Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a putative 2′-5′ RNA ligase from *Pyrococcus horikoshii*

Peter H. Rehse and Tahir H. Tahirov*‡

RIKEN Harima Institute, 1-1-1 Kouto, Mikazuki-cho, Sayo-gun, Hyogo 679-5148, Japan

‡ Current address: Eppley Institute for Research in Cancer and Allied Diseases, LTC Room 10737A, University of Nebraska Medical Center, 986805 Nebraska Medical Center, Omaha, NE 68198-7696, USA.

Correspondence e-mail: tahir@spring8.or.jp

Received 13 April 2005
Accepted 6 June 2005

PDB Reference: putative 2′-5′ RNA ligase, 1vdx, r1vdxsf.

Cyclic phosphodiesterase and 2′-5′ RNA ligase are members of a superfamily of proteins which share structural similarities even though their homology may be very low. A putative 2′-5′ RNA ligase from *Pyrococcus horikoshii* has been crystallized and its X-ray crystallographic structure determined to 2.4 Å. The protein crystallized in the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 44.07, b = 45.47, c = 93.17 Å and one protein monomer in the asymmetric unit. The molecular-replacement probe was a 2′-5′ RNA ligase from *Thermus thermophilus* which shares 30% sequence identity. The *P. horikoshii* RNA ligase has some structural features that have more in common with a cyclic phosphodiesterase from *Arabidopsis thaliana* with which it has no significant homology, yet an examination of the electrostatic surface potential clearly defines its relationship to the *T. thermophilus* RNA ligase. However, the size of the active-site cleft is smaller and less positively charged than that of the *T. thermophilus* homologue, suggesting that the actual substrate may be smaller than that previously postulated for the latter.

1. Introduction

RNA-ligase activity has been found in a wide range of organisms (Arn & Abelson, 1996). In eukaryotic species, most RNA-ligase functions have been attributed to known RNA-processing events such as intron removal from tRNA (Greer, Peebles *et al.*, 1983) or the maturation of mitochondrial RNA of trypanosomatids (Deng *et al.*, 2004). In bacteria, the activity was first encountered through the 2′-5′ ligation of *Saccharomyces cerevisiae* tRNA-splicing intermediates (Greer, Javor *et al.*, 1983) and an RNA ligase was first cloned and purified from *Escherichia coli* (Arn & Abelson, 1996). RNA-processing events requiring RNA-ligase activity have not been discovered in bacteria, although products that presumably arise from post-transcriptional processing by a 2′-5′ RNA-ligase activity have been found (Trujillo *et al.*, 1987). The *in vivo* function of this activity remains obscure. Unlike RNA ligases from eukaryotic sources, activity does not require a nucleoside triphosphate cofactor and the resultant phosphodiester bond is considered unusual.

The crystal structure of a 2′-5′ RNA ligase from *Thermus thermophilus* HB8 has been published (Kato *et al.*, 2003). The 2′-5′ RNA-ligase sequences are generally conserved among bacteria and archea and are characterized by two H*X*T*X* motifs, where *X* is a small hydrophobic residue. The proteins are all about 200 residues in length. An examination of the overall fold allowed Kato *et al.* (2003) to determine the protein's structural relatedness to a superfamily of enzymes with cyclic phosphodiesterase or related activities (Nasr & Filipowicz, 2000) which also contain the above-mentioned

H*X*T*X* motifs. The sequence identity between the 2′-5′ RNA ligases and the cyclic phosphodiesterases are very low at around 9%. Crystal structures of the latter enzyme from *Arabidopsis thaliana* (CPDase; Hofmann *et al.*, 2000, 2002) have been solved. Based on this structure and mutagenesis work carried out on the homologous protein from *S. cerevisiae* (Nasr & Filipowicz, 2000), Hofmann and coworkers proposed an enzymatic mechanism. This mechanism was adapted by Kato *et al.* (2003) to 2′-5′ RNA ligase.

Here, we report the three-dimensional structure of a 2′-5′ RNA ligase from *Pyrococcus horikoshii* which shares 30% identity with the *T. thermophilus* RNA ligase. The *P. horikoshii* and *T. thermophilus* RNA ligases share 27 and 31% sequence identity with the 2′-5′ RNA ligase from *E. coli*, respectively. Kinetic experiments have only been performed on the *E. coli* RNA ligase (Arn & Abelson, 1996). The relationship between the two 2′-5′ RNA-ligase structures with regard to overall fold, active site and sequence comparisons will be discussed.

Table 1
Processing and refinement statistics.

Values in parentheses are for the highest resolution shell.

Processing	
Resolution (Å)	39.84–2.40 (2.49–2.40)
R_{merge} (%)	8.0 (42.3)
$\langle I/\sigma(I)\rangle$	10.0 (3.36)
Completeness (%)	99.1 (98.5)
Multiplicity	3.41 (3.33)
Refinement	
Resolution (Å)	39.84–2.40 (2.55–2.40)
Reflections	7386
Reflections in R_{free}	552 (83)
Protein atoms	1488
Solvent atoms	41
Other atoms	1
R factor (%)	22.5 (27.7)
R_{free} (%)	27.4 (32.6)
Ramachandran plot, residues in	
Most favoured region (%)	91.6
Additionally allowed region (%)	7.8
Generously allowed region (%)	0.6
Disallowed region (%)	0.0
R.m.s. deviations	
Bonds (Å)	0.007
Angles (°)	1.40
Average B factor (Å^2)	
Main chain	41.1
Side chain	46.0
Solvent	42.5
Others	52.1
All	43.6

2. Materials and methods

2.1. Protein expression and purification

The polymerase chain reaction (PCR) was used for gene amplification of *P. horikoshii* OT3 genomic DNA. The recombinant plasmid was constructed using the super-rare-cutter system (Hayashizaki *et al.*, manuscript in preparation) and transformed into *E. coli* strain BL21-CodonPlus (DE3)-RIL cells. The cells were grown for 20 h at 310 K in 4.5 l medium. The cells (20.9 g) were harvested by centrifugation at 6500 rev min^{-1} for 5 min and suspended in 38 ml 20 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 5 m*M* 2-mercaptoethanol and 1 m*M* phenylmethylsulfonyl fluoride. The cells were disrupted by sonication followed by heat treatment at 363 K (*P. horikoshii* proteins are not expected to denature at this temperature) for 13 min. The cell debris and denatured proteins were removed by centrifugation (15 000 rev min^{-1}, 30 min, 277 K) and the supernatant was applied onto a HiPrep 26/10 desalting column (53 ml, Amersham Biosciences) using 20 m*M* Tris–HCl pH 8.0 (buffer *A*). The elutant was applied onto a SuperQ Topopearl 650M column (80 ml, Tosoh) equilibrated with buffer *A*. The fraction containing protein was eluted with a 0–0.3 *M* NaCl linear gradient. The main protein peak was desalted using a HiPrep 26/10 column with buffer *A* and applied onto a Resource Q column (6 ml, Amersham Biosciences) equilibrated in buffer *A* and eluted with a 0–0.2 *M* NaCl linear gradient. The main protein fraction was desalted with a HiPrep 26/10 column equilibrated in 10 m*M*

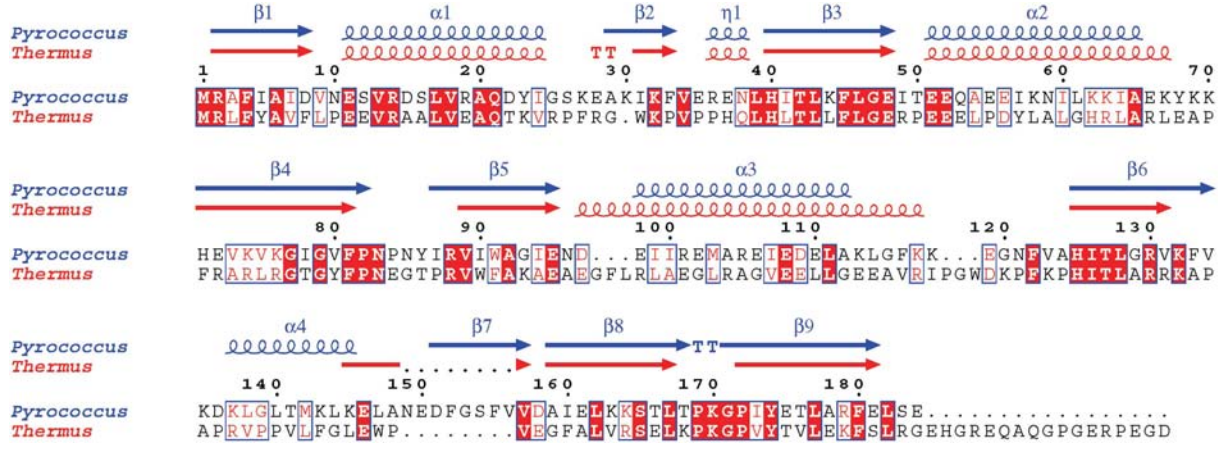

Figure 1
Sequence alignment of *P. horikoshii* RNA ligase (blue) with *T. thermophilus* RNA ligase (red). The identity is 30%. Alignment was performed using *CLUSTALW* (Thompson *et al.*, 1994), with minor adjustment of the *T. thermophilus* RNA-ligase sequence such that the strand corresponding to the *P. horikoshii* RNA ligase β7 was unbroken. This figure was prepared using *ESPript* (Gouet *et al.*, 1999).

sodium phosphate pH 7.0 (buffer *B*) and applied onto a Bio-Scale CHT-20-I column (Bio-Rad) equilibrated with buffer *B*. The fractions containing protein were eluted with a linear gradient of 10–200 m*M* potassium phosphate, pooled, concentrated by ultrafiltration (Vivaspin, 5 kDa cutoff) and loaded onto a HiLoad 16/60 Superdex 75pg column (120 ml, Amersham Biosciences) equilibrated with 20 m*M* Tris–HCl pH 8.0 and 0.2 *M* NaCl. The purified protein was homogeneous on SDS–PAGE. The protein was concentrated to 44.4 mg ml^{-1} by ultrafiltration (Vivaspin, 5 kDa cutoff).

2.2. Crystallization

Very small poorly shaped protein crystals were initially obtained by the microbatch method (Chayen *et al.*, 1990) using a TERA crystallization robot and a screening kit designed for high-throughput protein crystallization (Sugahara & Miyano, 2002). The initial conditions consisted of 100 m*M* MES pH 5.8 and 22.5%(*w*/*v*) PEG 4000. Diffraction-quality crystals were obtained under essentially the same conditions using sitting drops (1 µl protein solution plus 1 µl reservoir solution) and a slightly lower PEG concentration (20%). These crystals appeared overnight and grew into blocks (0.2 × 0.3 × 0.1 mm) over the course of a week.

2.3. Data collection

The solution surrounding the crystals was changed to cryoprotectant solution over the course of 10 min. The cryoprotectant buffer was identical to the reservoir buffer except for the addition of 30%(*v*/*v*) PEG 400. Initially, 8 µl reservoir solution was added to the drop, followed by increasing amounts of cryoprotectant solution (2–5 µl) coupled with the removal of excess drop solution. A single crystal was mounted in a cryoloop and flash-frozen at 100 K in an N_2 stream. Data were collected on a Rigaku model RU-300 rotating copper-anode diffractometer using an R-AXIS V image-plate detector. *DENZO* and *SCALEPACK* implemented in the *HKL*2000 program package (Otwinowski & Minor, 1997) were used to process this data. Data-collection statistics are summarized in Table 1.

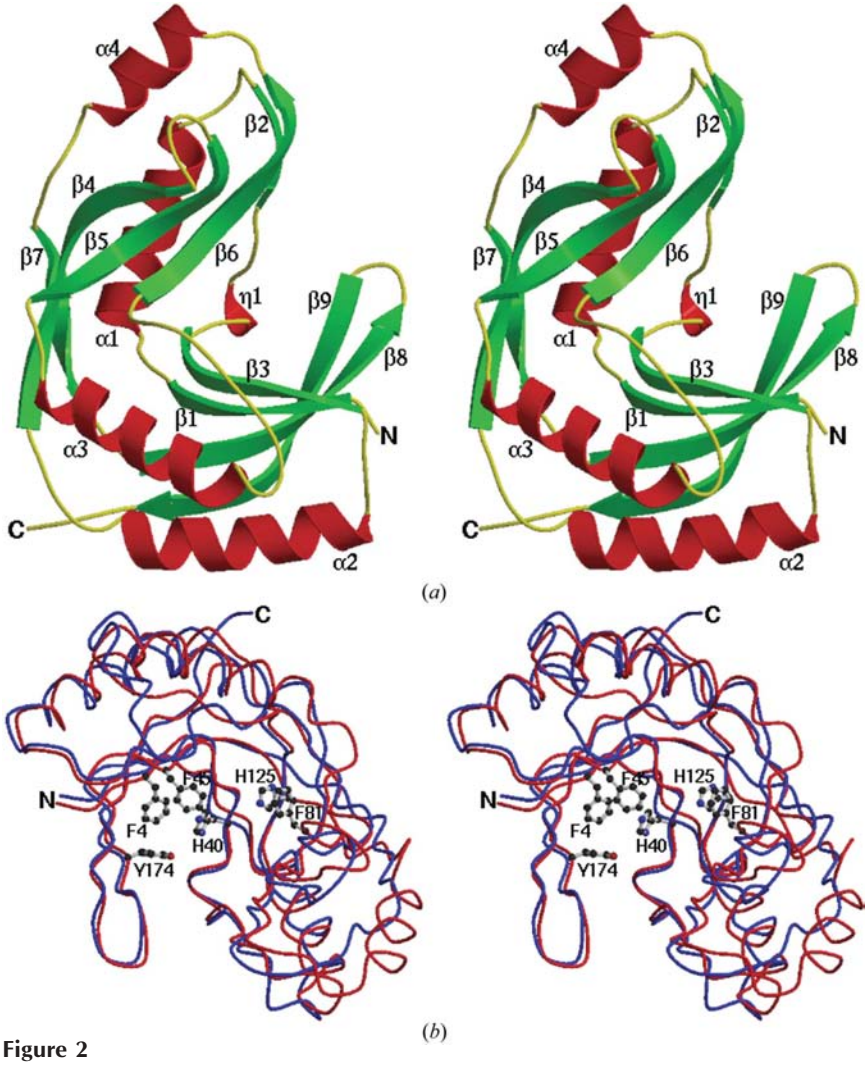

Figure 2
(*a*) Stereoview of the monomer structure of *P. horikoshii* RNA ligase. The β-sheet structures are in green and the helices in red. (*b*) Overlapped *P. horikoshii* RNA ligase (red) and *T. thermophilus* RNA ligase (blue) structures. Selected residues are from *P. horikoshii* RNA ligase.

2.4. Structure determination and model refinement

The structure was solved by molecular replacement using *MOLREP* (Collaborative Computational Project, Number 4, 1994; Vagin & Teplyakov, 1997) using RNA ligase from *T. thermophilus* (PDB code 1iuh; Kato *et al.*, 2003), with which it shares 30% sequence identity, as a model. The structure was built and modified using the program *QUANTA* (Accelrys). Refinement was carried out using *CNS* (Brünger *et al.*, 1998 and stereochemical analysis of the structure was performed using *PROCHECK* (Laskowski *et al.*, 1993). Refinement statistics and the model-quality assessment are summarized in Table 1. Figures were generated with *ESPript* (Gouet *et al.*, 1999), *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Bacon, 1997). Sequence alignments were performed using *CLUSTALW* (Thompson *et al.*, 1994) and secondary-structure analysis using *DSSP* (Kabsch & Sander, 1983). Homologous proteins were found using a *BLAST* (Altschul *et al.*, 1997) search of sequence and structure data bases.

3. Results and discussion

The *P. horikoshii* RNA ligase crystallized in the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 44.07, b = 45.47, c = 93.17 Å and one protein monomer in the asymmetric unit. The R factor for the final

model to 2.4 Å is 22.5%. The model consists of a single 184-residue monomer in the asymmetric unit with no missing atoms or residues. The model contains 1488 protein atoms, 41 water molecules and one chloride ion.

The maximum interaction between crystallographically related monomers results in a buried surface area of 10 006 Å^2 (calculated using *GRASP*; Nicholls *et al.*, 1991), which is well below the value described by Janin (1997) as being biologically significant.

3.1. Overall fold

The structure shown in Fig. 2 is characterized by two lobes of similar architecture coming together to form a β-barrel. As with similar folds found in the structures of CPDase (Hofmann *et al.*, 2000) and *T. thermophilus* RNA ligase (Kato *et al.*, 2003), the main chain weaves between the two domains to such an extent that lobes is a more appropriate term than domain. Following previous nomenclature, the 'terminal lobe' is that which contains the N- and C-terminal ends of the protein, as opposed to the 'transit lobe'. The *DSSP* program was used to calculate the secondary-structure elements (Kabsch & Sander, 1983) and the relationship to the sequence is shown in Fig. 1.

The 'terminal lobe' of *P. horikoshii* RNA ligase consists of a four-stranded antiparallel β-sheet ($\beta3\beta1\beta8\beta9$) and helices $\alpha2$ and $\alpha3$. The 'transit lobe' of the protein contains a five-stranded antiparallel β-sheet ($\beta2\beta6\beta5\beta4\beta7$) and helices $\alpha1$ and $\alpha4$. The formation of a β-barrel places all four helices on the outside of the structure. There is a relatively good correspondence in the sequence alignment of *T. thermophilus* RNA ligase with that of *P. horikoshii* (30% sequence identity) and the alignment of secondary structure (see Fig. 2*a*). The *P. horikoshii* RNA ligase contains an extra α-helix ($\alpha4$), while the third α-helix ($\alpha3$) is about 50% shorter. This is a similar relationship to that seen between CPDase and *T. thermophilus* RNA ligase, even though there is no significant sequence homology between the CPDase and either of the other two proteins. The extra $\alpha4$ helix is far removed from the active-site cleft, as is the $\alpha3$ helix. In contrast to the CPDase structure, both the *P. horikoshii* and *T. thermophilus* RNA-ligase structures have a short β-strand ($\beta2$) and 3_{10}-helix ($\eta1$) inserted after the first α-helix, providing a little extra structure for the 'transit lobe'. The $\eta1$ helix immediately precedes the first H*X*T*X* catalytic motif (see below). Finally, β-strands $\beta7$ and $\beta8$ are reported as a single strand in the *T. thermophilus* RNA ligase, whereas the *DSSP* program of Kabsch & Sander (1983) calculates there to be two strands in both structures. The two distinct strands are quite clear in the *P. horikoshii* RNA ligase and the β-strand numbering reflects this. A sequence alignment of *T. thermophilus* RNA ligase and CPDase based on secondary-structural elements and postulated active-site residues was attempted by Kato *et al.* (2003). Using this alignment, the r.m.s. deviation between the *T. thermophilus* RNA ligase and CPDase structures ranged between 2.46 and 3.19 Å.

The 15 C-terminal residues that were disordered in the *T. thermophilus* RNA ligase correspond to a sequence extension compared with the *P. horikoshii* RNA ligase. The full sequences are shown in Fig. 1.

Fig. 2(*b*) shows the overlapped *P. horikoshii* and *T. thermophilus* RNA-ligase structures. In order to maximize the agreement between the two central histidines (His40 and His75) the region 30–81 and 121–131 was used for alignment. There is very little structural variance in the active-site region, although there is much more elsewhere. Most notable is a loop protrusion just above residue Phe45 in Fig. 3, corresponding to a three-residue insertion between residues 118 and 119 of the *P. horikoshii* RNA ligase. The orientation of the side chains is such that there is no effect on the size of the active-site cleft. The overall r.m.s. deviation between these two structures is 0.88 Å.

3.2. Active site

The catalytic mechanism described by Kato *et al.* (2003) for the putative active site of *T. thermophilus* 2′-5′ RNA has the two H*X*T*X* motifs acting in concert. His130 activates the ribose hydroxyl group, which performs a nucleophilic attack on the cyclic phosphate. His39 assists the leaving of the phosphate group by providing protonation of one of the O atoms. The hydroxyl groups of the threonines from both motifs (Thr132, Thr41) and the guanidino group of Arg135 interact with the leaving phosphate group. A stereoview of the active site of *P. horikoshii* RNA ligase is given in Fig. 3, with residues of the *T. thermophilus* RNA ligase that differ in type or significantly in position shown in red. The two H*X*T*X*

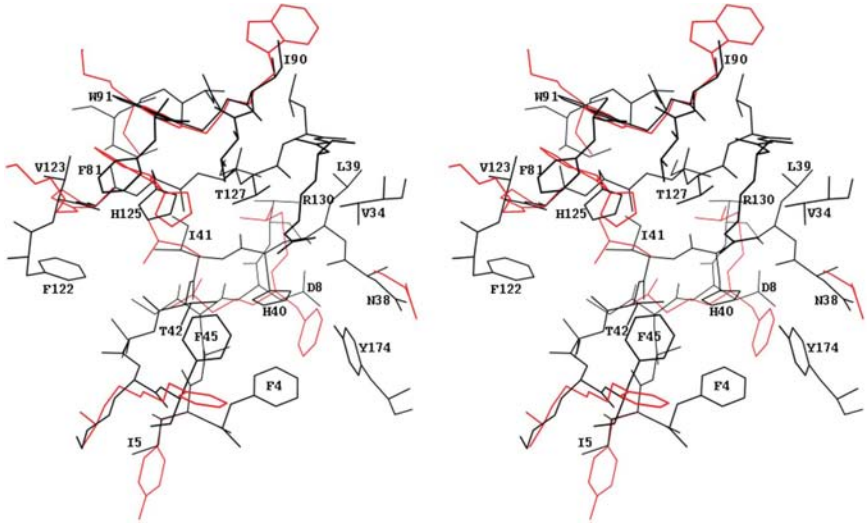

Figure 3
Active site of the *P. horikoshii* RNA ligase. Residues in red refer to *T. thermophilus* RNA-ligase residues that differed in type or significantly in position. Residue labelling refers to the *P. horikoshii* RNA ligase.

motifs come together so that they form a pseudo-twofold axis which is further enhanced by a clustering of aromatic groups (two or three) to side of the motif histidines (His40 and His125). The aromatic cluster associated with His40 is increased by one residue in the *T. thermophilus* RNA-ligase structure, with an Asp8Phe substitution and to a lesser extent Ile5Tyr. Residue Phe45 is found as a different rotamer in the *T. thermophilus* RNA ligase, possibly owing to the interaction with the latter changed residue. In *P. horikoshii* RNA ligase, His125 is found as a single rotamer interacting with Ile41. In *T. thermophilus* RNA ligase this isoleucine is mutated to leucine and the corresponding histidine is found as two rotamers, one being similar to that found in the *P. horikoshii* RNA ligase. The other rotamer would clash with Trp91. In the *T. thermophilus* RNA ligase, this residue corresponds to a phenylalanine. The guanidino group of Arg130, implicated in catalysis by Kato *et al.* (2003), approaches His40 and Thr127 in the same way as the *T. thermophilus* RNA ligase. Generally speaking, there is very little variation between the residues directly involved in catalysis; however, the active-site region is not completely conserved.

3.3. Surface potentials

The electrostatic surface potentials of the *P. horikoshii* and *T. thermophilus* RNA ligases and the CPDase are shown in Fig. 4. The drastic difference between the latter two, even though the active-site homology is high, was pointed out by Kato *et al.* (2003), who surmised that the large positive region was essential for recognizing the large negatively charged tRNA molecule. An examination of the *P. horikoshii* RNA ligase reveals a much narrower and deeper active-site cleft with a substantially smaller positively charged patch. This suggests that the substrate may in fact differ from tRNA, but not as much as a single cyclic nucleotide. As mentioned above, there is very little variation around the active site with respect to the main chain. The differences primarily arise from to side-chain type and orientation. For example, the increased positive charge of the *T. thermophilus* RNA ligase can be attributed to Glu37His and Gly120Lys changes, along with the movement of the Arg2 and Arg88 side chains deeper into the cleft. The narrowing of the cleft can to a large extent be attributed to the rotation of Phe45.

4. Conclusion

Based on active-site topology and an examination of the electrostatic surface potentials, the putative *P. horikoshii* RNA ligase described in this paper is clearly related to *T. thermophilus* RNA ligase. Both these proteins show a medium (30%) identity to an *E. coli* protein known to catalyze a 2′-5′ ligation of yeast tRNA. An examination of the genomes of *T. thermophilus* and *P. horikoshii* do not indicate proteins that are more homologous to the *E. coli* protein, suggesting that their function is the same. Kato *et al.* (2003) suggested that the topology of the *T. thermophilus* RNA-ligase active site could bind tRNA. This appears less likely with *P. horikoshii* RNA ligase, since its active-site cleft appears to be smaller and less positively charged, although still capable of binding single-stranded RNA. The exact role in bacterial RNA processing remains obscure and therefore the protein

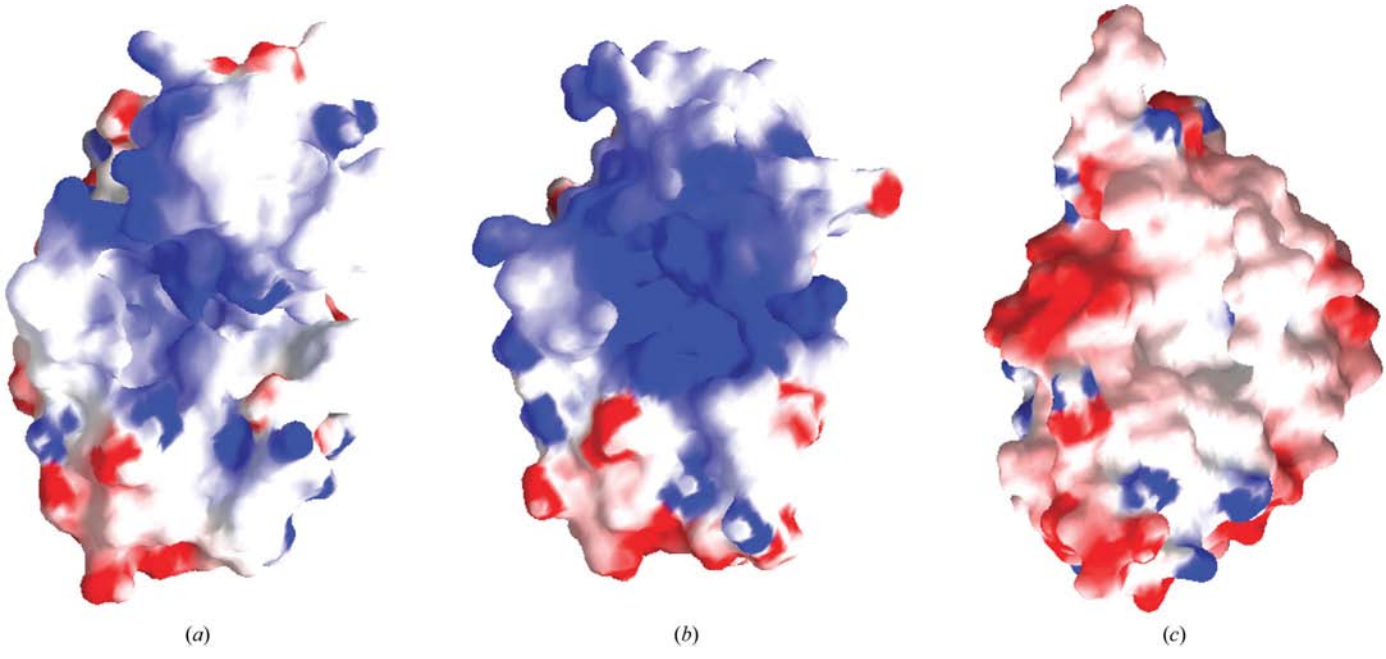

Figure 4
The surface potentials of the three related proteins. (*a*) *P. horikoshii* RNA ligase, (*b*) *T. thermophilus* RNA ligase, (*c*) CPDase. The two histidines modelled as alternate conformations in *T. thermophilus* RNA ligase were reduced to the conformation most closely matching the equivalent residues in the *P. horikoshii* RNA ligase. The orientation of the two RNA ligases is identical and placed to enhance the positive patch of the *T. thermophilus* RNA ligase. The orientation of the CPDase was chosen to match the active-site histidines of the RNA-ligase structures.

can only be assigned as a putative RNA ligase. A comparison of secondary-structural elements and their folding shows that the *P. horikoshii* RNA ligase shares several features in common with CPDase, with which it has no significant homology, that differ from the *T. thermophilus* RNA ligase. In this sense, the *P. horikoshii* RNA ligase can be considered to be intermediate between the two.

We would like to thank Seiki Kuramitsu and Shigeyuki Yokoyama for the plasmid vector used in the protein expression and Yayoi Fujimoto for technical help during purification. We would also like to thank Mitsuaki Sugahara for the TERA contribution. This work was supported by Protein 3000 Japan grant (Project No. PH0099/HTPF10159).

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Arn, E. A. & Abelson, J. N. (1996). *J. Biol. Chem.* **271**, 31145–31153.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Deng, J., Schnaufer, A., Salavati, R., Stuart, K. D. & Hol, W. G. (2004). *J. Mol. Biol.* **343**, 601–613.
Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). *Bioinformatics*, **15**, 305–308.
Greer, C. L., Javor, B. & Abelson, J. (1983). *Cell*, **33**, 899–906.
Greer, C. L., Peebles, C. L., Gegenheimer, P. & Abelson, J. (1983). *Cell*, **32**, 537–546.
Hofmann, A., Grella, M., Botos, I., Filipowicz, W. & Wlodawer, A. (2002). *J. Biol. Chem.* **277**, 1419–1425.
Hofmann, A., Zdanov, A., Genschik, P., Ruvinov, S., Filipowicz, W. & Wlodawer, A. (2000). *EMBO J.* **19**, 6207–6217.
Janin, J. (1997). *Nature Struct. Biol.* **4**, 973–974.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Kato, M., Shirouzu, M., Terada, T., Yamaguchi, H., Murayama, K., Sakai, H., Kuramitsu, S. & Yokoyama, S. (2003). *J. Mol. Biol.* **329**, 903–911.
Kraulis, J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Nasr, F. & Filipowicz, W. (2000). *Nucleic Acids Res.* **28**, 1676–1683.
Nicholls, A., Sharp, K. & Honig, B. (1991). *Proteins*, **11**, 281–296.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sugahara, M. & Miyano, M. (2002). *Tanpakushitsu Kakusan Koso*, **47**, 1026–1032.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Trujillo, M. A., Roux, D., Fueri, J. P., Samuel, D., Cailla, H. L. & Rickenberg, H. V. (1987). *Eur. J. Biochem.* **169**, 167–173.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.