Structure and two-metal mechanism of a eukaryal nick-sealing RNA ligase

Mihaela-Carmen Unciuleac[a], Yehuda Goldgur[b], and Stewart Shuman[a,1]

[a]Molecular Biology Program, Sloan-Kettering Institute, New York, NY 10065; and [b]Structural Biology Program, Sloan-Kettering Institute, New York, NY 10065

Edited by James M. Berger, The Johns Hopkins University School of Medicine, Baltimore, MD, and approved October 1, 2015 (received for review August 19, 2015)

ATP-dependent RNA ligases are agents of RNA repair that join 3′-OH and 5′-PO_4 RNA ends. *Naegleria gruberi* RNA ligase (NgrRnl) exemplifies a family of RNA nick-sealing enzymes found in bacteria, viruses, and eukarya. Crystal structures of NgrRnl at three discrete steps along the reaction pathway—covalent ligase-(lysyl-Nζ)–AMP•Mn^{2+} intermediate; ligase•ATP•$(Mn^{2+})_2$ Michaelis complex; and ligase•Mn^{2+} complex—highlight a two-metal mechanism of nucleotidyl transfer, whereby (*i*) an enzyme-bound "catalytic" metal coordination complex lowers the pK_a of the lysine nucleophile and stabilizes the transition state of the ATP α phosphate; and (*ii*) a second metal coordination complex bridges the β- and γ-phosphates. The NgrRnl N domain is a distinctively embellished oligonucleotide-binding (OB) fold that engages the γ-phosphate and associated metal complex and orients the pyrophosphate leaving group for in-line catalysis with stereochemical inversion at the AMP phosphate. The unique domain architecture of NgrRnl fortifies the theme that RNA ligases have evolved many times, and independently, by fusions of a shared nucleotidyltransferase domain to structurally diverse flanking modules. The mechanistic insights to lysine adenylylation gained from the NgrRnl structures are likely to apply broadly to the covalent nucleotidyltransferase superfamily of RNA ligases, DNA ligases, and RNA capping enzymes.

RNA repair | covalent nucleotidyltransferase | lysyl-AMP

Biochemically diverse RNA repair systems rely on RNA ligases to maintain or manipulate RNA structure in response to purposeful RNA breakage events (1–5). RNA breaks destined for repair are inflicted by sequence-specific or structure-specific endoribonucleases during physiological RNA processing (e.g., tRNA splicing; kinetoplast mRNA editing) and under conditions of cellular stress (e.g., virus infection, starvation, unfolded protein response). RNA cleavage can occur either: (*i*) by a transesterification mechanism (generally metal-independent) that yields 2′,3′-cyclic-PO_4 and 5′-OH ends; or (*ii*) via a hydrolytic mechanism (typically metal-dependent) that leaves 3′-OH and 5′-PO_4 ends. RNA repair enzymes capable of sealing 2′,3′-cyclic-PO_4/5′-OH breaks or 3′-OH/5′-PO_4 breaks are distributed widely in all phylogenetic domains of life.

ATP-dependent RNA ligases join 3′-OH and 5′-PO_4 RNA termini via a series of three nucleotidyl transfer steps similar to those of DNA ligases (6). In step 1, RNA ligase reacts with ATP to form a covalent ligase-(lysyl-Nζ)–AMP intermediate plus pyrophosphate. In step 2, AMP is transferred from ligase-adenylate to the 5′-PO_4 RNA end to form an RNA–adenylate intermediate, AppRNA. In step 3, ligase catalyzes attack by an RNA 3′-OH on the RNA–adenylate to seal the two ends via a phosphodiester bond and release AMP.

The autoadenylylation reaction of RNA ligases is performed by a nucleotidyltransferase (NTase) domain that is conserved in ATP-dependent and NAD^+-dependent DNA ligases and GTP-dependent mRNA capping enzymes (6, 7). The NTase domain includes six peptide motifs (I, Ia, III, IIIa, IV, and V) that form the nucleotide-binding pocket (Fig. S1). Motif I contains the lysine that becomes covalently attached to the NMP. It is thought that modern RNA and DNA ligases and RNA capping enzymes evolved from an ancestral stand-alone ATP-using NTase domain via clade-specific acquisitions of flanking domain modules (7). DNA ligases and RNA capping enzymes have a shared core domain structure in which a conserved oligonucleotide-binding (OB)–fold domain is linked to the C terminus of the NTase domain (8–14). The scheme depicted in Fig. 1 posits that: (*i*) ATP-dependent DNA ligases evolved before divergence of bacteria, archaea, and eukarya; (*ii*) NAD^+-dependent ligases evolved in bacteria by acquisition of a signature N-terminal module that confers NAD^+ specificity (15, 16); and (*iii*) capping enzymes evolved uniquely in eukarya, via innovations within the NTase and OB domain that impart GTP specificity and recognition of diphosphate-terminated RNA ends.

In contrast, ATP-dependent RNA ligases appear to have evolved many times, and independently, by NTase fusions to structurally diverse C-terminal domain modules (Fig. 1). There are presently four structurally characterized RNA ligase families, exemplified by: (*i*) bacteriophage T4 RNA ligase 1 (Rnl1 family) (17); (*ii*) T4 RNA ligase 2 (Rnl2 family) (18); (*iii*) *Pyrococccus abyssi* RNA ligase (Rnl3 family) (19); and (*iv*) *Clostridium thermocellum* RNA ligase (Rnl4 family) (20, 21). The C-domain folds of these four RNA ligases are unrelated to one another and to the OB domains of DNA ligases or capping enzymes. The physiology and biochemistry of the Rnl families suggest a division of labor in RNA repair, whereby Rnl1 and Rnl4 are tailored to seal single-strand breaks in the loop of RNA stem-loops (22, 23), whereas Rnl2 is designed to seal 3′-OH/5′-PO_4 nicks in duplex RNAs and RNA:DNA hybrids (24). The specificity of T4 Rnl1 for tRNA repair is conferred by its C domain (22). T4 Rnl2 depends on its C domain for binding and sealing a duplex RNA nick (7, 25). Our hypothesis is that there are more RNA ligase families to be discovered and that new flavors of

Significance

Polynucleotide ligases are an ancient superfamily of nucleic acid repair enzymes that join 3′-OH and 5′-PO_4 DNA or RNA ends. Ligases react with ATP or NAD^+ to form a covalent enzyme–adenylate intermediate in which AMP is linked via a P–N bond to a lysine side-chain. This paper reports crystal structures of a eukaryal ATP-dependent RNA ligase (*Naegleria gruberi* RNA ligase, NgrRnl) that illuminate the stereochemistry and two-metal catalytic mechanism of the lysine adenylylation reaction. A signature N-terminal domain of NgrRnl binds the ATP γ-phosphate and orients the pyrophosphate leaving group apical to the lysine nucleophile. NgrRnl is the founder of a distinct RNA ligase clade, with homologs in diverse bacterial, viral, and eukaryal proteomes.

Author contributions: M.-C.U. and S.S. designed research; M.-C.U. and Y.G. performed research; M.-C.U., Y.G., and S.S. analyzed data; and M.-C.U., Y.G., and S.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 5COT, 5COU, and 5COV).

[1]To whom correspondence should be addressed. Email: s-shuman@ski.mskcc.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516536112/-/DCSupplemental.

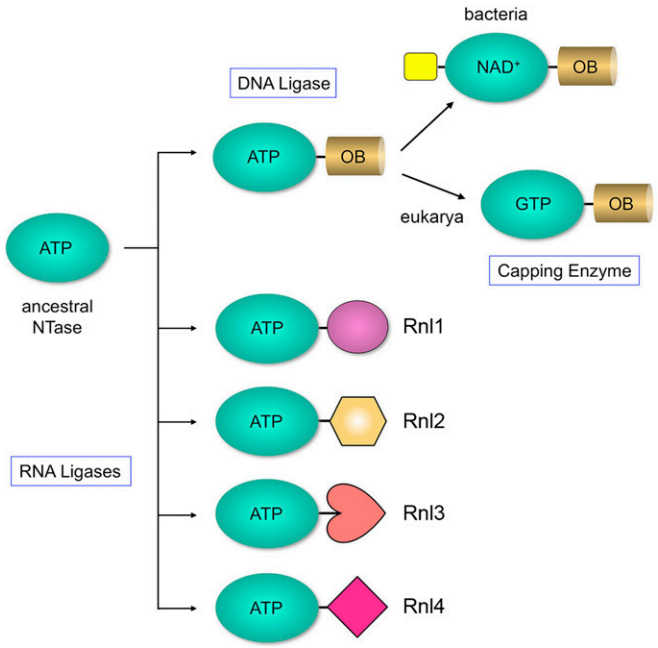

Fig. 1. Scheme of ligase evolution. RNA ligases, DNA ligases, and mRNA capping enzymes comprise a superfamily of covalent nucleotidyltransferases that act via enzyme-(lysyl-Nζ)–NMP intermediates. They are thought to descend from an ancestral ATP-using NTase domain (turquois oval) by fusions to structurally diverse flanking domain modules (depicted in various shapes and colors), as discussed in the text.

RNA ligases will inform the biochemistry, biology, and evolutionary history of RNA repair.

Deinococcus radiodurans RNA ligase (DraRnl) and *Naegleria gruberi* RNA ligase (NgrRnl) represent a recently appreciated, though structurally uncharacterized, Rnl5 family of ATP-dependent RNA ligase (26–29). DraRnl and NgrRnl are template-directed RNA ligases capable of sealing nicked duplexes in which the 3′-OH strand is RNA. DraRnl and NgrRnl can join a RNA_{OH} end to a 5′-pRNA or 5′-pDNA strand, but are unable to join when the 3′-OH strand is DNA. In this respect, they are similar to T4 Rnl2. However, unlike members of the four known RNA ligase families, DraRnl and NgrRnl lack a C-terminal appendage to their NTase domain. Instead, they contain a defining N-terminal domain that is important for overall 3′-OH/5′-PO_4 nick sealing and ligase adenylylation (step 1), but dispensable for phosphodiester synthesis at a preadenylylated nick (step 3). DraRnl and NgrRnl prefer manganese as the metal cofactor for nick sealing, although either Mg^{2+} or Mn^{2+} can support the formation of the Rnl–AMP intermediate (26, 29). Manganese exerts profound effects on *Deinococcus* resistance to high-dose ionizing radiation and oxidative stress (30, 31). The gene encoding DraRnl is transiently up-regulated during recovery of *Deinococcus* from radiation exposure (32). It was speculated that DraRnl might contribute to the extreme radiation resistance of *Deinococcus*, by either repairing broken RNAs or by sealing broken DNAs that have acquired 3′-OH RNA termini by ribonucleotide addition (perhaps as a stop-gap measure during break and gap repair) (28).

Rnl5 homologs are encoded by 35 bacterial genera, representing 10 different phyla (28). Bacteriophages that prey on *Aeromonas*, *Caulobacter*, *Mycobacterium*, and *Sinorhizobium* also encode Rnl5 homologs. An archaeal homolog is present in *Methanobrevibacter ruminantium*. We were especially intrigued by the presence of Rnl5 homologs in many unicellular eukarya, including 13 genera of fungi, the amoebae *Dictyostelium* and *Polysphondylium*, and the amoebo-flagellate *N. gruberi*. (*N. gruberi* is an avirulent cousin of the human pathogen *Naegleria fowleri*, known as the "brain-eating amoeba," which causes a devastating infection, primary amoebic meningoencephalitis.) We have purified and characterized recombinant NgrRnl, the exemplary eukaryal Rnl5 ligase (29). Here we report atomic structures of NgrRnl, captured at three discrete steps along the lysine–adenylylation reaction pathway. The structures suggest a two-metal mechanism of nucleotidyl transfer, in which a catalytic metal coordination complex lowers the pK_a of the lysine nucleophile and stabilizes the transition state of the ATP α-phosphate, and a second metal coordination complex bridges the ATP β- and γ-phosphates. The NgrRnl N domain is a uniquely embellished OB-fold that binds the ATP γ-phosphate and associated metal complex and orients the pyrophosphate leaving group for in-line catalysis.

Results

Overview of the NgrRnl Structure. Crystallization of native and SeMet-substituted NgrRnl and structure determination of the native enzyme at 1.7 Å resolution are described in *Methods* and summarized in Table S1. A stereoview of the NgrRnl tertiary structure is shown in Fig. 2*A*. It is composed of 21 β-strands, seven α-helices, and two 3_{10} helices, which are displayed over the amino acid sequence in Fig. 2*B*. NgrRnl consists of two domains. As expected, the C-terminal module (amino acids 123–339) is a classic NTase domain. A DALI search (33) of the PDB recovered the NTase domain of T4 Rnl2 as the top "hit," with a *z*-score of 18.4 and 2.4 Å rmsd at 186 Cα positions (Table S2). The aligned tertiary structures of the NgrRnl and T4 Rnl2 NTase domains are shown in Fig. S1*A*. The primary structure alignment in Fig. S1*B*

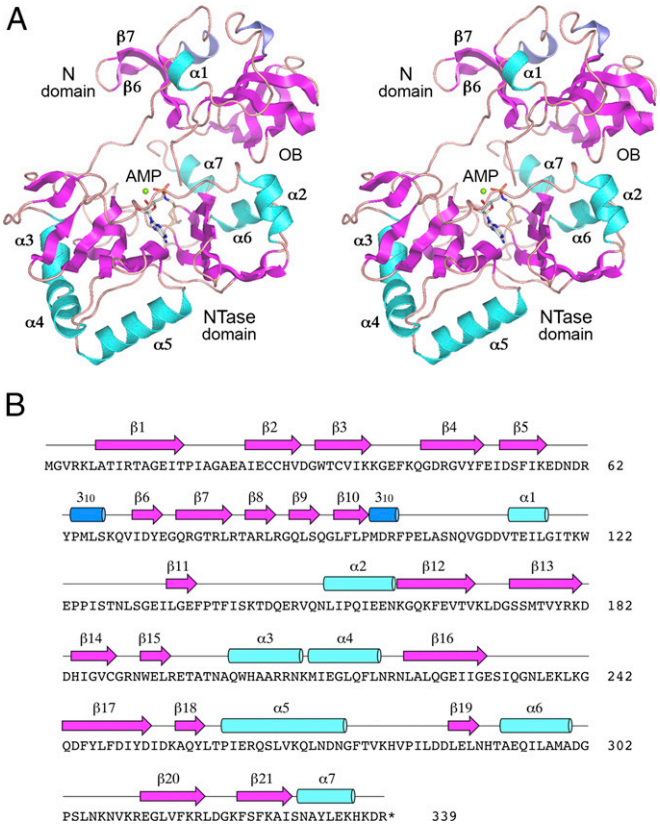

Fig. 2. Overview of NgrRnl structure. (*A*) Stereoview of the tertiary structure of NgrRnl, which consists of an OB-containing N domain and a C-terminal NTase domain, depicted as a ribbon model with magenta β strands, cyan α helices, and blue 3_{10} helices. The lysyl–AMP adduct in the active site is rendered as a stick model. Mn^{2+} is depicted as a green sphere. (*B*) Secondary structure elements (colored as in *A*) are displayed above the NgrRnl amino acid sequence.

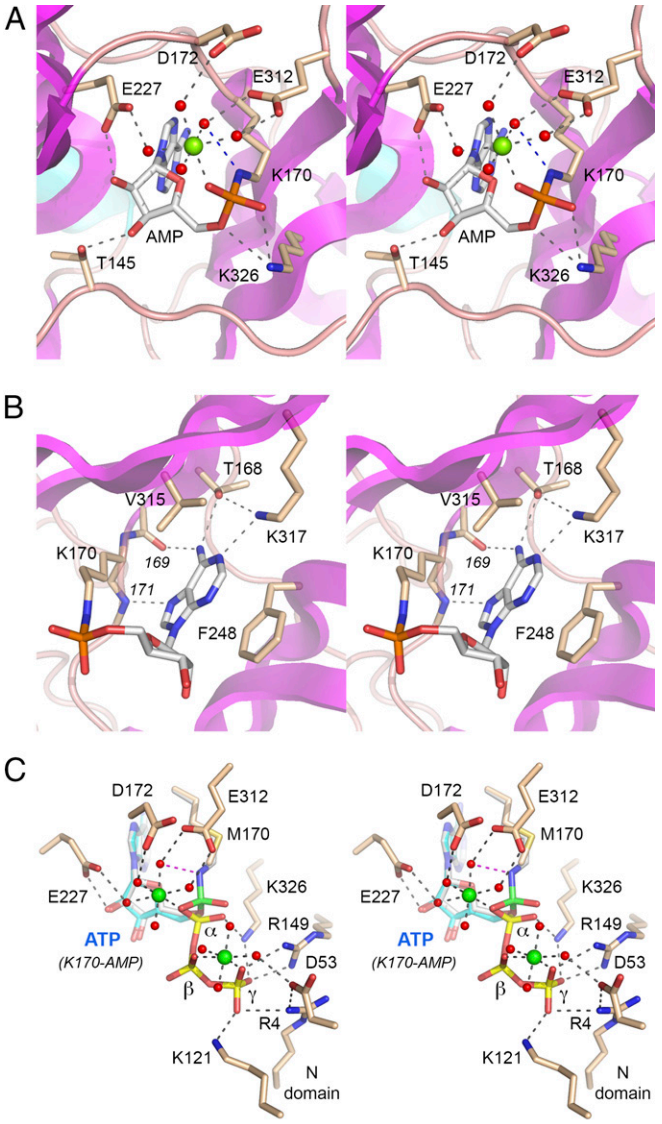

Fig. 3. Structural insights to the mechanism of lysine adenylylation. (*A* and *B*) Stereo views of the active site of the NgrRnl-(lysyl)–AMP•Mn^{2+} intermediate in different orientations highlighting the Mn^{2+} coordination complex and enzymic contacts to the AMP ribose and phosphate (*A*) and contacts to the adenine nucleobase (*B*). Amino acids and AMP are shown as stick models with beige and gray carbons, respectively. A single Mn^{2+} ion and associated waters are depicted as green and red spheres, respectively. Atomic contacts are indicated by dashed lines. (*C*) Stereoview of the active site of the NgrRnl-K170M•ATP•$(Mn^{2+})_2$ Michaelis complex. Amino acids and ATP are shown as stick models with beige and blue carbons, respectively; the ATP phosphorus atoms are colored yellow. Two Mn^{2+} ions and associated waters are depicted as green and red spheres, respectively. Atomic contacts are indicated by dashed lines. The superimposed lysyl–AMP adduct from the covalent intermediate structure is shown with gray carbons and a green α phosphorus atom (to highlight the stereochemical inversion of the phosphorus center after lysine adenylylation).

highlights the conservation of the signature NTase motifs. The NgrRnl NTase domain consists of two central antiparallel β-sheets, a six-strand sheet (with topology β15↓•β14↑•β13↓•β16↑•β17↓•β18↑) and a four-strand sheet (with topology β21↑•β20↓•β12↑•β19↓) that form the adenylate-binding pocket. Six α-helices decorate the lateral surfaces of the NTase domain. Electron density showed that AMP is covalently linked to the motif I Lys170 side-chain in the active site of the NTase domain (Fig. S2), consistent with our biochemical evidence that at least half of the recombinant NgrRnl protein was NgrRnl–AMP (29). The adenosine nucleotide is in the *syn* conformation. The AMP phosphate is coordinated by a divalent cation (green sphere), which we modeled as Mn^{2+} in light of the anomalous difference electron density overlying the metal atom (Fig. S2).

Adenine Nucleotide Specificity. Enzymic contacts to the adenine base are highlighted in Fig. 3*B*. The purine ring is sandwiched between the aromatic ring of Phe248 and the hydrophobic side chain of Val315. Adenine specificity is conferred by amino acid side-chain hydrogen bonds from adenine-N6 to Thr168-Oγ and from Lys317-Nζ to adenine-N1. There are additional main-chain hydrogen bonds from adenine-N6 to Val169 carbonyl and from Leu171 amide to adenine-N7. The ribose 2′-OH and 3′-OH make hydrogen bonds to Glu227 and Thr145, respectively (Fig. 3*A*).

Key Role of the Metal Ion in Catalysis. Waters occupy five of the ligand sites in the octahedral Mn^{2+} coordination complex in the NgrRnl–AMP structure (Fig. 3*A*). The ligase binds the Mn^{2+} complex via water-mediated contacts to Glu227, Glu312, and Asp172. [Glu227 and Glu312 are conserved in DraRnl as Glu230 and Glu305, respectively; their mutation to alanine in DraRnl effaced nick ligation and ligase adenylylation activity (27).] The sixth Mn^{2+} ligand site entails direct metal contact to one of the nonbridging AMP phosphate oxygens. The two other AMP phosphate oxygens are contacted by Lys326. The structure suggests that Lys326 and the metal stabilize a pentavalent transition state of the ATP α-phosphate during the lysine-adenylylation reaction. Our structure also reveals a likely role for the metal complex in stabilizing the unprotonated state of the lysine nucleophile before catalysis, via atomic contact of Lys170-Nζ to one of the waters in the metal complex (indicated by the blue dashed line in Fig. 3*A*).

The N Domain Is a Distinctively Embellished OB-Fold. A DALI search of the NgrRnl N domain (amino acid 1–122) retrieved a C-terminal domain of a "putative tRNA binding protein" from *Staphylococcus saprophyticus* (PDB ID code 3BU2) as the top hit, with a *z*-score of 9.4 and 2.9 Å rmsd at 93 Cα positions. The aligned tertiary structures of the NgrRnl N-domain and 3BU2 are shown in Fig. 4*A*; the aligned secondary and primary structures are shown in Fig. 4*B*. The shared segment (formed by seven of the NgrRnl β-strands) is an OB-fold, an ancient domain found in many nucleic acid binding proteins. Indeed, the second best DALI hit against the NgrRnl N domain, with a *z*-score of 8.8, is an OB module found in *Thermus thermophilus* phenylalanyl-tRNA synthetase (34) (Table S3). NgrRnl's OB-fold is embellished by a distinctive insert module (in blue in Fig. 4*A*), consisting of a three-strand antiparallel β-sheet (β6↑•β7↓•β5↑) and a 3_{10} helix. To our knowledge, this is the first example of an RNA ligase that contains an OB domain and the first instance in which an OB domain is situated upstream of an NTase domain in a member of the covalent NTase superfamily. It is to be emphasized that the N-terminal NgrRnl domain has minimal structural similarity to the C-terminal OB-folds of DNA ligases and mRNA capping enzymes. For example, it aligns very weakly to the OB domains of *Chlorella* virus DNA ligase (9) (*z*-score 2.8; rmsd 6.5 Å at 60 Cα positions) and *Chlorella* virus mRNA capping enzyme (12) (*z*-score 2.1; rmsd 4.9 Å at 58 Cα positions). Moreover, there is virtually no amino acid identity between the NgrRnl N domain and the OB domains of DNA ligase and mRNA capping enzyme. This signifies to us that the N-terminal domain of the Rnl5 family of RNA ligases evolved separately from the C-terminal OB domains of DNA ligases and capping enzymes.

A Michaelis Complex of NgrRnl with ATP and Manganese. To capture a mimetic of the Michaelis complex, we exploited a mutated version, K170M, in which the lysine nucleophile was replaced with

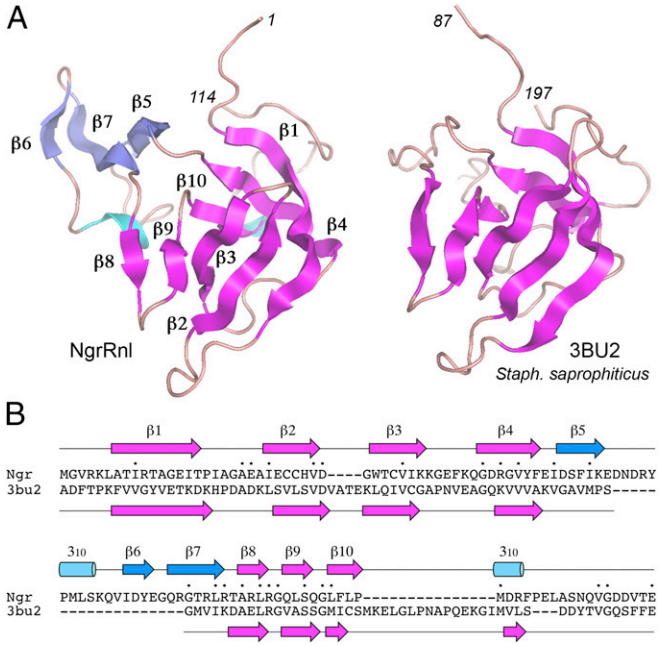

Fig. 4. The N domain is a distinctively embellished OB-fold. (*A*) The tertiary structures of the NgrRnl N domain and the top DALI hit *S. saprophyticius* 3BU2 were aligned and then offset horizontally. The N- and C-terminal amino acid numbers are italicized. The NgrRnl secondary structure elements are labeled. The magenta β-strands are common to both proteins; the blue β-strands and the 3_{10} helices are unique to NgrRnl. (*B*) The NgrRnl and 3BU2 primary structures are aligned according to DALI. Gaps in the alignment are indicated by dashes. Positions of amino acid side chain identity/similarity are indicated by a dot (•) above the sequence. The secondary structure elements of NgrRnl and 3BU2 (colored as in *A*) are shown above and below the respective amino acid sequences.

methionine (effectively isosteric to lysine, minus the ε-amino group). NgrRnl-K170M was preincubated with 2 mM ATP and 5 mM $MnCl_2$ before crystallization. The refined 1.9 Å structure (R/R_{free} = 0.176/0.247) (Table S1) comprised the entire NgrRnl polypeptide, with electron density for ATP and two manganese ions in the active site (Fig. S3). The tertiary structure of the ligase protomer in the ATP complex was essentially identical to that of the ligase–AMP intermediate (*z*-score 51.0; rmsd of 0.6 Å at 339 Cα positions). Fig. 3*C* shows a stereoview of the active site of the ATP complex (ATP rendered with cyan carbons and yellow phosphorus atoms; Mn^{2+} as green spheres; water as red spheres), highlighting atomic interactions relevant to catalysis. For comparison, we included just the lysyl–AMP adduct from the NgrRnl–AMP structure (AMP with gray carbons and green Pα atom in Fig. 3*C*). The adenosine nucleosides superimpose almost perfectly and the Met170 side-chain is indeed virtually isosteric to Lys170. The salient mechanistic insights are summarized below.

Geometry and Stereochemistry of Lysine Adenylylation. The Lys170-Nζ is situated 2.9 Å from the ATP α phosphorus, in an apical orientation to the pyrophosphate leaving group (Nζ–Pα–O3α angle = 169°). Consistent with a single-step in-line mechanism, we see that the α-phosphate undergoes stereochemical inversion during the transition from ATP Michaelis complex to lysyl–AMP intermediate (Fig. 3*C*).

Two-Metal Mechanism of Nucleotidyl Transfer. The "catalytic" octahedral Mn^{2+} complex (Mn1) that engages the ATP α-phosphate in the Michaelis complex is identical to that seen in the lysyl–AMP intermediate. The ATP-bound structure revealed a second Mn^{2+} atom (Mn2), coordinated octahedrally to four waters and to ATP β- and γ-phosphate oxygens. We infer that the second Mn^{2+} promotes lysine adenylylation by ensuring a proper conformation of the ATP triphosphate moiety conducive to expulsion of the pyrophosphate (PP_i) leaving group.

Role of the N Domain in Orienting the PP_i Leaving Group. There are no direct enzymic contacts to the ATP β-phosphate. Lys326 and Arg149 are the only side chains of the NTase domain that engage the ATP phosphates, via a bifurcated interaction of Lys326 with α- and γ-phosphates and a bidentate interaction of Arg149 with the γ-phosphate. The structure highlights direct contacts to the ATP γ-phosphate from the N domain of NgrRnl, via Arg4 and Lys121, and to the second Mn^{2+} coordination complex via Asp53. The N domain contacts to ATP and Mn2 in the Michaelis complex aid in orienting the PP_i leaving group during step 1 lysine adenylylation, and account for the severe decrement in step 1 activity when the N domain is deleted (29).

"Who's on First"—Does Manganese Precede ATP? Because the catalytic Mn^{2+} complex makes only one contact to the ATP α-phosphate, versus five ligand sites filled by waters, four of which are coordinated by NgrRnl active site side chains (Asp172, Glu227, Glu312), we hypothesize that the first metal cofactor Mn1 binds to the ligase before ATP. Because the second Mn^{2+} complex that bridges the β- and γ-phosphates makes only one water-mediated contact to the ligase (Asp53), and because a second metal is absent from the ligase-AMP intermediate (presumably having departed in a complex with the PP_i leaving group), we envision that ATP•Mn(2) binds to ligase•Mn(1) to form a ligase•ATP•$(Mn^{2+})_2$ Michaelis complex. To test the "Mn1 first" hypothesis, we grew crystals of NgrRnl-K170M that had been preincubated with 5 mM $MnCl_2$ (without ATP), collected diffraction data to 2.2 Å resolution, and refined the structure to R/R_{free} of 0.190/0.266 (Table S1). The difference density maps indicate that there is a single manganese ion, but no nucleotide, in the active site. The Mn^{2+} is coordinated to three waters, signifying that the metal coordination complex is incomplete in the absence of nucleotide.

Fig. 5 shows a stereoview of the metals and metal ligands in the superimposed NgrRnl•Mn^{2+} (colored blue), NgrRnl•ATP•Mn^{2+} (colored gold), and NgrRnl–AMP•Mg^{2+} (colored green) structures. The five waters in the Mn^{2+} coordination complex of the adenylate-bound ligases are labeled "a" to "e." The three water sites that are occupied in the NgrRnl•Mn^{2+} binary complex are "a," "b," and "c." Water sites "d" and "e" that are vacant in the NgrRnl•Mn^{2+} binary complex occupy apical positions in the octahedron. Water "d" is coordinated jointly by Glu227-Oε and the adenosine ribose O2′, which might explain why this site is not filled before ATP binding. Two of the sites that are filled in the NgrRnl•Mn^{2+} binary complex make pairwise contacts with the ligase. Water "b" is engaged by hydrogen bonds to Asp172-Oδ and the Gly173 main-chain carbonyl. Water "a" (the one that contacts Lys170-Nζ) is engaged via hydrogen bonds to Glu312-Oε and the Leu171 main-chain carbonyl. We surmise that the initial binding of the catalytic metal Mn1 to the ligase stabilizes the unprotonated form of the lysine nucleophile in advance of ATP binding.

Discussion

The present study provides new insights to the structure, mechanism, and evolution of RNA ligases. We report the atomic structure of NgrRnl, the eukaryal founder of a Rnl5 family of nick sealing enzymes. The distinctive domain composition of NgrRnl fortifies the theme that multiple different clades of RNA ligases emerged by fusions of diverse accessory modules to a shared NTase domain. To our knowledge, NgrRnl is the first instance in which an OB domain is intrinsic to an RNA ligase polypeptide.

The NgrRnl•ATP•$(Mn^{2+})_2$ structure, when superimposed on the NgrRnl–AMP•Mn^{2+} structure, captures (for the first time to our knowledge) a mimetic of the Michaelis complex of the ligase

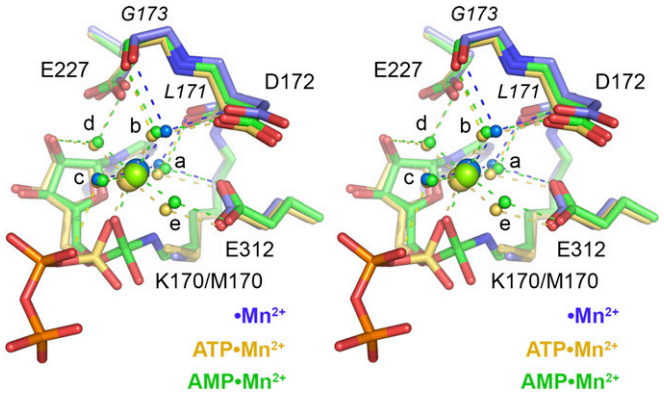

Fig. 5. Comparison of the catalytic Mn^{2+} site with and without adenylate. Stereoview of the superimposed active sites of the NgrRnl-K170M•Mn^{2+} binary complex (with carbon atoms, waters, and Mn^{2+} colored blue), the NgrRnl-K170M•ATP•Mn^{2+} Michaelis complex (gold), and the NgrRnl-(lysyl)–AMP•Mn^{2+} intermediate (green). The five waters in the octahedral Mn^{2+} complex of the ATP and lysyl–AMP structures are labeled "a" to "e." Only the "a," "b," and "c" water sites are occupied in the Mn^{2+} binary complex. Atomic contacts to the metal and waters are shown as dashed lines.

adenylylation reaction in which the nucleophile and PP_i leaving group are oriented correctly for catalysis and the proper catalytic metal cofactor is coordinated in a mechanistically instructive fashion to ATP, the lysine nucleophile, and essential carboxylate side-chains of the NTase domain. Early structures of the covalent ligase–AMP intermediates of *Chlorella* virus DNA ligase (9) and *Mycobacterium tuberculosis* DNA ligase D (35) obtained from crystals that were soaked or grown in the presence of non-physiological metals (e.g., lutetium or zinc) suggested the location of a "catalytic" metal near the AMP phosphate (corresponding to Mn1 in the NgrRnl structures). The structure of T4 Rnl1 with the unreactive analog AMPCPP in the active site obtained from crystals grown in the presence of calcium revealed a calcium ion in the catalytic metal site, coordinated to six waters and one AMPCPP α-phosphate oxygen (17). The closest counterpart to the present NgrRnl structures is that of the bacterial Rnl4 family RNA ligase–AMP•Mg^{2+} covalent intermediate formed *in crystallo* (21), in which a catalytic Mg^{2+} ion contacts the AMP phosphate directly and the remaining five positions in the octahedral metal complex are filled by waters, two of which are engaged by motif III and IV glutamate side chains (the equivalents of Glu227 and Glu312 in NgrRnl).

In addition to highlighting how direct metal interaction with the ATP α-phosphate would stabilize the transition state during lysine adenylylation, the NgrRnl structures offer clues to a long-standing conundrum in ligase biochemistry. As Robert Lehman pointed out in 1974 (36), it is a mystery how lysine (with a predicted pK_a value of ∼10.5) loses its proton at physiological pH to attain the unprotonated state required for attack on the α-phosphorus of ATP or NAD^+. In principle, ligase might use a general base to deprotonate the lysine. Alternatively, the active site milieu lowers the pK_a of the lysine so that a significant fraction of the side-chain is unprotonated at physiological pH. Reduction of the lysine pK_a could be driven by positive charge potential surrounding lysine-Nζ, analogous to how the pK_a of the cysteine nucleophile in *Yersinia* protein tyrosine phosphatase is lowered to 4.7 by a network of positive potential from main-chain amides surrounding the sulfur (37, 38). Several crystal structures of ligase and capping enzyme absent metals (7, 12, 16, 19) provided scant support for either explanation. In these structures, the motif I lysine nucleophile is located next to the motif IV glutamate or aspartate side-chain (7, 12, 16, 19). The lysine and the motif IV carboxylate form an ion pair, the anticipated effect of which is to increase the pK_a of lysine by virtue of surrounding negative charge. It is unlikely that a glutamate or aspartate anion could serve as general base to abstract a proton from the lysine cation. A potential solution to the problem would be if a divalent cation (coordinated by the motif IV carboxylate and the ATP α phosphate) abuts the lysine-Nζ and drives down its pK_a.

An explicit model was proposed for the atomic contacts of the catalytic metal in the transition state of T4 DNA ligase (39), although there is no direct structural information on this enzyme in any state along its reaction pathway. According to the model, the catalytic metal makes direct contacts with lysine-Nζ (the nucleophile) and with the α–β bridging oxygen of ATP (the leaving group) as two of the component ligands of an octahedral metal coordination complex that also includes four waters. The model does not mandate a direct contact between the metal and nonbridging ATP α-phosphate oxygens; instead the nonbridging phosphate oxygens are depicted as accepting hydrogen bonds from two of the metal-bound waters (39). The NgrRnl structures presented here militate strenuously against such a model. Our structures implicate a metal-bound water, coordinated by the motif IV glutamate and situated within hydrogen-bonding distance of the lysine-Nζ, in stabilizing the unprotonated state of the lysine before ATP binding. Also, our structures are not consistent with the catalytic metal being involved in expulsion of the PP_i leaving group, with which it makes no contacts.

Indeed, the NgrRnl structures instate a two-metal mechanism of lysine adenylylation, in which a second Mn^{2+} ion engages the ATP β- and γ-phosphates to attain a catalytically conducive orientation of the PP_i leaving group apical to the lysine nucleophile. An earlier crystal structure of the isolated NTase domain of trypanosome RNA editing ligase 1 (TbrREL1, an Rnl2-family enzyme) in complex with ATP had revealed a single Mg^{2+} ion bridging the β- and γ-phosphates (40), analogous to the second metal in the NgrRnl•ATP structure. However, the TbrREL1 NTase•ATP structure is missing a catalytic metal ion that engages the α phosphate, and it represents an off-pathway state in which: (*i*) the noncatalytic Mg^{2+} ion occupies the position of the catalytic metal in the NgrRnl Michaelis complex; (*ii*) the ATP β- and γ-phosphates are in an unreactive orthogonal orientation to the lysine nucleophile (Nζ–Pα–O3α angle = 96°); and (*iii*) the motif IV glutamate forms a salt bridge with the lysine nucleophile that disfavors its deprotonation (Fig. S4).

In NgrRnl, the Mn^{2+}-bound γ-phosphate makes electrostatic and hydrogen-bonding interactions with the signature N-terminal domain of NgrRnl, thereby explaining the importance of the N domain for ligase adenylylation via its role in binding ATP and orienting the PP_i leaving group. [A second metal was seen previously in the T4 Rnl1•AMPCPP structure (17). Modeled as magnesium, it makes direct contact to just one of the AMPCPP β-phosphate oxygens; this Mg^{2+} ion is coordinated directly and via water to amino acids in the distinctive C-terminal domain of T4 Rnl1.]

Our findings highlight how Nature has adopted diverse structural solutions to the problem of leaving group geometry in the chemistry of lysine nucleotidylylation. In the case of NgrRnl and NAD^+-dependent DNA ligases, the respective PP_i and NMN leaving groups are oriented for in-line catalysis by clade-specific modules fused to the N terminus of the NTase domain: a unique OB domain in NgrRnl (29) and a unique "Ia domain" in NAD^+-dependent DNA ligases (15, 16, 41). In the case of exemplary ATP-dependent DNA ligases and GTP-dependent RNA capping enzymes, the PP_i leaving group is positioned by a conserved motif VI (RxDK) that defines a clade-specific OB module fused to the C terminus of the NTase domain (12, 42, 43).

Methods

NgrRnl and NgrRnl-K170M crystals were grown by hanging-drop vapor diffusion after mixing purified recombinant protein with 0.1 M Hepes pH 6.5, 30% PEG6000 precipitant solution. The crystals belonged to space group $P3_2$ and had one ligase protomer per asymmetric unit. The structure of native

NgrRnl–AMP was solved using single-wavelength anomalous dispersion data from a single crystal of SeMet-NgrRnl. The final model of native NgrRnl–AMP was refined to R_{work}/R_{free} of 0.167/0.209 (Table S1). The NgrRnl structure was used as a starting model against which the diffraction data for single NgrRnl-K170M•ATP•Mn^{2+} and NgrRnl-K170M•Mn^{2+} crystals were refined (Table S1). Full details of protein purification, crystallization, diffraction data collection, and structure determination are provided in *SI Methods*.

ACKNOWLEDGMENTS. This research was supported by National Institutes of Health Grant GM42498 (to S.S.) and Memorial Sloan Kettering Cancer Center Core Grant P30-CA008748.

1. Amitsur M, Levitz R, Kaufmann G (1987) Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. *EMBO J* 6(8): 2499–2503.
2. Sidrauski C, Cox JS, Walter P (1996) tRNA ligase is required for regulated mRNA splicing in the unfolded protein response. *Cell* 87(3):405–413.
3. Schnaufer A, et al. (2001) An RNA ligase essential for RNA editing and survival of the bloodstream form of *Trypanosoma brucei*. *Science* 291(5511):2159–2162.
4. Schwer B, Sawaya R, Ho CK, Shuman S (2004) Portability and fidelity of RNA-repair systems. *Proc Natl Acad Sci USA* 101(9):2788–2793.
5. Nandakumar J, Schwer B, Schaffrath R, Shuman S (2008) RNA repair: An antidote to cytotoxic eukaryal RNA damage. *Mol Cell* 31(2):278–286.
6. Shuman S, Lima CD (2004) The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Curr Opin Struct Biol* 14(6):757–764.
7. Ho CK, Wang LK, Lima CD, Shuman S (2004) Structure and mechanism of RNA ligase. *Structure* 12(2):327–339.
8. Subramanya HS, Doherty AJ, Ashford SR, Wigley DB (1996) Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell* 85(4):607–615.
9. Odell M, Sriskanda V, Shuman S, Nikolov DB (2000) Crystal structure of eukaryotic DNA ligase-adenylate illuminates the mechanism of nick sensing and strand joining. *Mol Cell* 6(5):1183–1193.
10. Nandakumar J, Nair PA, Shuman S (2007) Last stop on the road to repair: Structure of *E. coli* DNA ligase bound to nicked DNA-adenylate. *Mol Cell* 26(2):257–271.
11. Nair PA, et al. (2007) Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nat Struct Mol Biol* 14(8):770–778.
12. Håkansson K, Doherty AJ, Shuman S, Wigley DB (1997) X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell* 89(4):545–553.
13. Fabrega C, Shen V, Shuman S, Lima CD (2003) Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Mol Cell* 11(6):1549–1561.
14. Doamekpor SK, Sanchez AM, Schwer B, Shuman S, Lima CD (2014) How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. *Genes Dev* 28(12):1323–1336.
15. Sriskanda V, Shuman S (2002) Conserved residues in domain Ia are required for the reaction of *Escherichia coli* DNA ligase with NAD^+. *J Biol Chem* 277(12):9695–9700.
16. Gajiwala KS, Pinko C (2004) Structural rearrangement accompanying NAD^+ synthesis within a bacterial DNA ligase crystal. *Structure* 12(8):1449–1459.
17. El Omari K, et al. (2006) Molecular architecture and ligand recognition determinants for T4 RNA ligase. *J Biol Chem* 281(3):1573–1579.
18. Nandakumar J, Shuman S, Lima CD (2006) RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. *Cell* 127(1): 71–84.
19. Brooks MA, et al. (2008) The structure of an archaeal homodimeric ligase which has RNA circularization activity. *Protein Sci* 17(8):1336–1345.
20. Smith P, Wang LK, Nair PA, Shuman S (2012) The adenylyltransferase domain of bacterial Pnkp defines a unique RNA ligase family. *Proc Natl Acad Sci USA* 109(7): 2296–2301.
21. Wang P, et al. (2012) Molecular basis of bacterial protein Hen1 activating the ligase activity of bacterial protein Pnkp for RNA repair. *Proc Natl Acad Sci USA* 109(33): 13248–13253.
22. Wang LK, Nandakumar J, Schwer B, Shuman S (2007) The C-terminal domain of T4 RNA ligase 1 confers specificity for tRNA repair. *RNA* 13(8):1235–1244.
23. Zhang C, Chan CM, Wang P, Huang RH (2012) Probing the substrate specificity of the bacterial Pnkp/Hen1 RNA repair system using synthetic RNAs. *RNA* 18(2):335–344.
24. Nandakumar J, Ho CK, Lima CD, Shuman S (2004) RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *J Biol Chem* 279(30):31337–31347.
25. Nandakumar J, Shuman S (2004) How an RNA ligase discriminates RNA versus DNA damage. *Mol Cell* 16(2):211–221.
26. Martins A, Shuman S (2004) An RNA ligase from *Deinococcus radiodurans*. *J Biol Chem* 279(49):50654–50661.
27. Raymond A, Shuman S (2007) *Deinococcus radiodurans* RNA ligase exemplifies a novel ligase clade with a distinctive N-terminal module that is important for 5′-PO_4 nick sealing and ligase adenylylation but dispensable for phosphodiester formation at an adenylylated nick. *Nucleic Acids Res* 35(3):839–849.
28. Schmier BJ, Shuman S (2014) Effects of 3′-OH and 5′-PO_4 base mispairs and damaged base lesions on the fidelity of nick sealing by *Deinococcus radiodurans* RNA ligase. *J Bacteriol* 196(9):1704–1712.
29. Unciuleac MC, Shuman S (2015) Characterization of a novel eukaryal nick-sealing RNA ligase from *Naegleria gruberi*. *RNA* 21(5):824–832.
30. Daly MJ (2009) A new perspective on radiation resistance based on *Deinococcus radiodurans*. *Nat Rev Microbiol* 7(3):237–245.
31. Slade D, Radman M (2011) Oxidative stress resistance in *Deinococcus radiodurans*. *Microbiol Mol Biol Rev* 75(1):133–191.
32. Liu Y, et al. (2003) Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc Natl Acad Sci USA* 100(7):4191–4196.
33. Holm L, Kääriäinen S, Rosenström P, Schenkel A (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics* 24(23):2780–2781.
34. Mosyak L, Reshetnikova L, Goldgur Y, Delarue M, Safro MG (1995) Structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus*. *Nat Struct Biol* 2(7): 537–547.
35. Akey D, et al. (2006) Crystal structure and nonhomologous end-joining function of the ligase component of *Mycobacterium* DNA ligase D. *J Biol Chem* 281(19):13412–13423.
36. Lehman IR (1974) DNA ligase: Structure, mechanism, and function. *Science* 186(4166): 790–797.
37. Zhang ZY, Dixon JE (1993) Active site labeling of the *Yersinia* protein tyrosine phosphatase: The determination of the pKa of the active site cysteine and the function of the conserved histidine 402. *Biochemistry* 32(36):9340–9345.
38. Stuckey JA, et al. (1994) Crystal structure of *Yersinia* protein tyrosine phosphatase at 2.5 A and the complex with tungstate. *Nature* 370(6490):571–575.
39. Cherepanov AV, de Vries S (2002) Kinetic mechanism of the Mg^{2+}-dependent nucleotidyl transfer catalyzed by T4 DNA and RNA ligases. *J Biol Chem* 277(3):1695–1704.
40. Deng J, Schnaufer A, Salavati R, Stuart KD, Hol WG (2004) High resolution crystal structure of a key editosome enzyme from *Trypanosoma brucei*: RNA editing ligase 1. *J Mol Biol* 343(3):601–613.
41. Sriskanda V, Moyer RW, Shuman S (2001) NAD^+-dependent DNA ligase encoded by a eukaryotic virus. *J Biol Chem* 276(39):36100–36109.
42. Sawaya R, Shuman S (2003) Mutational analysis of the guanylyltransferase component of mammalian mRNA capping enzyme. *Biochemistry* 42(27):8240–8249.
43. Samai P, Shuman S (2012) Kinetic analysis of DNA strand joining by *Chlorella* virus DNA ligase and the role of nucleotidyltransferase motif VI in ligase adenylylation. *J Biol Chem* 287(34):28609–28618.
44. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
45. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
46. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3): 240–255.
47. Jones TA, Zou J-Y, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47(Pt 2):110–119.