>50</sup>, 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<sup>51</sup>. 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<sup>52</sup>.

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. 5, right), is more complex than strategies proposed for less efficient artificial ribozymes that promote similar reactions<sup>21,23</sup>, suggesting that the relatively complicated catalytic strategy of the ligase might be tied to its catalytic power<sup>53</sup>. 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

### Sample preparation for crystallography

The transcription template for the unreacted C47U mutant ligase, p307HU\_C37U, 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<sup>17</sup>, 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<sup>54</sup> was expressed and purified as described<sup>17</sup>.

### Crystallization and soaking

Approximately 200  $\mu\text{M}$  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  $\text{CaCl}_2$ , 1 mM DTT). Sample preparation then proceeded as described<sup>17</sup>. Crystal clusters were grown at 22°C from 1  $\mu\text{L}$  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 1mM 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 equilibrating over wells containing 20–24% (v/v) MPD. Crystals were cryo-stabilized 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 prior to freezing. To equilibrate, stabilized drops were serially supplemented with eight two-fold additions of an isotonic solution in which  $\text{Mg}^{2+}$  replaced  $\text{Ca}^{2+}$ , after each addition incubating for 10 minutes and then withdrawing liquid so as to bring the volume back to 4  $\mu\text{L}$ . Following a final two-hour incubation, crystals were frozen as above. Because the ligation rate of the C47U mutant is  $\sim 1.1 \times 10^{-5} \text{ min}^{-1}$  at pH 6.0, between 0.1–0.3% presumably ligated during equilibration with  $\text{Mg}^{2+}$ .

### Data collection, structure solution and refinement

Data were collected at New England 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<sup>55</sup>. Each structure was solved independently; each monomer was built separately. Initial phases were obtained by molecular replacement in PHASER<sup>56</sup>, 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<sup>57</sup>. Model building was in COOT<sup>58</sup>. 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  $\text{Ca}^{2+}$ – $\text{Sr}^{2+}$ -complex and  $\text{Mg}^{2+}$ – $\text{Sr}^{2+}$ -complex models have estimated coordinate errors (maximum likelihood<sup>57</sup>) of 0.47 Å and 0.40 Å, respectively. Structural figures were made in PyMol (<http://www.pymol.org>, Schrödinger LLC). Additional description of model building and refinement are provided (Supplementary Methods).

### Sample preparation for biochemistry

Ligase constructs bearing modified nucleotides were assembled using DNA-splinted RNA ligation<sup>60</sup>. 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 10-mer spanning residues 24–33, and synthetic 9-mers 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 performed under single-turnover conditions at 22°C, as described<sup>24,31</sup>, with final concentrations of 1 μM ribozyme, 0.5 μM oligonucleotide substrate supplemented with trace 5'-<sup>32</sup>P-radiolabeled substrate, and 10 mM MgCl<sub>2</sub>. In primer-extension assays, 1 mM GTP (or analog) was added with an additional 1 mM MgCl<sub>2</sub>. Reaction timepoints 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 performed in triplicate; reported errors are standard deviations. 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  $\sigma_A$ ,  $\sigma_B$  and  $\sigma_R$  are their associated uncertainties. 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<sup>2+</sup> alone:

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

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

$$Rescue = \frac{\left(\frac{X k_{thio}}{X k_{oxy}}\right)}{\left(\frac{Mg k_{thio}}{Mg k_{oxy}}\right)}$$

Metal-rescue experiments were attempted at several MgCl<sub>2</sub> concentrations, titrating CdCl<sub>2</sub>, CoCl<sub>2</sub>, MnCl<sub>2</sub> and ZnCl<sub>2</sub> as potential rescuing metals. The most substantial rescue was observed when 50 mM MgCl<sub>2</sub> was supplemented with 2 mM CdCl<sub>2</sub>. 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)<sup>31</sup>. Values for GTPγS were measured using both assays.

### Proton Inventories

All reaction components were reconstituted in both D<sub>2</sub>O and H<sub>2</sub>O, and self-ligation reactions were measured under varying molar ratios of the two solvents. Data were fit to the modified Gross-Butler equations<sup>37</sup>

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

where  $k_n$  is the observed rate constant at  $n$  molar fraction of D<sub>2</sub>O,  $k_{H_2O}$  is the rate constant in pure water, each  $\Phi_i$  is the inverse of the of the SKIE for an individual ionizable group, and  $i$  is one or two.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgements

We thank F. Eckstein for the gift of GTP analogs, R.A. Grant, D. Lim, T. Schwartz and K.R. Rajashankar for assistance with data collection and processing, H. Mackie, E. Roesch and J. De Luca for advice on oligonucleotide synthesis, J. Chen, A. Ricardo, K. Frederick, N. Yoder and W. Johnston for help with HPLC purification, E. Spooner for performing mass spectrometry, U. RajBhandary, C. Drennan, J. Piccirilli, J. Szostak and members of the Bartel laboratory for helpful discussions. Supported by National Institutes of Health (NIH) grant GM061835 to D.B. This 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 U.S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357.

### References

1. Joyce, GF.; Orgel, LE. Prospects for Understanding the Origin of the RNA World. In: Gesteland, RF.; Cech, TR.; Atkins, JF., editors. *The RNA World*, Second Edition. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1999. p. 49-77.

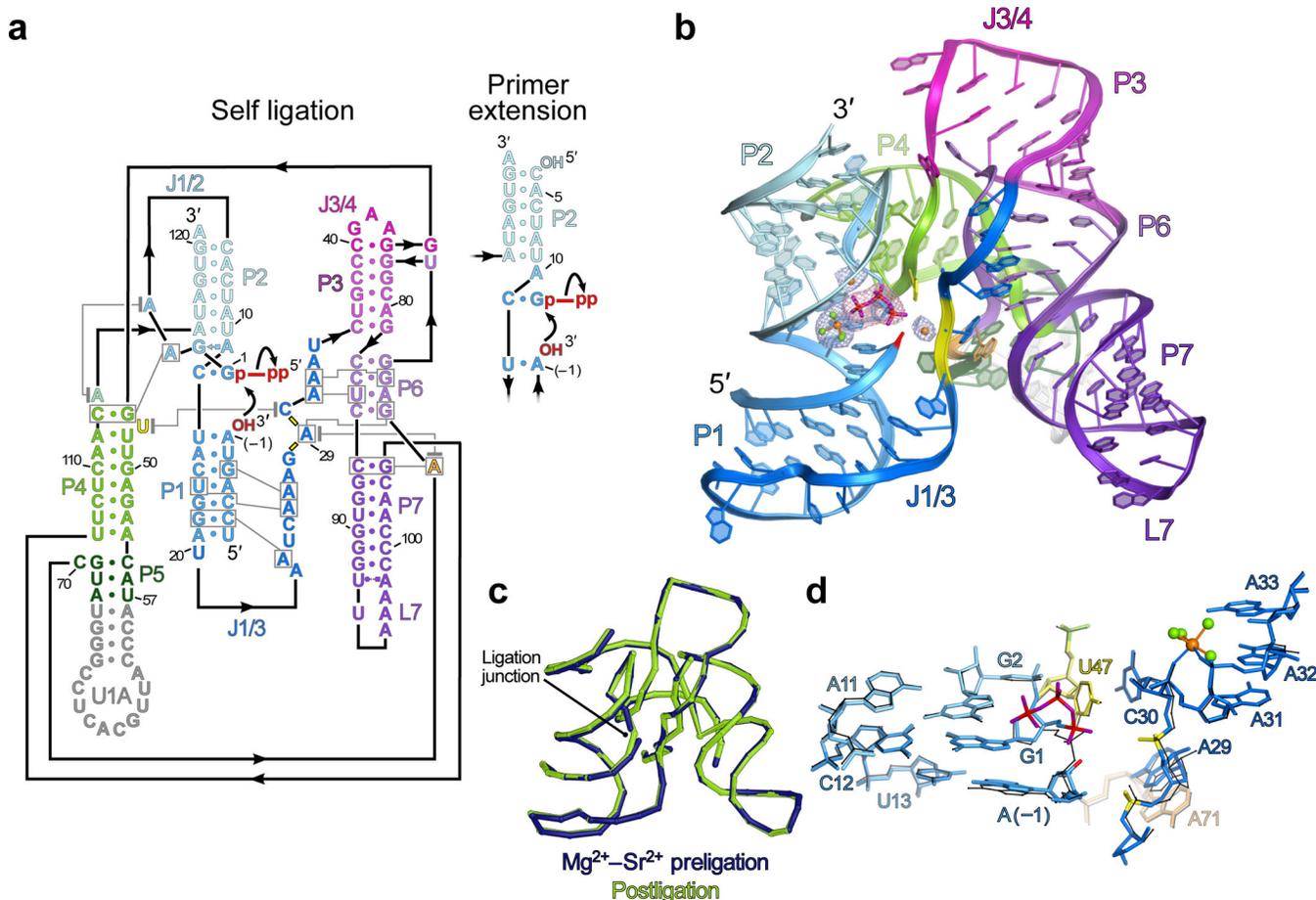
2. Orgel LE. Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol.* 2004; 39:99–123. [PubMed: 15217990]
3. White HB 3rd. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol.* 1976; 7:101–104. [PubMed: 1263263]
4. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. *Science.* 2000; 289:920–930. [PubMed: 10937990]
5. Fedor MJ, Williamson JR. The catalytic diversity of RNAs. *Nat Rev Mol Cell Biol.* 2005; 6:399–412. [PubMed: 15956979]
6. Chen X, Li N, Ellington AD. Ribozyme catalysis of metabolism in the RNA world. *Chem Biodivers.* 2007; 4:633–655. [PubMed: 17443876]
7. Lincoln TA, Joyce GF. Self-sustained replication of an RNA enzyme. *Science.* 2009; 323:1229–1232. [PubMed: 19131595]
8. Bartel, DP. Re-creating an RNA Replicase. In: Gesteland, RF.; Cech, TR.; Atkins, JF., editors. *The RNA World, Second Edition.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1999. p. 143-162.
9. McGinness KE, Joyce GF. In search of an RNA replicase ribozyme. *Chem Biol.* 2003; 10:5–14. [PubMed: 12573693]
10. Bartel DP, Szostak JW. Isolation of new ribozymes from a large pool of random sequences. *Science.* 1993; 261:1411–1418. [PubMed: 7690155]
11. Eklund EH, Szostak JW, Bartel DP. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science.* 1995; 269:364–370. [PubMed: 7618102]
12. Eklund EH, Bartel DP. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature.* 1996; 382:373–376. [PubMed: 8684470]
13. Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science.* 2001; 292:1319–1325. [PubMed: 11358999]
14. Lawrence MS, Bartel DP. New ligase-derived RNA polymerase ribozymes. *RNA.* 2005; 11:1173–1180. [PubMed: 15987804]
15. Zaher HS, Unrau PJ. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA.* 2007; 13:1017–1026. [PubMed: 17586759]
16. Wochner A, Attwater J, Coulson A, Holliger P. Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science.* 2011; 332:209–212. [PubMed: 21474753]
17. Shechner DM, et al. Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science.* 2009; 326:1271–1275. [PubMed: 19965478]
18. Bagby SC, Bergman NH, Shechner DM, Yen C, Bartel DP. A class I ligase ribozyme with reduced Mg<sup>2+</sup> dependence: Selection, sequence analysis, and identification of functional tertiary interactions. *RNA.* 2009; 15:2129–2146. [PubMed: 19946040]
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.* 2002; 124:3238–3244. [PubMed: 11916406]
20. Serganov A, et al. Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat Struct Mol Biol.* 2005; 12:218–224. [PubMed: 15723077]
21. Robertson MP, Scott WG. The structural basis of ribozyme-catalyzed RNA assembly. *Science.* 2007; 315:1549–1553. [PubMed: 17363667]
22. Xiao H, Murakami H, Suga H, Ferre-D'Amare AR. Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature.* 2008; 454:358–361. [PubMed: 18548004]
23. Pitt JN, Ferre-D'Amare AR. Structure-guided engineering of the regioselectivity of RNA ligase ribozymes. *J Am Chem Soc.* 2009; 131:3532–3540. [PubMed: 19220054]
24. Glasner ME, Bergman NH, Bartel DP. Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry.* 2002; 41:8103–8112. [PubMed: 12069603]
25. Harding MM. Geometry of metal-ligand interactions in proteins. *Acta Crystallogr D Biol Crystallogr.* 2001; 57:401–411. [PubMed: 11223517]

26. Chen JH, et al. A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry*. 2010; 49:6508–6518. [PubMed: 20677830]
27. Steitz TA. A mechanism for all polymerases. *Nature*. 1998; 391:231–232. [PubMed: 9440683]
28. Castro C, et al. Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol*. 2009; 16:212–218. [PubMed: 19151724]
29. Hsiao, C., et al. Complexes of Nucleic Acids with Group I and II Cations. In: Hud, NV., editor. *Nucleic Acid-Metal Ion Interactions*. Vol. Vol. 1. Cambridge, UK: RSC Publishing; 2009. p. 1-38.
30. Harding MM. 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*. 2000; 56:857–867. [PubMed: 10930832]
31. Glasner ME, Yen CC, Ekland EH, Bartel DP. Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. *Biochemistry*. 2000; 39:15556–15562. [PubMed: 11112542]
32. Basu S, Strobel SA. Thiophilic metal ion rescue of phosphorothioate interference within the Tetrahymena ribozyme P4–P6 domain. *RNA*. 1999; 5:1399–1407. [PubMed: 10580468]
33. Sträter N, Lipscomb WN, Klabunde T, Krebs B. Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions. *Angew. Chem. Int. Ed. Engl*. 1996; 35:2024–2055.
34. Frederiksen JK, Piccirilli JA. Identification of catalytic metal ion ligands in ribozymes. *Methods*. 2009; 49:148–166. [PubMed: 19651216]
35. Saenger, W. *Principles of Nucleic Acid Structure*. New York: Springer-Verlag; 1984.
36. Jencks, WP. *Catalysis in chemistry and enzymology*. Vol. xvi. New York: McGraw-Hill; 1969. p. 644
37. Venkatasubban KS, Schowen RL. The proton inventory technique. *CRC Crit Rev Biochem*. 1984; 17:1–44. [PubMed: 6094099]
38. Schowen KB, Schowen RL. Solvent isotope effects of enzyme systems. *Methods Enzymol*. 1982; 87:551–606. [PubMed: 6294457]
39. Das SR, Piccirilli JA. General acid catalysis by the hepatitis delta virus ribozyme. *Nat Chem Biol*. 2005; 1:45–52. [PubMed: 16407993]
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*. 2007; 129:13335–13342. [PubMed: 17924627]
41. Stahley MR, Strobel SA. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science*. 2005; 309:1587–1590. [PubMed: 16141079]
42. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. *Science*. 2008; 320:77–82. [PubMed: 18388288]
43. Yang W. An equivalent metal ion in one- and two-metal-ion catalysis. *Nat Struct Mol Biol*. 2008; 15:1228–1231. [PubMed: 18953336]
44. Fedor MJ. Comparative enzymology and structural biology of RNA self-cleavage. *Annu Rev Biophys*. 2009; 38:271–299. [PubMed: 19416070]
45. Bevilacqua PC, Yajima R. Nucleobase catalysis in ribozyme mechanism. *Curr Opin Chem Biol*. 2006; 10:455–464. [PubMed: 16935552]
46. Ke A, Zhou K, Ding F, Cate JH, Doudna JA. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature*. 2004; 429:201–205. [PubMed: 15141216]
47. Luptak A, Ferre-D'Amare AR, Zhou K, Zilm KW, Doudna J. Direct pK(a) measurement of the active-site cytosine in a genomic hepatitis delta virus ribozyme. *J Am Chem Soc*. 2001; 123:8447–8452. [PubMed: 11525650]
48. Klein DJ, Ferre-D'Amare AR. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science*. 2006; 313:1752–1756. [PubMed: 16990543]
49. Cochrane JC, Lipchock SV, Strobel SA. Structural investigation of the GlnS ribozyme bound to its catalytic cofactor. *Chem Biol*. 2007; 14:97–105. [PubMed: 17196404]
50. Nesbitt S, Hegg LA, Fedor MJ. An unusual pH-independent and metal-ion-independent mechanism for hairpin ribozyme catalysis. *Chem Biol*. 1997; 4:619–630. [PubMed: 9281529]

51. Rupert PB, Massey AP, Sigurdsson ST, Ferre-D'Amare AR. Transition state stabilization by a catalytic RNA. *Science*. 2002; 298:1421–1424. [PubMed: 12376595]
52. Salter J, Krucinska J, Alam S, Grum-Tokars V, Wedekind JE. Water in the active site of an all-RNA hairpin ribozyme and effects of Gua8 base variants on the geometry of phosphoryl transfer. *Biochemistry*. 2006; 45:686–700. [PubMed: 16411744]
53. Emilsson GM, Nakamura S, Roth A, Breaker RR. Ribozyme speed limits. *RNA*. 2003; 9:907–918. [PubMed: 12869701]

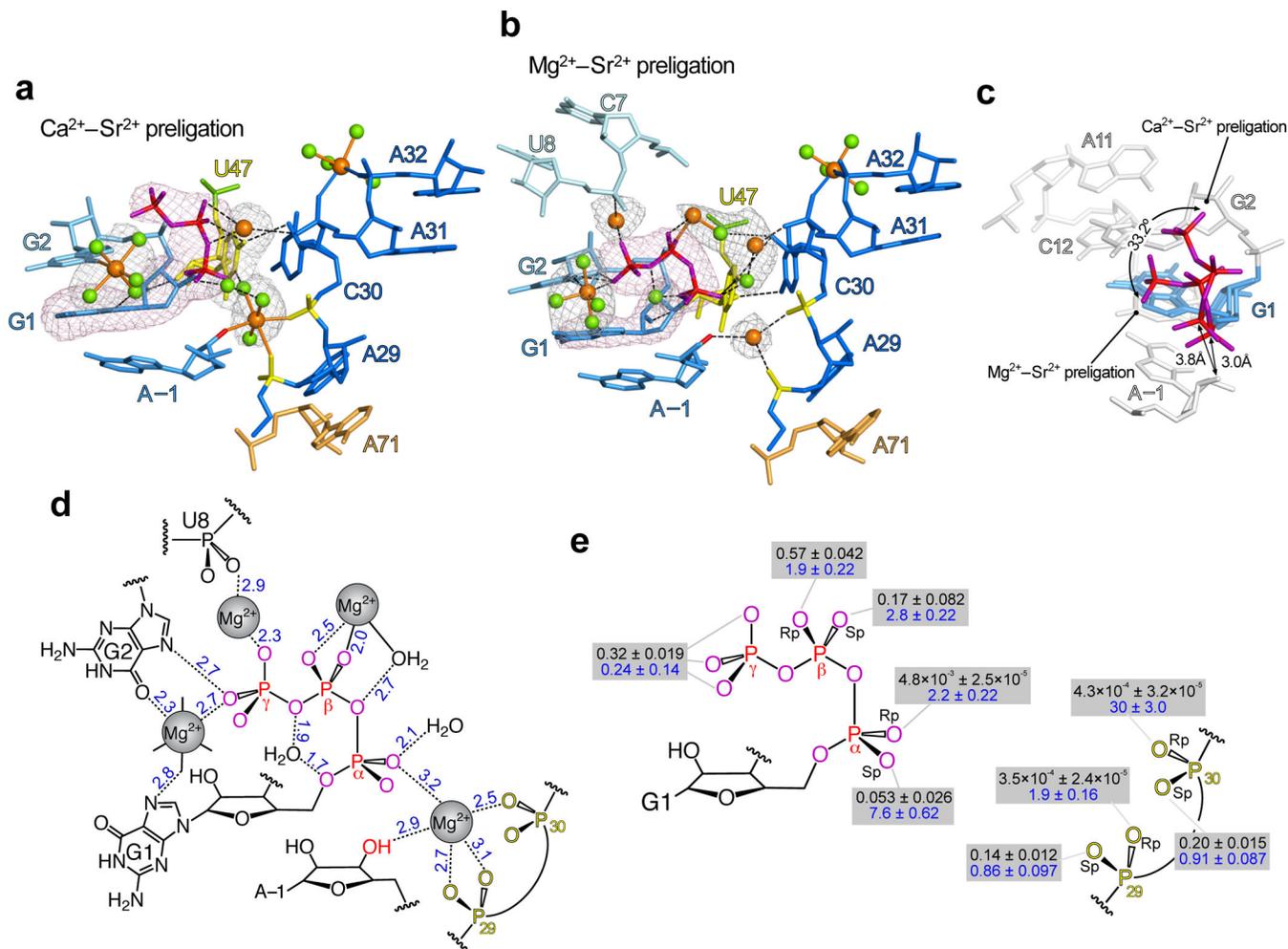
## Methods-only references

54. Ferré-D'Amaré AR, Doudna JA. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J Mol Biol*. 2000; 295:541–556. [PubMed: 10623545]
55. Otwinowski Z, Minor W. Processing of X-ray diffraction data collection in Oscillation Mode. *Methods Enzymol*. 1997; 276:307–326.
56. McCoy AJ, et al. Phaser crystallographic software. *J Appl Crystallogr*. 2007; 40:658–674. [PubMed: 19461840]
57. Adams PD, et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*. 2002; 58:1948–1954. [PubMed: 12393927]
58. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004; 60:2126–2132. [PubMed: 15572765]
59. Bullard DR, Bowater RP. Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4. *Biochem J*. 2006; 398:135–144. [PubMed: 16671895]

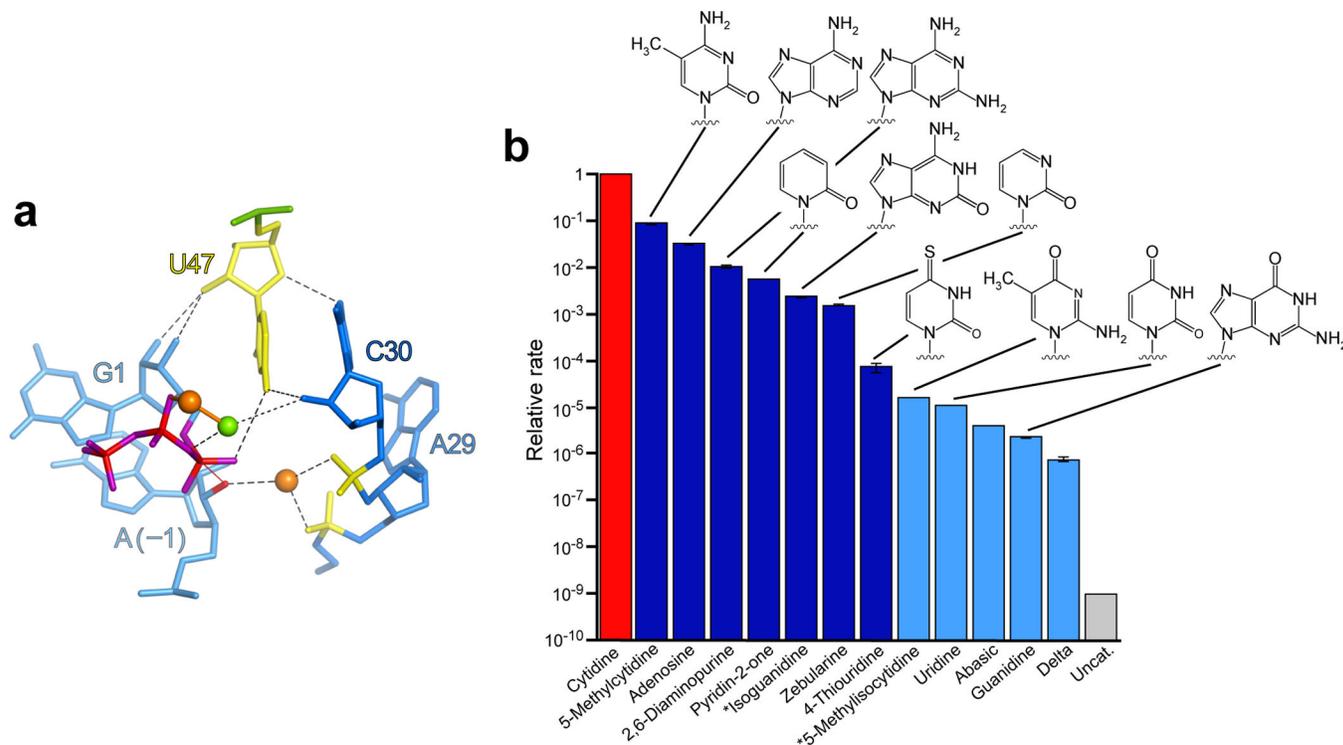


**Figure 1.**

The Class I ligase ribozyme preligation complex. **a.** Secondary structure of the C47U crystallization construct<sup>17</sup>, 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 ref. 17. Inset, primer extension catalyzed by a ligase variant resembling the self-ligation construct, except exogenous GTP replaces G1–A4. **b.** Overview of the Mg<sup>2+</sup>–Sr<sup>2+</sup>-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<sup>2+</sup>; green, water). Meshes are simulated-annealing  $|F_o| - |F_c|$  OMIT maps calculated without G1 (magenta, contoured at  $5\sigma$ ) or active-site solvent atoms (dark blue, contoured at  $4.5\sigma$ ). **c.** Superposition of Mg<sup>2+</sup>–Sr<sup>2+</sup>-preligation and product (PDB ID: 3HHN)<sup>17</sup> structures. **d.** Superposition of Mg<sup>2+</sup>–Sr<sup>2+</sup>-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.

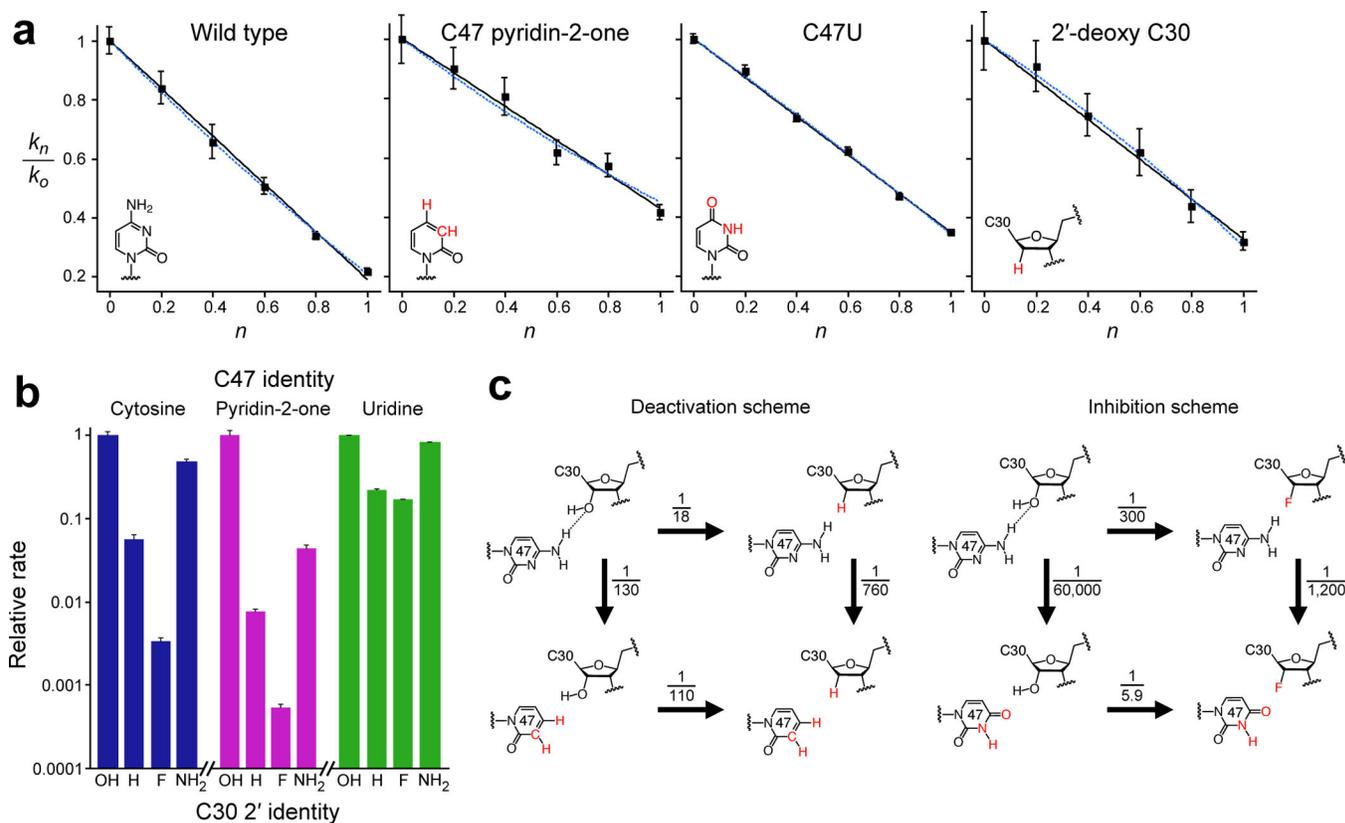
**Figure 2.**

The Class I ligase active site. **a**. The active site, as observed in the Ca<sup>2+</sup>–Sr<sup>2+</sup>–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  $5\sigma$ ) or active site solvent atoms (navy, contoured at  $4.5\sigma$ ). **b**. The active site, as observed in the Mg<sup>2+</sup>–Sr<sup>2+</sup>–preligation structure. As in **a**, except orange spheres depict Mg<sup>2+</sup> ions. **c**. Reorientation of the 5'-GTP during metal-ion exchange. **d**. Solvent interactions in active site of the Mg<sup>2+</sup>–Sr<sup>2+</sup>–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,  $0.4 \text{ \AA}$ . Waters hydrating the Mg<sup>2+</sup> bound by G1 and G2 are represented by black lines. **e**. Summary of phosphorothioate interference (black,  $Mg_{k_s}Mg_{k_{oxy}}$ ) and metal rescue [blue,  $(X_{k_s}X_{k_{oxy}})/(Mg_{k_s}Mg_{k_{oxy}})$ ] 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 standard deviations from three independent experiments.



**Figure 3.**

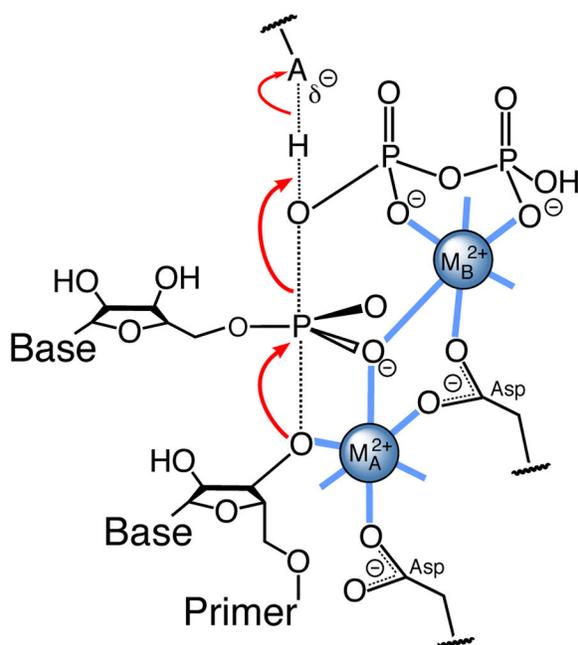
Biochemical interrogation of C47. **a.** View of the active site observed in the  $Mg^{2+}$ - $Sr^{2+}$  complex, roughly orthogonal to that in Figure 2b, highlighting interactions between C47, C30 and the 5'-triphosphate. Depiction and coloring are as in Figure 2b, with the thin red line indicating 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 hours, the rates of slower constructs (light blue bars) were measured at pH 8.0 and renormalized, assuming log-linear pH dependencies<sup>24</sup>. Constructs marked with an asterisk were synthesized with a 2'-deoxy ribose at position 47, with rates normalized to account for this modification. The 4-thiouridine variant exhibited biphasic kinetics in which approximately 10% reacted in a burst phase with a relative rate of  $0.23 \pm 0.04$  (s.d.); shown is the relative rate relevant to the majority of the molecules. Error bars, standard deviation from three independent experiments.

**Figure 4.**

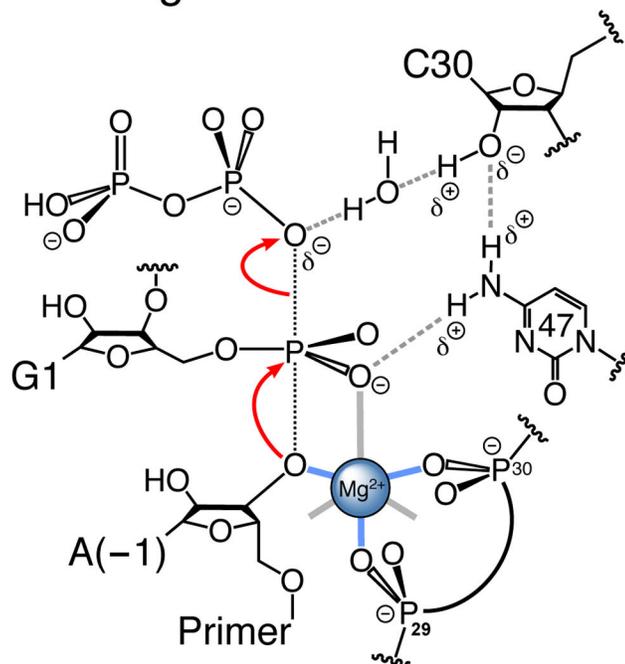
Catalytic roles of active-site functional groups. **a**. Proton inventories for self-ligation by the indicated variants. The solid black line indicates the best fit for a single-proton-transfer model; the dashed blue line indicates the best fit for a two-proton-transfer model<sup>28,37,38</sup>.

Error bars, standard deviation 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 substitutions at position 47. For each C47 substitution, rates are normalized to that of the C30 2'-OH construct. Error bars, standard deviation 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, denoted in red. 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  $0.93 \pm 0.009 \text{ min}^{-1}$  (s.d.).

## Proteins:



## Class I ligase:



**Figure 5.** Transition-state stabilization by protein and RNA active sites. (Left) Catalysis by proteinaceous polymerases<sup>28</sup>. Black dotted lines indicate bonds formed or broken during the transition state. Red arrows indicate the reaction direction during polymerization. (Right) 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.

**Table 1**

## Data Collection and refinement statistics

	Ca <sup>2+</sup> -Sr <sup>2+</sup> -Preligation Complex	Mg <sup>2+</sup> -Sr <sup>2+</sup> -Preligation Complex
<b>Data collection</b>		
Space group	P1	P1
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	58.70, 69.99, 71.86	59.19, 70.24, 71.21
$\alpha$ , $\beta$ , $\gamma$ (°)	99.86, 99.73, 103.68	99.86, 99.34, 103.81
Resolution (Å)	30–3.15 (3.26–3.15) <sup>a</sup>	40–3.15 (3.26–3.15)
<i>R</i> <sub>Merge</sub> (%)	9.8 (39.8)	13.5 (42.9)
Mean <i>I</i> / $\sigma$ ( <i>I</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)
<b>Refinement</b>		
Resolution (Å)	29.9–3.15	37.9–3.15
No. Reflections	18,190	18,739
<i>R</i> <sub>Work</sub> / <i>R</i> <sub>Free</sub> (%)	21.2/25.5	19.7/24.0
No. atoms		
RNA and Protein	7231	7254
Ligand/ion	26	46
Water	38	51
Mean <i>B</i> -factors (Å <sup>2</sup> )		
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.

<sup>a</sup>Values in parenthesis are for highest-resolution shell.