The structure of an archaeal homodimeric ligase which has RNA circularization activity

MARK ADRIAN BROOKS,^{1,3} LAURENCE MESLET-CLADIÉRE,^{2,3,4} MARC GRAILLE,¹
JOËLLE KUHN,² KARINE BLONDEAU,¹ HANNU MYLLYKALLIO,²
AND HERMAN VAN TILBEURGH¹

¹IBBMC-CNRS, Université de Paris-Sud, CNRS-UMR8619, IFR115, 91405 Orsay, France

²Institut de Génétique et Microbiologie, Université Paris-Sud, CNRS-UMR8621, IFR115, 91405 Orsay, France

(RECEIVED April 1, 2008; FINAL REVISION May 27, 2008; ACCEPTED May 28, 2008)

Abstract

The genome of *Pyrococcus abyssi* contains two open reading frames encoding proteins which had been previously predicted to be DNA ligases, Pab2002 and Pab1020. We show that while the former is indeed a DNA ligase, Pab1020 had no effect on the substrate deoxyoligo-ribonucleotides tested. Instead, Pab1020 catalyzes the nucleotidylation of oligo-ribonucleotides in an ATP-dependent reaction, suggesting that it is an RNA ligase. We have solved the structure of Pab1020 in complex with the ATP analog AMPPNP by single-wavelength anomalous dispersion (SAD), elucidating a structure with high structural similarity to the catalytic domains of two RNA ligases from the bacteriophage T4. Additional carboxy-terminal domains are also present, and one of these mediates contacts with a second protomer, which is related by noncrystallographic symmetry, generating a homodimeric structure. These C-terminal domains are terminated by short domain swaps which themselves end within 5 Å of the active sites of the partner molecules. Additionally, we show that the protein is indeed capable of circularizing RNA molecules in an ATP-dependent reaction. These structural and biochemical results provide an insight into the potential physiological roles of Pab1020.

Keywords: *Pyrococcus abyssi*; RNA, crystallization; nucleotidyl transferase; ligase; gene expression; X-ray crystallography

Supplemental material: see www.proteinscience.org

³These authors contributed equally to this work.

⁴Present address: Institut de Biologie Physico-Chimique, F-75005 Paris, France.

Reprint requests to: Herman van Tilbeurgh, Université de Paris-Sud, CNRS-UMR8619, IFR115, Bâtiment 430, 91405 Orsay, France; e-mail: Herman.Van-Tilbeurgh@u-psud.fr; fax: (33) 1-69-85-37-15; or Hannu Myllykallio, Institut de Génétique et Microbiologie, Université Paris-Sud, CNRS-UMR8621, IFR115, Bâtiment 409, 91405 Orsay, France; e-mail: hannu.myllykallio@igmors.u-psud.fr; fax: (33) 1-69-15-78-08.

Abbreviations: AMPPNP, phosphoamino phosphonic acid-adenylate ester; ATP, adenosine triphosphate; T4 Rnl, bacteriophage T4 RNA ligase; AMPcPP, adenosine 5'-(α,β -methylene)triphosphate; SAD, single-wavelength anomalous dispersion.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.035493.108>.

Polynucleotide ligases catalyze phosphodiester bond formation using nicked nucleic acid substrates with the high energy nucleotide of either ATP or NAD⁺ as a cofactor (Tomkinson et al. 2006; Pascal 2008) and are members of the nucleotidyl transferase superfamily, which also encompasses mRNA capping enzymes (Shuman and Schwer 1995). Nucleotidyl transferases catalyze ligation via multistep reactions. Firstly, covalent reaction of a nucleotide cofactor with a conserved lysine residue at the active site yields an enzyme-nucleotide intermediate. Secondly, the nucleotide is transferred from the enzyme to the 5'-phosphate of the polynucleotide that is to be either ligated or capped. Comparison of the sites of

adenylation of various nucleotidyl transferases allowed a conserved Kx(D/N)G motif to be identified (motif “I,” Supplemental Fig. S1) (Shuman and Schwer 1995). A further four other motifs which are conserved throughout polynucleotide ligases and RNA capping enzymes have been subsequently identified by sequence analysis (motifs “III,” “IIIA,” “IV,” and “V”; see Figs. 2A, 3B and Supplemental Fig. S1).

ATP-dependent DNA ligases are present in many organisms and include human DNA ligases I, III, and IV, which act during the replication and repair of DNA, as well as during genetic recombination and nonhomologous end-joining (Tomkinson et al. 2006). ATP-dependent DNA ligases have also been identified from the virus *Chlorella*, as well as bacteria (Ho et al. 1997; Nandakumar et al. 2007). NAD⁺-dependent ligases are confined to a much narrower range of organisms; ligases that are able to use this cofactor have only been found in bacteria (in which some such ligases have been found to be essential), as well as archaea and certain insect poxviruses (Lee et al. 2000; Nakatani et al. 2000; Sriskanda et al. 2001).

In contrast to the diverse classes of DNA ligases, far fewer ligases that use ribonucleic acid as a substrate have been characterized. Examples of such enzymes include RNA ligase 1 from the bacteriophage T4 (T4 Rnl1), which act to negate a bacterial defense mechanism by repairing tRNA^{Lys} (Silber et al. 1972). The structure of this ligase has recently been solved, giving an insight into its probable mode of interaction with tRNA (El Omari et al. 2006). Transfer RNA ligases from baker's yeast and wheat germ have also been shown to religate tRNA after the removal of introns from precursors during tRNA splicing, generating a characteristic 2'-phosphomonoester, 3',5'-phosphodiester bond (Konarska et al. 1981; Westaway et al. 1988).

More recently, another pathway called “RNA editing” has been discovered in trypanosomatids, wherein guide RNAs direct the multimeric “editosome” complex to sites on mitochondrial RNAs at which UMP nucleotides are added or deleted (Sabatini and Hajduk 1995). Within editosomes, ligase subunits religate the substrate mRNA fragments after such processing, and the crystal structure of an exemplary member of this family has been described (Deng et al. 2004). It has also emerged that further ligases exist, for example in the radiation-resistant bacterium *Deinococcus radiodurans* and certain archaea (Ho and Shuman 2002; Martins and Shuman 2004).

During the course of our research into DNA replication in the archaeon *Pyrococcus abyssi* (Meslet-Cladière et al. 2007), it came to our attention that its genome encodes a ligase which had not previously been characterized, called Pab1020. Pab1020 is a member of the Clusters of

Orthologous Groups of proteins (COG), number 1423 (Tatusov et al. 1997), which includes proteins not only from archaea such as *Archaeoglobus fulgidus* and *Pyrococcus horikoshii*, but also the bacterium *Aquifex aeolicus*, and is annotated as being a family of ATP-dependent DNA ligases which are homologous to DNA ligase III. Sequence analysis has previously shown that Pab1020 has a catalytic domain of an ATP-dependent ligase which showed significant sequence similarity to T4 Rnl2 (Ho and Shuman 2002), though the remaining sequence was of unknown fold.

We decided to investigate the catalytic activity of Pab1020, and found that when incubated with adenosine triphosphate (ATP), but not other nucleotide triphosphates, Pab1020 was adenylated, in common with ATP-dependent ligases. We therefore resolved to gain a further enzymatic and structural understanding of this ligase, as well as to identify the functional niche of this ATP-dependent ligase within *Pyrococcus abyssi*.

In this work, we have determined the crystal structure of Pab1020, and we show that it shares highest structural similarity with certain RNA ligases. Concordant with these results, enzyme assays demonstrate that it is capable of circularizing oligo-ribonucleotides in vitro, in an ATP-dependent manner. Pab1020 is homodimeric, and is the first structure of an oligomeric RNA ligase to be determined. As such, the twin active sites provide a clue as to the potential roles of Pab1020 as an RNA ligase which may perform two reactions concomitantly.

Results

Adenylyltransferase activity

Sequence similarity indicated that Pab1020 possesses a ligase catalytic domain, a fold which is a member of the nucleotidyl-transferase domain superfamily and includes mRNA capping enzymes. We first investigated the specificity of the reaction of Pab1020 with nucleotide triphosphates, in which a reactive enzyme-nucleotide intermediate is formed. Incubation of Pab1020 with each of α -³²P radio-labeled ATP, GTP, UTP, and CTP showed a clear selectivity for adenylation (Fig. 1A), and no reaction was evident when UTP or CTP were input. A slight reaction with GTP was apparent upon prolonged exposure of the phosphorimaging plate to the gel, perhaps due to a minor reaction product with GTP.

The fact that Pab1020 is adenylated by ATP in a similar way to DNA ligases, coupled with the bioinformatics evidence suggesting the presence of a ligase catalytic domain, prompted us to attempt to determine its structure. This strategy has allowed us to combine biochemical characterization with structural analysis to try to identify the functions of Pab1020.

Figure 1. Pab1020 adenylation reaction and ligation assays. (A) Reaction mixtures containing 0.2 pmol of Pab1020 in 50 mM of Tris-HCl pH 6.8, 5 mM of DTT, 1 mM of $MgCl_2$, and 0.1 μM of α - ^{32}P -labeled nucleoside triphosphates were incubated for 15 min at 75°C. Reactions were analyzed under denaturing conditions after SDS-PAGE and autoradiography. Reaction products in panels A, B, and C are marked with arrows. (B) Adenylation assays using oligoribonucleotides were performed using 2 pmol of Pab2002 or Pab1020. The reaction mixtures containing 1 pmol of labeled RNA oligo in buffer R were incubated for 30 min at 75°C. Formation of adenylated oligo-ribonucleotide was only detected when Pab1020 was input. (C) DNA ligation assay using an ATP-dependent DNA ligase Pab2002 and Pab1020. The reaction mixtures contained 50 mM Tris-HCl pH 7, 5 mM DTT, 1 mM $MgCl_2$ (buffer R), and 1 mM ATP. Two picomoles of Pab1020 or Pab2002 were included in reaction mixtures containing 1 pmol of labeled “nicked” DNA substrate, and were incubated 30 min at 75°C. Reaction substrates and products were analyzed on a denaturing gel. The primer Dn1 was labeled at the 5' end as indicated by an asterisk. A molecular weight shift indicating DNA ligation was only observed in the case of Pab2002.

Overall structure of Pab1020

Crystals of full-length recombinant Pab1020 were obtained by vapor diffusion and diffracted to $\sim 3.0 \text{ \AA}$ resolution. Phases were obtained from selenomethionyl-substituted Pab1020 in a single-wavelength anomalous dispersion experiment (cf. Table 1 and Materials and Methods). The crystal structure contains two protomers per asymmetric unit, and each one is bound to a single molecule of phosphoamino phosphonic acid-adenylate ester (referred to as AMPPNP hereafter) that had been co-crystallized with Pab1020 (Fig. 2A). Of the 382 amino acids, continuous density is apparent for all except for the first 10 amino acids, which were not modeled. Furthermore, there is no clear electron density for four side chains of the N-terminal domain, probably due to disorder, and were therefore modeled as alanine residues (residues 10 and 36–38). The two chains are almost identical, with a root mean square displacement (RMSD) of 0.092 $\text{\AA}$ over 377 protein backbone $C\alpha$ atoms, although this value will have been biased by the tight NCS restraints that had been maintained throughout the refinement. Analysis with PROCHECK indicates that the model is of excellent stereochemistry; the Ramachandran plot indicates that of the nonglycine and nonproline residues, none are in disallowed regions, 1.7% are “generously allowed,” and 11.1% are in the “additionally allowed” region (Laskowski et al. 1993).

Each monomer of Pab1020 consists of four domains which pack together in a generally globular arrangement (Fig. 2A). The catalytic domain envelops the bound nucleotide in a similar manner to other members of the nucleotidyl-transferase superfamily, which include DNA and RNA ligases as well as capping enzymes (Shuman and Schwer 1995; Tomkinson et al. 2006). However, the flanking domains differentiate this structure from other members of this superfamily. Residues 10–61 form a small independent domain which consists of two α -helices followed by a four-stranded antiparallel β -sheet of strand order 1234. Downstream from the catalytic domain is a three-helix bundle (from residues 245 to 313 and hereafter referred to as the “dimerization domain”) that mediates the interaction between the two monomers. This is followed by a C-terminal domain with a β - α - β - β - α - β topology, the first three β -strands of which combine to form a continuous antiparallel β -sheet, onto which the first α -helix abuts.

Structurally similar folds to the catalytic domain of Pab1020 present in the Protein Data Bank (PDB) were identified using secondary structure matching (SSM) (Krissinel and Henrick 2004). This algorithm identified the closest three structural homologs as: T4 RNA ligase 1 (Z-score 6.4, RMSD 2.78 $\text{\AA}$ over 200 aligned residues), an RNA editing ligase from *Trypanosoma brucei* (Z-score

Figure 2. Architecture of the Pab1020 homodimer. (A) Domain structure of Pab1020. The domains of Pab1020 are colored as follows: N-terminal domain, red; catalytic domain, green; dimerization domain, blue; C-terminal domain, orange. At left is a view facing the active site, and at right the molecule has been rotated by 90° around the axis shown. (Below) Schematic diagram of the domain boundaries. Locations of the conserved motifs I, III, IIIA, IV, and V of the nucleotidyl transferase superfamily are shown in dark green. (B) Molecule A is shown in green and molecule B in red as cartoons, and the molecules of AMPPNP as ball-and-stick models. The diagram demonstrates the “clamp” formation of the N-terminal domains and the cofactor analogs are highlighted. This figure, as well as Figures 3 and 5 were prepared with BOBSCRIPT (Esnouf 1997).

5.1, RMSD 2.42 Å over 162 residues), and T4 RNA ligase 2 (Z-score 6.2, RMSD 2.39 Å over 164 aligned residues) (PDB codes 2C5U, 1XDN, and 2HVQ, respectively). These three most similar structures are depicted in Figure 3 for comparison with Pab1020. Both RNA and DNA ligases are present among the entire list of structurally similar proteins identified by SSM, which is unsurprising, given the similarity of the fold of the catalytic domain among ligases.

T4 RNA ligase 1 (T4Rnl1) and Pab1020 both contain an N-terminal domain with a generally common topology

(Fig. 3A,B), but which is neither present in T4 RNA ligase 2 (T4 Rnl2) (Fig. 3A,C) nor in any other ligase structure which has been solved to date. This fold differs slightly from T4 Rnl1 in that it possesses an additional N-terminal helix, but otherwise can be structurally aligned reasonably well (RMSD 3.29 Å over 41 aligned Cα atoms). The N-terminal domains of both T4 Rnl1 and Pab1020 are similarly positioned with respect to the catalytic domains. The relatedness of the catalytic and N-terminal domains of T4 Rnl1 with those of Pab1020 suggests to us that the latter is an RNA ligase.

The amino-terminal section of the Pab1020 polypeptide appears to be the most mobile domain of the protein, with poor electron density for some of the side chains, and higher B-factors than found in the other domains. It is notable that sequences corresponding to either one or both of the first two α-helices are missing in some orthologs of Pab1020 (for example Q9YBN1, *Aeropyrum pernix*; Supplemental Fig. S1) indicating that these residues are probably functionally nonessential.

Pab1020 possesses two domains C-terminal to the nucleotidyl-transferase domain which have not been hitherto observed in ligase structures. The first is all α-helical, with no significant structural orthologs in PDB. T4 Rnl1 and T4 Rnl2 both have helical C-terminal domains but of different topology to that of Pab1020 and these domains are unlikely to be functionally equivalent. The final carboxy-terminal domain of Pab1020 has similarity to the copper-binding domain of amyloid precursor protein (APP) (Barnham et al. 2003; Kong et al. 2007) (Z-score 2.6, RMSD 3.07 Å over 46 residues, PDB code 1OWT), but the metal binding residues are not conserved in Pab1020. BLAST searches (Altschul et al. 1997) of public sequence databases indicate that the presence of both the dimerization domain and the second C-terminal domain of Pab1020 is specific to this protein, as well as the previously identified archaeal and bacterial orthologs (Ho and Shuman 2002), but we have not been able to identify homologous domains from eukarya.

Dimerization

Analytical gel filtration showed that the apparent molecular mass of Pab1020 was ~93 kDa, consistent with its formation of a homodimer in solution (Supplemental Fig. S2B,C). Analysis of the atomic model of Pab1020 with the program PISA (Krissinel and Henrick 2007) revealed that the largest interface between protomers in the crystals is of ~2194.5 Å² and connects two monomers that are related by non-crystallographic dyad symmetry, providing strong evidence that Pab1020 is a homodimer.

At the dimer interface, the most conserved residue is Gly296, which is strictly conserved among the homologs presented in Supplemental Figure S1. Interestingly, this

Figure 3. Comparison of Pab1020 monomer with other RNA ligases. Comparison of the domain organization of the closest crystal structures of Pab1020. (A) Pab1020, colored according to Figure 1. (B) T4 Rnl1 (PDB code 2C5U) (El Omari et al. 2006). (C) RNA editing ligase from *Trypanosoma brucei* (PDB 1XDN) (Deng et al. 2004). (D) T4 Rnl2 (PDB 2HVQ) (Nandakumar et al. 2006). (E) A structural alignment of the major motifs of polynucleotide ligases. Residues that were found to abrogate ligation activity of T4 Rnl1 are highlighted with a yellow background. Below, a sequence logo generated from these alignments displays the relative amino acid frequency for each motif.

residue generates a kink in helix $\alpha 10$, but remains in the helical region of the Ramachandran plot. However, the close proximity of the second molecule at this position means that mutation of this residue to even alanine is not tolerated, and would generate a significant clash with the backbone of the opposite molecule. Hence, formation of a dimer appears to have been a driving force in the conservation of this residue during the evolution of this enzyme.

The polypeptide chain of each monomer terminates in close proximity to the active site of the partner molecule. In each monomer, the carboxy-terminal domain of Pab1020 ends in a short β -strand that forms an interaction

with the second protomer, generating a short “3D domain swap” (Figs. 2A,B, 4B; Liu and Eisenberg 2002). However, the low level of sequence conservation in this sequence implies that such intertwining of the chains is not necessary for the function of ligases of this type (Supplemental Fig. S1).

The active site

In order to unambiguously identify the ligand at the active site of Pab1020, a simulated annealing composite omit map was calculated using CNS (Brünger et al. 1987). The resulting map demonstrates that there is clear and continuous density at each of the active sites which is consistent with AMPPNP molecules (Fig. 4C). The two NCS-related nucleotides are nearly identical, and are in a similar orientation to that of the AMPcPP [adenosine 5'-(α,β -methylenetriphosphate)] at the active site of T4 Rnl1 (Fig. 4D). The nucleosides of the AMPPNP molecules are also in a similar position to the adenyl group of T4 Rnl2, but due to the inversion at the β -phosphate, the pyrophosphate leaving group of Pab1020 points in an almost opposite direction to the reactive lysine of T4 Rnl2 (Fig. 4E; El Omari et al. 2006).

Adenylation of ATP-dependent ligases occurs by nucleophilic attack by the ζ -nitrogen of the reactive lysine on the α -phosphate of the nucleotide. Sequence comparisons of Pab1020 with other ligases suggest that the catalytic lysine is residue 95, and in the archaeal homologues of Pab1020, this residue is invariant (Fig. 3B, Supplemental Fig. S1). In this structure, the ζ -nitrogen of the Lys95 side chain is ~ 3.4 Å away from the α -phosphate of the nucleotide, well positioned for nucleophilic attack and formation of the phosphoramidate product. This residue is also in a similar position and orientation to the adenylated lysine of T4 Rnl2, thus it is likely to be the site of adenylation of Pab1020 (Fig. 4B,E).

The nucleotide analog is stabilized by a dense network of contacts around it, via interactions mediated by both lobes of the nucleotidyl-transferase domain. The AMPPNP molecule is in *syn* conformation, and is clamped on both faces of the nucleotide by hydrophobic side chains; the aromatic ring of Phe172 of motif IIIA (Fig. 4A) stacks onto one face, while the side chain of Ile230 (motif IV) contacts the opposite face in a similar manner to aromatic-purine-hydrophobic sandwiches which have been observed in other nucleotidyl transferases (Håkansson et al. 1997).

Further van der Waals contacts with the nucleotide moieties are contributed by Val96, which contacts the adenine via its backbone oxygen and nitrogen atoms. Hydrogen bonds are also formed between the exocyclic N6 atom of the purine with the main-chain oxygen of Glu94 as well as with an ϵ -oxygen of Glu93. Therefore,

Figure 4. The AMPPNP molecule bound at the catalytic site. (A) Amino acid residues contributed from motifs I, III, IIIA, IV, and V (colored gray) form a dense network of bonds with the nucleotide analog AMPPNP (colored sticks). (B) Similarity of T4 Rnl1 and Pab1020 around their superposed nucleotide-binding sites. Carbon atoms and cartoons are colored green (Pab1020 molecule B), orange (Pab1020 molecule A), and gray T4 Rnl1 (Krissinel and Henrick 2004) (PDB code 2C5U). Side chains of residues Arg54, Lys75, Glu159, and Asp272 of T4 Rnl1 are in similar positions to Arg115B, His70B, and Glu147B, as well as Asp382A of Pab1020, shown as sticks. The Mg^{2+} ion at the active site of T4 Rnl1 is also shown in gray. (C) A simulated annealing $2F_o - F_c$ composite omit map section contoured at 1.5σ around the nucleotide of Pab1020 molecule B. (D,E) Superposition of (D) AMPcPP from T4 Rnl1 (El Omari et al. 2006) (PDB code 2C5U) on the nucleotide bound to Pab1020; (E) the adenylated Lys35 of T4 Rnl2 (Nandakumar et al. 2006) (PDB code 2HVQ) on Lys95 and the bound nucleotide of Pab1020. See text for further discussion.

contacts with the nucleotide are via a combination of both amino acid main-chain and side-chain atoms. This is noteworthy since only main-chain atoms contact the nucleosides of T4 RNA ligases 1 and 2 (El Omari et al. 2006; Nandakumar et al. 2006). Evidently, nucleotide recognition by Pab1020 differs markedly from the T4 phage RNA ligases in this respect.

The carboxylate side chain of the motif III Glu147 interacts with the 2' oxygen of the ribose in an almost identical manner to glutamates 159 and 99 of T4 RNA ligases 1 and 2, respectively (Fig. 4; El Omari et al. 2006; Nandakumar et al. 2006). The γ -phosphate is contacted by His70, which is the most amino-terminal of the residues to contact the nucleotide, and is in a similar position to Lys75 of T4 Rnl1. Lys75 of T4 Rnl1 has been shown to be essential for ligase activity (Wang et al. 2003) and therefore it is likely that this residue will be similarly indispensable to the activity of Pab1020.

Arg115 of Pab1020 forms additional hydrogen bonds with the cofactor via both the ribose 3' oxygen as well as a bi-dentate interaction with an oxygen atom of the γ -phosphate. The guanidinium group of Arg115 is in a similar location to that of Arg54 of T4 Rnl1, performing hydrogen-bonding interactions with the β - and γ -phosphates of the nucleotide. This latter residue has been shown to be essential for step 2 of the ligation reaction (RNA adenylation), but was dispensable for steps 1 (enzyme adenylation) and 3 (phosphodiester-bond formation) (Wang et al. 2003). It is therefore possible that Arg115 of Pab1020 performs a similar role in step 2 chemistry of the ligation reaction.

In motif V of Pab1020, Lys241 contacts the α -phosphate of the AMPPNP residue (Fig. 4A), and again a similar contact has been observed in the structure of T4 Rnl1 (El Omari et al. 2006). This residue is highly conserved among the ligase structures that have been determined to date (Fig. 3E), and the corresponding residue (Lys240) has been shown to be essential for T4 Rnl1 activity, since its mutation to alanine prevented ligation from occurring. In this case, it was shown that the overall ligation and enzyme adenylation reactions were abrogated, even though the third step, phosphodiester-bond formation of a pre-adenylated substrate, could still occur in isolation (Wang et al. 2003). Therefore, it is likely that this lysine-mediated coordination of the α -phosphate plays an important role in stabilization of a transition state during transfer of the nucleotide to the active-site lysine and/or to the substrate RNA.

We do not observe density corresponding to metal ions at the active sites of Pab1020, despite having included magnesium chloride in the crystallization solution. Such ions have previously been observed at the active sites of both T4 Rnl1 and T4 Rnl2 (El Omari et al. 2006; Nandakumar et al. 2006). The lack of metal ions may be

due to the conformation of the phosphates of the ATP analog used in these crystallization trials, and a different orientation of the β - and γ -phosphates, such as that observed for AMPcPP at the active site of T4 Rnl1, may allow improved coordination of metal ions (Fig. 4D). In the case of T4 Rnl1, the magnesium is contacted by Gly269 and Asp272, the latter residue acting to position it correctly (Fig. 4B). Interestingly, Asp382 of the second protomer of the homodimer of Pab1020 is situated at a similar location to Asp272 of T4 Rnl1 (Fig. 4B). However, if it also serves to position a magnesium ion adjacent to the phosphates, one would expect it to be strictly conserved among orthologs of Pab1020, which is not the case (Supplemental Fig. S1).

Site of interaction with nucleic acid

The position of the nucleotide analog at the catalytic center localizes the site of nucleic acid binding on each Pab1020 monomer, but details of the mode of binding are otherwise unclear. The only RNA ligases whose structures have been determined thus far are not likely to represent similar binding mechanisms to Pab1020 because of the divergent C termini (Nandakumar et al. 2006). Instead, examination of the electrostatic potential surface of Pab1020 shows that there is a continuous, charged surface which spans the Pab1020 dimer, crossing at an angle in the cleft between the two N-terminal domains and close to the two nucleotide binding pockets (Supplemental Fig. S3). Therefore, the nucleic acid may cross the dimer from one active site to the other and contact available positively charged residues of the ligase via the phosphate backbone.

Circularization of RNA

The structural similarity of Pab1020 to T4 Rnl1 prompted us to assay Pab1020 for RNA adenylation and ligase activity. Addition of Pab1020 to 5'-labeled 18-mer RNA substrate yielded a partial product that migrated slightly more slowly than the input RNA, whereas the addition of Pab2002 did not (Fig. 1B). The small shift in electrophoretic mobility of RNA caused by incubation with Pab1020 is consistent with the formation of an adenylylated product. Incubation of Pab1020 with DNA did not ligate this substrate (Fig. 1C), showing that it is not a DNA ligase. In a control reaction in which a known DNA ligase (Pab2002) was input, a ligated species was indeed produced.

Incubation of Pab1020 with an 18-mer RNA substrate yielded not only an adenylylated product but a faster migrating species, albeit with a far lower yield (Fig. 5A). The observation that this product was resistant to treatment with alkaline phosphatase (data not shown), which catalyzes the hydrolysis of 5'-phosphate groups

Figure 5. Pab1020 adenylation reaction and circularization activity. (A) Ligation assays were performed as described in the Materials and Methods section, with 8 pmol of Pab1020 at 20°C, 85°C, 90°C, and 95°C. Pab1020 was found to catalyze the formation of adenylylated RNA oligo (AppRNA) and the RNA circle under the wide range of temperatures. (B) Effect of ATP on the formation of Pab1020 catalysis. The reaction mixtures contained 1 pmol of labeled RNA substrate and 2 pmol of Pab1020 in buffer R and various amounts of ATP (5 μ M, 10 μ M, 25 μ M, 50 μ M, 100 μ M, 500 μ M, and 1 mM) were incubated for 30 min at 75°C. The relative level of AppRNA (adenylylated RNA oligo) and RNA circle were expressed as percentages of the total labeled products, and plotted as function of ATP concentration.

from nucleic acid, demonstrated that the radiolabel was inaccessible post-reaction. It therefore is likely that this band corresponds to the product of the intra-molecular sealing reaction between the 5' and 3' ends of the substrate RNA, which forms a covalently closed RNA circle. It is also notable that no inter-molecular ligation was observed in these assays and the only reaction observed was of circularization. T4 Rnl2, for example, forms predominantly circular products when incubated with 18-mer substrate RNAs, although a concatenated product becomes the predominant species when the substrate used is shortened to 12 nucleotides (Yin et al. 2004).

The most likely reasons for the low yield of even this circular RNA product are twofold: Firstly, inclusion of

Table 1. X-ray data collection and refinement statistics

	SAD data set ^a (ID14-EH4)	Native data set (ID14-EH3)
Wavelength (Å)	0.9796	0.9310
Unit cell parameters (Å)	$a = 59.32, b = 104.03, c = 153.15$	$a = 58.96, b = 103.26, c = 150.93$
Space group	$P2_12_12_1$	$P2_12_12_1$
Resolution limits (Å) ^b	20.0–3.4 (3.6–3.4)	20.0–2.9 (3.1–2.9)
Number of observations measured ^b	188,288 (30,422)	100,621 (18,091)
Number of unique reflections measured ^b	24,911 (3936)	20,985 (3739)
Completeness (%) ^b	98.9 (99.0)	99.3 (99.2)
$\langle I \rangle / \langle \sigma(I) \rangle$ ^b	13.55 (4.34)	10.76 (2.33)
R_{sym} (%) ^{b,c}	10.7 (52.8)	13.2 (69.0)
Refinement statistics		
Number of non-hydrogen atoms (protein/water/other)		6094/10/118
Resolution range (Å)		20–2.9
R/R_{free} (%)		0.227/0.267
RMSD bonds (Å)/angles (°)		0.010/1.273
$\langle B \rangle$ (Å ²) Protein/AMPPNP/MPD/water		56.5/40.1/74.6/32.3

^aFriedel mates were treated as separate reflections.^bValues in parentheses refer to the highest resolution shell.^c $R_{\text{sym}} = \sum_h \sum_i |I_{hi} - \langle I \rangle_h| / \sum_h \sum_i I_{hi}$, where $\langle I \rangle_h$ is the mean intensity for reflection I_h and I_{hi} is the intensity of an individual measurement of reflection I_h .

ATP in the reaction may cause trapping of the reaction at the adenylylated RNA stage, as has previously been observed for T4 Rnl2 (Yin et al. 2004). The concentration of ATP used in these reactions is 10-fold higher than those of the enzyme, and it may be instructive to perform similar reactions with a pre-adenylylated RNA but without additional ATP. Secondly, because the RNA substrate used in this assay is probably not of the correct sequence or structure for correct ligation by Pab1020, the reaction may be suboptimal. Indeed, we currently do not know the cellular substrate for this enzyme.

The adenylation reaction showed a correlation between ATP concentration and formation of adenylylated product, in a similar way to the previously characterized ATP dependent adenylation of vibriophage KVP40 (Fig. 5B; Yin et al. 2004). Conversely, the production of the circular form of RNA correlates inversely with ATP concentration. Again, this may be indicative of a competitive inhibition of strand sealing by the nucleotide substrate.

Discussion

Both the similarity of the structure of Pab1020 to T4 Rnl1 and T4 Rnl2, and the fact that it has RNA ligase activity, indicate that it is an RNA ligase. The exact physiological substrate of Pab1020 has not yet been identified. However, the presence of pseudo-twofold symmetry of Pab1020 may imply that its substrate is either symmetrical or pseudo-symmetrical. Hence, we reason that it is possible that ligation is performed on a double-stranded rather than single-stranded substrate, perhaps by repairing two nicks concomitantly, one at each of the twin catalytic centers. The substrate nucleic acid may cross the dimer

symmetry axis, coincidental with its own axis of dyad (pseudo)-symmetry. Suitable double-stranded substrates would then include ribosomal RNA, transfer RNA, or indeed any other RNA molecule which possesses double-stranded elements of secondary structure.

To identify potential substrates of Pab1020, one may consider the roles of its structural homologs, because these may be involved in related processes. The protein with the highest structural similarity to Pab1020, T4 Rnl1, is able to ligate tRNA half molecules (Wang et al. 2007). Many tRNA molecules contain sequences in their anticodon stems which must be spliced out before the mature tRNA is produced. This is performed by the “splicing nuclease,” which is a pseudo-symmetrical complex that recognizes the cleavage sites at the junctions between the tRNA half molecule and its intervening sequence (“IVS” or “intron”) (Xue et al. 2006). In archaea, the cleavage site that is recognized has a characteristic pseudo-symmetrical secondary structure, termed the “bulge-helix-bulge” (BHB) motif, and the splicing endonuclease cleaves the phosphate backbone once at each of the bulges. Since this BHB recognition sequence is also pseudo-symmetrical, it may be a suitable substrate for this homodimeric *P. abyssi* RNA ligase (Marck and Grosjean 2003).

However, the question remains, what purpose does dimerization serve for the function of Pab1020? It seems plausible to us that a secondary reaction may occur concomitantly with the primary ligation event. An example of this type of symmetrical ligation reaction has been demonstrated in the archaeon *Haloferax volcanii*; it has been shown that during the splicing reaction of its tRNA, the intron is circularized (d’Orval et al. 2001; Salgia et al.

2003). Introns are also found in archaeal ribosomal RNA (rRNA), and during religation of spliced rRNA; and the circularization of these sequences has also been documented (Kjems and Garrett 1988; Burggraf et al. 1993).

It is then tempting to speculate that dimerization of Pab1020 may serve to perform not solely ligation of RNA such as tRNA, but also a secondary reaction, for example intron circularization, which has been documented in some organisms, albeit not to our knowledge in *P. abyssi*. RNA circularization may impart protection from digestion by exonucleases or prevent “breathing” of nucleic acid extremities that would otherwise occur, as has been suggested with respect to the circular box C/D RNAs of *P. furiosus* (Starostina et al. 2004).

In the structures of capping enzymes and DNA ligases, the domain downstream from the nucleotidyl-transferase domain frequently consists of an OB (oligonucleotide-binding) fold. The fold of the downstream domains in Pab1020 and both T4 RNA ligases is unrelated to the OB-domain. However, despite the lack of structural similarity, there is a recurring theme; the additional structures impart specificity to the ligase in question. In other members of the nucleotidyl-transferase superfamily, such as human DNA ligase I, the OB-fold has been shown to be essential for nucleic acid recognition (Pascal et al. 2004), and a similar role has been proposed for the C-terminal domain of T4 Rnl2, since its deletion abrogated RNA-binding activity (Nandakumar et al. 2006). T4 Rnl1 ligates tRNA half-molecules in preference to the circularization of the IVS, but this preference is negated when the C-terminal domain of the ligase is deleted, underlining the importance of this structure in tRNA processing (Wang et al. 2007).

Pab1020 possesses yet another domain structure and represents a further variation of the ligase core enzyme which has become functionally differentiated by the addition of supplementary structural elements. The functional role of these domains is not yet clear, but it is likely that these also confer substrate specificity as has been demonstrated in the preceding RNA ligase structures. Further experimentation is required before the exact nature of the specificity of this enzyme and its physiological function in this archaeon may be elucidated, and this work is under way.

Materials and Methods

Nucleotidyl transferase assays

Reactions containing varying amounts of [α - 32 P]NTP were incubated in a total volume of 20 μ L using the following buffer: 50 mM Tris-HCl, pH 7, 5 mM DTT and 1 mM MgCl₂, for 15 min at 70°C. Thirteen percent SDS-polyacrylamide gels were dried on filter paper (Whatman) and analyzed using a phosphor-imager to detect the formation of the covalent enzyme-NMP product (Molecular Dynamics).

RNA ligation assays

Activity assays using either Pab2002 or Pab1020 were performed using the following reaction buffer: 50 mM Tris-acetate, pH 6.5; 5 mM dithiothreitol (DTT), and 5 mM MgCl₂ in a total volume of 20 μ L. In assays, ATP and protein concentrations were systematically varied and were indicated in each figure. Samples were incubated at 75°C or at 50°C (other temperatures are indicated in each figure) for 30 min with 1 pmol of DNA or RNA substrates. All enzymatic reactions were stopped with stop buffer containing 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol. Reaction substrates and products were subjected to poly-acrylamide gel electrophoresis on denaturing 7 M urea, 18% or 20% gels (Sambrook et al. 1989). Radioactive signals were detected and quantitated with a Storm imaging system (Molecular Dynamics).

Crystallogenes

Following concentration of Pab1020 to 5–10 mg/mL, AMPPNP (Sigma) and magnesium chloride were added in fivefold molar excess. After initial crystallization condition screening using the vapor-diffusion method, a condition containing between 40% and 60% methane pentane-diol (MPD), 0–0.2 M NH₄H₂PO₄, and 0.1 M Tris-Cl, pH 8.5 was identified. Crystals of Pab1020 were also prepared which were labeled with selenium (Doublé 1997).

Experimental phasing

Structure solution was performed by means of a single-wavelength anomalous dispersion (SAD) experiment, using SHELXD (Schneider and Sheldrick 2002) to locate the heavy atom substructure. These sites were then refined using SHARP (de La Fortelle and Bricogne 1997) which was also used for calculation of initial phases. Solvent flipping and density modification performed by the programs SOLOMON and DM, respectively, allowed traceable maps to be generated (Cowtan 1994; Abrahams and Leslie 1996). Refinement was performed using CNS (Brünger et al. 1987), and using REFMAC (Murshudov et al. 1999), as described in the Supplemental material. The atomic coordinates have been deposited into the Protein Data Bank with the accession number 2VUG.

Acknowledgments

We thank the ESRF for provision of beam time on ID14-3 and ID14-4 in BAG MX554. We additionally thank Emmanuela Fioravanti and Raimond Ravelli for beamline support and Nathalie Ulrick for preparation of crystals. H.M. acknowledges the INSERM AVENIR program and Fondation Bettencourt Schueller for financial support.

References

- Abrahams, J.P. and Leslie, A.G. 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D Biol. Crystallogr.* **52**: 30–42.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, Z.W.M., and Lipman, D.J. 1997. Gapped Blast and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Barnham, K.J., McKinstry, W.J., Multhaup, G., Galatis, D., Morton, C.J., Curtain, C.C., Williamson, N.A., White, A.R., Hinds, M.G., Norton, R.S., et al. 2003. Structure of the Alzheimer's disease amyloid precursor protein

- copper binding domain. A regulator of neuronal copper homeostasis. *J. Biol. Chem.* **278**: 17401–17407.
- Brünger, A.T., Kuriyan, J., and Karplus, M. 1987. Crystallographic R factor refinement by molecular dynamics. *Science* **235**: 458–460.
- Burggraf, S., Larsen, N., Woese, C.R., and Stetter, K.O. 1993. An intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. *Proc. Natl. Acad. Sci.* **90**: 2547–2550.
- Cowtan, K. 1994. “dm”: An automated procedure for phase improvement by density modification. *Newsletter on Protein Crystallography* **31**: 34–38.
- de La Fortelle, E. and Bricogne, G. 1997. Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**: 472–494.
- Deng, J., Schnauffer, A., Salavati, R., Stuart, K.D., and Hol, W.G.J. 2004. High resolution crystal structure of a key editosome enzyme from *Trypanosoma brucei*: RNA editing ligase 1. *J. Mol. Biol.* **343**: 601–613.
- d’Orval, B.C., Bortolin, M.L., Gaspin, C., and Bachellerie, J.P. 2001. Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA^{Trp} intron guides the formation of two ribose-methylated nucleosides in the mature tRNA^{Trp}. *Nucleic Acids Res.* **29**: 4518–4529.
- Doublé, S. 1997. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**: 523–530.
- El Omari, K., Ren, J., Bird, L.E., Bona, M.K., Klarmann, G., LeGrice, S.F.J., and Stammers, D.K. 2006. Molecular architecture and ligand recognition determinants for T4 RNA ligase. *J. Biol. Chem.* **281**: 1573–1579.
- Esnouf, R.M. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**: 132–134.
- Håkansson, K., Doherty, A.J., Shuman, S., and Wigley, D.B. 1997. X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell* **89**: 545–553.
- Ho, C.K. and Shuman, S. 2002. Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc. Natl. Acad. Sci.* **99**: 12709–12714.
- Ho, C.K., Etten, J.L.V., and Shuman, S. 1997. Characterization of an ATP-dependent DNA ligase encoded by Chlorella virus PBCV-1. *J. Virol.* **71**: 1931–1937.
- Kjems, J. and Garrett, R.A. 1988. Novel splicing mechanism for the ribosomal RNA intron in the archaeobacterium *Desulfurococcus mobilis*. *Cell* **54**: 693–703.
- Konarska, M., Filipowicz, W., Domdey, H., and Gross, H.J. 1981. Formation of a 2'-phosphomonoester, 3',5'-phosphodiester linkage by a novel RNA ligase in wheat germ. *Nature* **293**: 112–116.
- Kong, G.K.W., Adams, J.J., Harris, H.H., Boas, J.F., Curtain, C.C., Galatis, D., Masters, C.L., Barnham, K.J., McKinstry, W.J., Cappai, R., et al. 2007. Structural studies of the Alzheimer's amyloid precursor protein copper-binding domain reveal how it binds copper ions. *J. Mol. Biol.* **367**: 148–161.
- Krissinel, E. and Henrick, K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 2256–2268.
- Krissinel, E. and Henrick, K. 2007. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**: 774–797.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**: 283–291.
- Lee, J.Y., Chang, C., Song, H.K., Moon, J., Yang, J.K., Kim, H.K., Kwon, S.T., and Suh, S.W. 2000. Crystal structure of NAD⁺-dependent DNA ligase: Modular architecture and functional implications. *EMBO J.* **19**: 1119–1129.
- Liu, Y. and Eisenberg, D. 2002. 3D domain swapping: As domains continue to swap. *Protein Sci.* **11**: 1285–1299.
- Marck, C. and Grosjean, H. 2003. Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: Evolutionary implications. *RNA* **9**: 1516–1531.
- Martins, A. and Shuman, S. 2004. An RNA ligase from *Deinococcus radiodurans*. *J. Biol. Chem.* **279**: 50654–50661.
- Meslet-Cladière, L., Norais, C., Kuhn, J., Briffotiaux, J., Sloostra, J.W., Ferrari, E., Hübscher, U., Flament, D., and Myllykallio, H. 2007. A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen. *J. Mol. Biol.* **372**: 1137–1148.
- Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., and Dodson, E.J. 1999. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 247–255.
- Nakatani, M., Ezaki, S., Atomi, H., and Imanaka, T. 2000. A DNA ligase from a hyperthermophilic archaeon with unique cofactor specificity. *J. Bacteriol.* **182**: 6424–6433.
- Nandakumar, J., Shuman, S., and Lima, C.D. 2006. RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. *Cell* **127**: 71–84.
- Nandakumar, J., Nair, P.A., and Shuman, S. 2007. Last stop on the road to repair: Structure of *E. coli* DNA ligase bound to nicked DNA-adenylate. *Mol. Cell* **26**: 257–271.
- Pascal, J.M. 2008. DNA and RNA ligases: Structural variations and shared mechanisms. *Curr. Opin. Struct. Biol.* **18**: 96–105.
- Pascal, J.M., O'Brien, P.J., Tomkinson, A.E., and Ellenberger, T. 2004. Human DNA ligase I completely encircles and partially unwinds nicked DNA. *Nature* **432**: 473–478.
- Sabatini, R. and Hajduk, S.L. 1995. RNA ligase and its involvement in guide RNA/mRNA chimera formation. Evidence for a cleavage-ligation mechanism of *Trypanosoma brucei* mRNA editing. *J. Biol. Chem.* **270**: 7233–7240.
- Salgia, S.R., Singh, S.K., Gurha, P., and Gupta, R. 2003. Two reactions of *Haloflex volcanii* RNA splicing enzymes: Joining of exons and circularization of introns. *RNA* **9**: 319–330.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular cloning: A laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schneider, T.R. and Sheldrick, G.M. 2002. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**: 1772–1779.
- Shuman, S. and Schwer, B. 1995. RNA capping enzyme and DNA ligase: A superfamily of covalent nucleotidyl transferases. *Mol. Microbiol.* **17**: 405–410.
- Silber, R., Malathi, V.G., and Hurwitz, J. 1972. Purification and properties of bacteriophage T4-induced RNA ligase. *Proc. Natl. Acad. Sci.* **69**: 3009–3013.
- Sriskanda, V., Moyer, R.W., and Shuman, S. 2001. NAD⁺-dependent DNA ligase encoded by a eukaryotic virus. *J. Biol. Chem.* **276**: 36100–36109.
- Starostina, N.G., Marshburn, S., Johnson, L.S., Eddy, S.R., Terns, R.M., and Terns, M.P. 2004. Circular box C/D RNAs in *Pyrococcus furiosus*. *Proc. Natl. Acad. Sci.* **101**: 14097–14101.
- Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. *Science* **278**: 631–637.
- Tomkinson, A.E., Vijayakumar, S., Pascal, J.M., and Ellenberger, T. 2006. DNA ligases: Structure, reaction mechanism, and function. *Chem. Rev.* **106**: 687–699.
- Wang, L.K., Ho, C.K., Pei, Y., and Shuman, S. 2003. Mutational analysis of bacteriophage T4 RNA ligase 1. Different functional groups are required for the nucleotidyl transfer and phosphodiester bond formation steps of the ligation reaction. *J. Biol. Chem.* **278**: 29454–29462.
- Wang, L.K., Nandakumar, J., Schwer, B., and Shuman, S. 2007. The C-terminal domain of T4 RNA ligase 1 confers specificity for tRNA repair. *RNA* **13**: 1235–1244.
- Westaway, S.K., Phizicky, E.M., and Abelson, J. 1988. Structure and function of the yeast tRNA ligase gene. *J. Biol. Chem.* **263**: 3171–3176.
- Xue, S., Calvin, K., and Li, H. 2006. RNA recognition and cleavage by a splicing endonuclease. *Science* **312**: 906–910.
- Yin, S., Ho, C.K., Miller, E.S., and Shuman, S. 2004. Characterization of bacteriophage KVP40 and T4 RNA ligase 2. *Virology* **319**: 141–151.