Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

The structure of *Pyrococcus horikoshii* 2′-5′ RNA ligase at 1.94 Å resolution reveals a possible open form with a wider active-site cleft

Yong-Gui Gao,[a] Min Yao,[a,b]* Ayuko Okada[a] and Isao Tanaka[a]

[a]Faculty of Advanced Life Sciences, Graduate School of Life Sciences, Hokkaido University, Sapporo 060-0810, Japan, and [b]RIKEN Harima Institute/Spring-8, Hyogo, Japan

Correspondence e-mail: yao@castor.sci.hokudai.ac.jp

Received 11 July 2006
Accepted 6 November 2006

PDB Reference: 2′-5′ RNA ligase, 1vgj, r1vgjsf.

Bacterial and archaeal 2′-5′ RNA ligases, members of the 2H phosphoesterase superfamily, catalyze the linkage of the 5′ and 3′ exons *via* a 2′-5′-phosphodiester bond during tRNA-precursor splicing. The crystal structure of the 2′-5′ RNA ligase PH0099 from *Pyrococcus horikoshii* OT3 was solved at 1.94 Å resolution (PDB code 1vgj). The molecule has a bilobal $\alpha+\beta$ arrangement with two antiparallel β-sheets constituting a V-shaped active-site cleft, as found in other members of the 2H phosphoesterase superfamily. The present structure was significantly different from that determined previously at 2.4 Å resolution (PDB code 1vdx) in the active-site cleft; the entrance to the cleft is wider and the active site is easily accessible to the substrate (RNA precursor) in our structure. Structural comparison with the 2′-5′ RNA ligase from *Thermus thermophilus* HB8 also revealed differences in the RNA precursor-binding region. The structural differences in the active-site residues (tetrapeptide motifs H-*X*-T/S-*X*) between the members of the 2H phosphoesterase superfamily are discussed.

1. Introduction

The tRNA precursors (pre-tRNA), the primary transcripts of tRNA genes, must undergo processing before becoming functional. This processing involves the removal of introns by splicing, the cleavage of extra 5′ nucleotides, the replacement of UU residues at the 3′ end with a CCA sequence and the modification of numerous bases (Abelson *et al.*, 1998). To remove the introns, the pre-tRNA is cleaved by endonuclease to produce a 5′ exon with a 2′,3′-cyclic monophosphate end and a 3′ exon with a 5′-OH end (Peebles *et al.*, 1983). In eukaryotes, the two half-molecules are joined by a 3′-5′-phosphodiester bond linkage, which is catalyzed by ATP-dependent tRNA ligase through three steps of nucleotidyl transfer. The yeast tRNA ligase consists of an N-terminal adenyltransferase domain, a central domain with kinase activity and a C-terminal domain that is similar to the 2H phosphotransferase superfamily (Sawaya *et al.*, 2003; Wang & Shuman, 2005). However, in bacteria and archaea the 5′ and 3′ exons are linked *via* a 2′-5′-phosphodiester bond, the formation of which is catalyzed by the 2′-5′ RNA ligase, possibly through a one-step reaction involving hydrolysis of the 2′,3′-cyclic phosphodiester bond and the simultaneous formation of a 2′-5′-phosphodiester bond (Kato *et al.*, 2003). Unlike yeast tRNA ligase, the bacterial and archaeal 2′-5′ RNA ligase is a single-domain protein of about 200 residues.

In addition to 2′-5′ RNA ligases (Kato *et al.*, 2003), plant and yeast 1′′, 2′′-cyclic nucleotide phosphodiesterases (CPDases; Hofmann *et al.*, 2000; Nasr & Filipowicz, 2000), fungal RNA ligases (Xu *et al.*, 1990) and human/rat brain 2′,3′-cyclic nucleotide 3′-phosphodiesterases (CNPases; Sakamoto *et al.*, 2005; Kozlov *et al.*, 2003) are classified into the same protein superfamily (the 2H phosphoesterase superfamily). They share common characteristics: two conserved H-*X*-T/S-*X* motifs and protein or catalytic domains consisting of about 200 residues that catalyze the hydrolysis of 2′,3′-cyclic phosphodiester bonds (Mazumder *et al.*, 2002). Members of this superfamily occur across a vast range of organisms from bacteria to

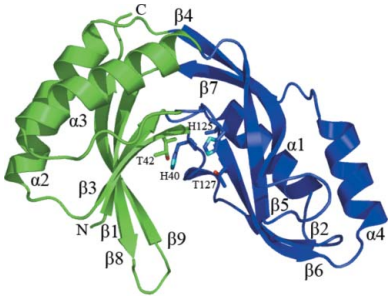

Table 1
Crystal parameters and data-collection statistics.

Values in parentheses are for the highest resolution shell.

Crystal	SeMet
Beamline	BL44B2 (SPring-8)
Wavelength	0.9795
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 41.5, b = 45.7, c = 97.6
Resolution range (Å)	50–1.94 (2.01–1.94)
No. of unique reflections	14330 (1394)
Completeness (%)	99.9 (99.6)
Redundancy	7.0 (6.6)
$I/\sigma(I)$	16.5 (3.8)
$R_{\rm merge}$† (%)	4.7 (38.4)

† $R_{\rm merge} = \sum_h \sum_j |\langle I\rangle_h - I_{hj}| / \sum_h \sum_j I_{hj}$, where $\langle I\rangle_h$ is the mean intensity of symmetry-equivalent reflections.

Table 2
Summary of refinement statistics.

Resolution range (Å)	10–1.94
No. of reflections (total/test)	12780/1414
No. of residues	184
No. of water molecules	41
$R/R_{\rm free}$ (%)	21.6/26.5
R.m.s.d. bond lengths (Å)	0.027
R.m.s.d. angles (°)	1.966
Average B factor (Å^2)	
Protein molecules	33.7
Water molecules	38.1
Ramachandran plot (%)	
Most favoured regions	93.4
Additionally allowed regions	6.6
Generously allowed regions	0
Disallowed regions	0

mammals and their sequence similarity is low. However, the three-dimensional structures of these proteins or corresponding domains are similar. It has been suggested that the four subclasses of this superfamily originated from a common ancestor (Kozlov *et al.*, 2003). In this superfamily, CPDases (Hofmann *et al.*, 2000), fungal RNA ligases (Xu *et al.*, 1990) and CNPases (Sakamoto *et al.*, 2005; Kozlov *et al.*, 2003) have one domain that possesses hydrolysis activity, while in yeast tRNA ligase the C- and N-terminal domains possess hydrolysis and ligation activities, respectively (Sawaya *et al.*, 2003; Wang & Shuman, 2005). Only the subclass of bacteria and archaeal 2′-5′ RNA ligases, which are single-domain proteins, possess both hydrolysis activity towards 2′,3′-cyclic phosphodiester bonds and ligation activity towards two RNA exons with 2′-phosphate and 5′-OH ends. It is possible that the 2′-5′ RNA ligase acts as a bridge between the structural homology of the 2H phosphoesterase superfamily and yeast tRNA ligase.

Pyrococcus horikoshii protein PH0099 is assigned to the 2′-5′ RNA ligases and belongs to the cluster of orthologous groups 1154 (COGs 1154), which has 21 members. The structure of PH0099 was recently determined independently by Rehse and coworkers (PDB code 1vdx, 2.4 Å resolution; Rehse & Tahirov, 2005) and by our research group (PDB code 1vgj). Here, we report the crystal structure of PH0099 at 1.94 Å resolution. Compared with the structure 1vdx, regions (β–turn–β 165–175 and loop 81–88) of the present structure shift toward opposite sides at the entrance to the active-site cleft, forming an open conformation that is beneficial in allowing RNA precursors to approach the active site. The structures of the active sites of the 2H phosphoesterase superfamily were compared.

2. Materials and methods

2.1. Protein production

The recombinant protein PH0099 was expressed in *Escherichia coli* strain B834 (DE3) by adding 1 m*M* IPTG to LB broth (310 K) at an OD_{600} of approximately 0.6. After induction for 5 h, the cells were harvested and then resuspended in buffer *A* (50 m*M* Tris pH 9.0) and disrupted with a French press. The lysate was incubated at 343 K for 30 min and centrifuged (40 000*g* for 30 min at 277 K). The supernatant was applied onto a 5 ml HiTrap QXL column (Amersham Bioscience) which had been pre-equilibrated with buffer *A*. The bound protein was eluted with a linear gradient of NaCl (0–1 *M*, 20 column volumes). The recovered fraction was loaded onto a HiLoad 26/60 Superdex 75pg column (Amersham Bioscience) which had been pre-equilibrated with buffer *B* (50 m*M* Tris pH 9.0, 0.3 *M* NaCl) and the target peak was pooled. After addition of 2 *M* ammonium sulfate, the samples were filtered (0.45 µm, Millipore) and applied onto a 1 ml Resource Phenyl column (Amersham Bioscience) and highly purified PH0099 was eluted. The buffer was exchanged and the sample was concentrated to 10 mg ml^{-1} (in 20 m*M* Tris pH 9.0) using an Amicon column. The target protein was confirmed by MALDI–TOF mass spectrometry. For the production of SeMet-substituted PH0099, cells were cultured in minimum medium containing SeMet. The procedure for the purification of SeMet PH0099 was the same as that used for the native protein.

2.2. Crystallization

Preliminary crystallization trials were performed using crystallization screening kits from Hampton Research. Each drop consisted of 1 µl sample solution and 1 µl reservoir solution and was equilibrated against 100 µl reservoir solution at 293 K. Initial crystals were obtained with Hampton Research PEG/Ion Screen (No. 3; 20% PEG 3350, 0.2 *M* ammonium fluoride pH 6.2). After optimization by changing the buffer and precipitant, crystals that diffracted well were obtained by the hanging-drop method with 2 µl protein solution mixed with an equal volume of mother-liquor solution containing 0.2 *M* ammonium fluoride, 24% PEG 4000 and 0.1 *M* sodium acetate pH 4.6.

2.3. Data collection and processing

X-ray diffraction data were collected at BL44B2 of SPring-8 (Hyogo, Japan). A single crystal was mounted in a loop and flash-cooled in a stream of nitrogen gas after soaking in mother liquor containing 15% glycerol. The X-ray diffraction data were processed using the *HKL*-2000 package (Otwinowski & Minor, 1997). The crystals belong to space group $P2_12_12_1$, with unit-cell parameters a = 41.5, b = 45.7, c = 97.6 Å. The asymmetric unit contains one molecule of PH0099, which corresponds to a solvent content of 43.5% ($V_{\rm M}$ = 2.1 Å^3 Da^{-1}). A summary of data-collection and processing statistics is given in Table 1.

2.4. Structure solution and refinement

The structure of PH0099 was solved by molecular replacement with *MOLREP* (Vagin & Teplyakov, 1997) using the homologous (32.6% sequence identity) protein structure of *T. thermophilus* 2′-5′ RNA ligase (PDB code 1iuh; Kato *et al.*, 2003) as a search model. Data in the resolution range 20–3.0 Å were used in both rotation and translation calculations, which gave an obvious solution with significant contrast. The initial model, which comprised 78.8% of 184 residues, was rebuilt automatically using *ARP*/*wARP* (Lamzin & Wilson, 1993). The fragmented initial model was extended and refined automatically using *LAFIRE* (Yao *et al.*, 2006) with *CNS*

(Brünger *et al.*, 1998) with slight manual intervention (*R* factor and R_{free} = 23.8 and 27.2%, respectively). After a final check and correction, the *R* and free *R* factors for the model consisting of 184 amino-acid residues and 41 water molecules were 21.6 and 26.5%, respectively. The stereochemical quality of the final model was validated with *PROCHECK* (Laskowski *et al.*, 1993). The Ramachandran plot indicated that 93.4% of the residues lie in the most favoured regions, with the remaining 6.6% in additionally allowed regions (Table 2).

3. Results and discussion

The final model of PH0099 is composed of one molecule in the asymmetric unit, containing all residues (184) and 41 water molecules. The overall structure of PH0099, which has dimensions 45 × 40 × 25 Å, contains one 3_{10}-helix, four α-helices and nine β-strands and resembles two lobes positioned symmetrically (Fig. 1*a*). The lobe harbouring the N- and C-terminal regions (the terminal lobe) consists of $\beta1$, $\beta3$, $\alpha2$, $\alpha3$, $\beta8$ and $\beta9$ and forms a four-stranded antiparallel β-sheet. The other lobe (the transit lobe) possesses a 3_{10}-helix, $\alpha1$, $\alpha4$ and a five-stranded antiparallel β-sheet ($\beta4$, $\beta5$, $\beta6$, $\beta2$, $\beta7$). The two antiparallel β-sheets ($\beta1$-$\beta3$-$\beta8$-$\beta9$ and $\beta4$-$\beta5$-$\beta6$-$\beta2$-$\beta7$) constitute a V-shaped active-site cleft (about 25 Å in width and 15 Å in depth), with two antiparallel α-helices wrapping around the outer side of each β-sheet (Fig. 1*a*). Two structures of PH0099 were deposited in the PDB at around the same time: 1vgj (1.94 Å resolution) and 1vdx (Rehse & Tahirov, 2005; 2.4 Å resolution). The two crystals used for structural determination were obtained under different crystallization conditions and belonged to the same space group with slightly different unit-cell parameters [a difference of 2.6 Å (6%) in the *a* axis and 4.4 Å (4.5%) in the *c* axis]. With the exception of two regions, residues 165–175 forming a β–turn–β structure ($\beta8$ and $\beta9$) and a loop containing residues 81–88 linking $\beta4$ to $\beta5$, the two structures superposed very well, with an r.m.s.d. (root-mean-square deviation) of 0.67 Å for the main chain. Compared with 1vdx, the maximum C^{α} shifts of these two regions (residues 165–175 and 81–88) in the present structure are 5.8 and 1.6 Å, respectively. These shifts broaden the width of the entrance of the active-site cleft. In particular, the different orientation of the side chain of Lys170 in our structure almost doubled the opening width of the cleft in comparison with that of 1vdx (Fig. 1*b*).

There is no direct contact between residues situated on opposite sides of the V-shaped active-site cleft in both structures. In the above-mentioned regions (residues 165–175 and 81–88), only one water molecule (W230) was found in the cleft of the structure determined at the higher resolution (1vgj), which makes an interaction with residue Pro172. It is unlikely that this interaction is capable of causing such a conformational change of the residues 165–175 (particularly the side chain of Lys170). The plots of the *B* factors of the main chain for the two structures are shown in Fig. 2. For both structures, the average *B* factors of the main chain of the two regions (residues 81–88 and 165–175) are higher than that overall. Residues 168–171 (the turn between $\beta8$ and $\beta9$) constitute the region with the highest average *B* factors in 1vgj and 1vdx (50.0 and 82.9 Å^2, respectively). The high *B* factors and the structural variations possibly suggest that conformational changes occur in these regions on binding RNA substrate. These changes occurring at the entrance of the active-site cleft can broaden its width (>8 Å) and form an open conformation that is beneficial for allowing RNA precursors to approach the active site. It is possible that the present structure (1vgj) and the previous one (1vdx) are open and closed forms of the *P. horikoshii* 2′-5′ RNA ligase, respectively.

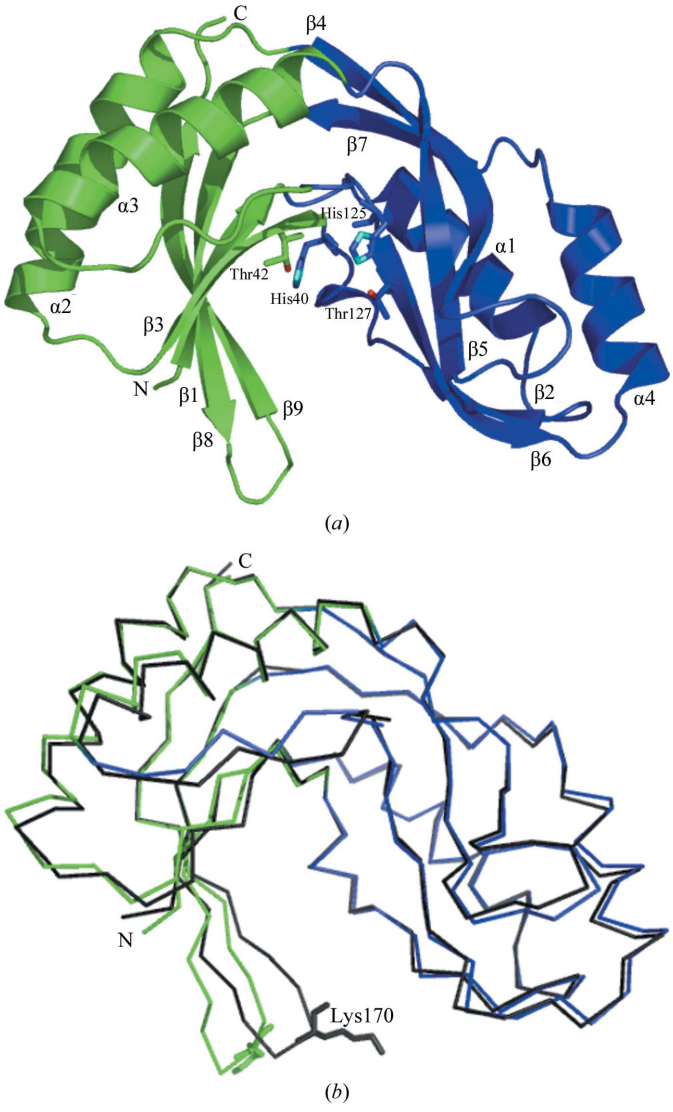

Figure 1
Structure of PH0099 and the active site. The terminal lobe is shown in green and the transit lobe in blue. (*a*) Ribbon representation of PH0099. The active-site residues His40, Thr42, His125 and Thr127 in the tetrapeptide motif H-*X*-T/S-*X* are depicted explicitly in stick representation. (*b*) Comparison of this structure (green and blue) with that of PDB entry 1vdx (black).

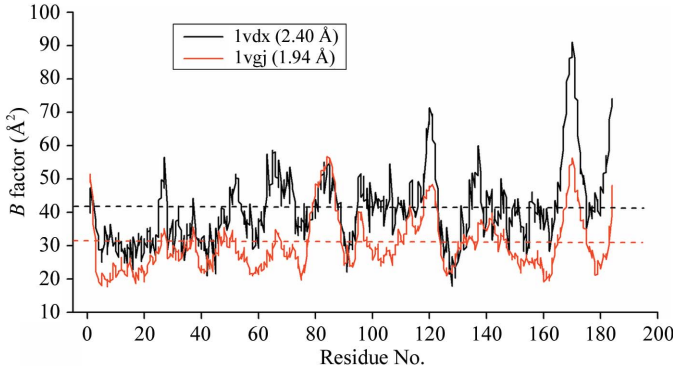

Figure 2
Plot of the *B* factors of the main chains of 1vgj (red) and 1vdx (black). The dashed lines indicate the average *B* factor (1vgj, 31.6 Å^2, red; 1vdx, 41.8 Å^2, black).

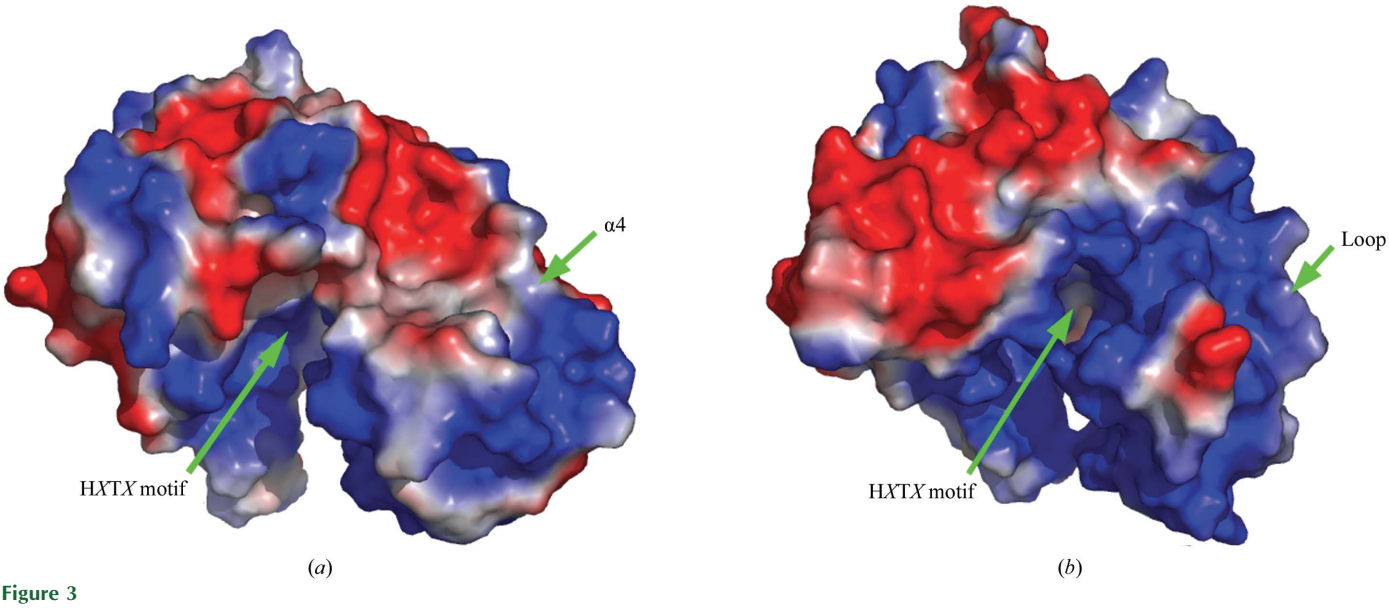

Figure 3
Electrostatic surface potentials of (*a*) *P. horikoshii* and (*b*) *T. thermophilus* 2′-5′ RNA ligase. The electrostatic potential was calculated using *PyMOL* (DeLano, 2002); positively charged regions are in blue and negatively charged regions in red.

The *T. thermophilus* 2′-5′ RNA ligase has a loop at residues 138–148; however, PH0099, CPDase (Hofmann *et al.*, 2000) and CNPase (Sakamoto *et al.*, 2005; Kozlov *et al.*, 2003) have helices α4 corresponding to these positions. PH0099 possess two α-helices distributed on each of the outer sides of the bilobal arrangement, similar to the structures of the 2H phosphoesterase superfamily members CPDase (Hofmann *et al.*, 200) and CNPase (Sakamoto *et al.*, 2005; Kozlov *et al.*, 2003). Compared with *T. thermophilus* 2′-5′ RNA ligase (Kato *et al.*, 2003), the electrostatic surface potential shows that PH0099 has a smaller positively charged patch and that the centre of the active-site cleft is also positively charged, which is essential for recognizing the negatively charged RNA precursors (Fig. 3). The region containing β4, β5 and the C-terminus of α4 in PH0099 is mostly negatively charged (Fig. 3*a*), which implies that it is not involved in binding the RNA precursor. The corresponding region (β4, β5 and loop 138–148) in *T. thermophilus* 2′-5′ RNA ligase is positively charged (Fig. 3*b*) and hence a larger positively charged region is formed compared with that of PH0099. In the case of *T. thermophilus* 2′-5′ RNA ligase, the loop (138–148) may contribute to binding substrate by possibly changing its conformation. A loop is more suitable than a helix for such conformational changes. The different charged regions suggest that the substrates and their recognition may be different for the two ligases. The structure of PH0099 with two α-helices distributed in each lobe, similar to CPDase and CNPase, may act as a bridge between the structural homology of the 2H phosphoesterase superfamily.

The members of the 2H phosphoesterase superfamily share two prominent tetrapeptide motifs, H-*X*-T/S-*X* (*X* is generally a hydrophobic residue), in the antiparallel β-sheets. At the predicted active sites containing the two motifs, the structures of the *T. thermophilus* 2′-5′ RNA ligase (Kato *et al.*, 2003), plant CPDase in two forms (Hofmann *et al.*, 2000, 2002) and the catalytic domain of human/rat CNPase (Sakamoto *et al.*, 2005; Kozlov *et al.*, 2003) were superimposed onto PH0099 and the result showed that the conserved residues in the tetrapeptide motifs (His, Thr or Ser) superimposed well in all proteins, with the exception of one conformation of His130 in *T. thermophilus* 2′-5′ RNA ligase (data not shown). The structural similarity of the tetrapeptide motifs in those proteins is consistent with the common characteristic that the members of the 2H phosphoesterase superfamily catalyze the hydrolysis of 2′,3′-cyclic phosphodiester bonds. In the tetrapeptide motifs H-*X*-T/S-*X*, the distances between the two NE atoms of the two histidines and the two OG1 atoms of the two threonines (Thr to Ser in CPDase) vary in the range 6.7–8.2 Å and 7.0–7.6 Å, respectively. It seems that the distance between the threonines is less varied. This is in agreement with the observation that all threonines are located within β-strands except for Thr132 in *T. thermophilus* 2′-5′ RNA ligase, while all histidines lie in the loop except for the histidine corresponding to His40 (PH0099) in CNPase.

We thank R. Ogawa for help with protein purification. We also thank the staff of beamline BL44B2, SPring-8, Japan for their kind help with data collection. This work was funded by the National Project on Protein Structure and Functional Analyses from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Abelson, J., Trotta, C. R. & Li, H. (1998). *J. Biol. Chem.* **273**, 12685–12688.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.
Hofmann, A., Grella, M., Botos, I. & Filipowicz, W. (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.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Kato, M., Shirouzu, M., Terada, T., Yamaguchi, H., Murayama, K., Sakai, H., Kuramitsu, S. & Yokoyama, S. (2003). *J. Mol. Biol.* **329**, 903–911.
Kozlov, G., Lee, J., Elias, D., Gravel, M., Gutierrez, P., Ekiel, I., Braun, P. E. & Gehring, K. (2003). *J. Biol. Chem.* **278**, 46021–46028.
Mazumder, R., Iyer, L. M., Vasudevan, S. & Aravind, L. (2002). *Nucleic Acids Res.* **30**, 5229–5243.

Nasr, F. & Filipowicz, W. (2000). *Nucleic Acids Res.* **28**, 1676–1683.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Peebles, C. L., Gegenheimer, P. & Abelson, J. (1983). *Cell*, **32**, 525–536.
Rehse, P. H. & Tahirov, T. H. (2005). *Acta Cryst.* D**61**, 1207–1212.
Sakamoto, Y., Tanaka, N., Ichimiya, T., Kurihara, T. & Nakamura, K. T. (2005). *J. Mol. Biol.* **346**, 789–800.
Sawaya, R., Schwer, B. & Shuman, S. (2003). *J. Biol. Chem.* **278**, 43928–43938.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Wang, L. K. & Shuman, S. (2005). *RNA*, **11**, 966–975.
Xu, Q., Teplow, D., Lee, T. D. & Abelson, J. (1990). *Biochemistry*, **29**, 6132–6138.
Yao, M., Zhou, Y. & Tanaka, I. (2006). *Acta Cryst.* D**62**, 189–196.