The structure of Pyrococcus horikoshii 2'-5' RNA ligase at 1.94Å resolution reveals a possible open form with a wider active-site cleft.
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, Min Yao, Ayuko Okada and Isao Tanaka
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

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 α+β 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. Members of this superfamily undergo 2′,3′-cyclic 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., 1999) 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...
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',5'-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 solved by molecular replacement (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',5'-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 solved by molecular replacement (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',5'-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.

2. Materials and methods

2.1. Protein production

The recombinant protein PH0099 was expressed in Escherichia coli strain B834 (DE3) by adding 1 mM IPTG to LB broth (310 K) at an OD600 of approximately 0.6. After induction for 5 h, the cells were harvested and then resuspended in buffer A (50 mM Tris pH 9.0) and disrupted with a French press. The lysate was incubated at 343 K for 30 min and centrifuged (40 000g 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 mM 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 mM 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. Crystalization

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 (Otwinski & Minor, 1997). The crystals belong to space group P2_1_2_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_M = 2.1 Å³ 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 1hu; 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.
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 \( \times \) 40 \( \times \) 25 Å, contains one \( \beta \)-helix, four \( \alpha \)-helices and nine \( \beta \)-strands and resembles two lobes positioned symmetrically (Fig. 1a). The lobe harbouring the N- and C-terminal regions (the terminal lobe) consists of \( \beta_1, \beta_3, \alpha_2, \alpha_3, \beta_8 \) and \( \beta_9 \) and forms a four-stranded antiparallel \( \beta \)-sheet. The other lobe (the transit lobe) possesses a \( \beta_4 \)-helix, \( \alpha_1, \alpha_4 \) and a five-stranded antiparallel \( \beta \)-sheet (\( \beta_4, \beta_5, \beta_6, \beta_2, \beta_7 \)). The two antiparallel \( \beta \)-sheets (\( \beta_1-\beta_3-\beta_8-\beta_9 \) and \( \beta_4-\beta_5-\beta_6-\beta_2-\beta_7 \)) constitute a V-shaped active-site cleft (about 25 Å in width and 15 Å in depth), with two antiparallel \( \alpha \)-helices wrapping around the outer side of each \( \beta \)-sheet (Fig. 1a). 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 \( \beta \)-turn–\( \beta \) structure (\( \beta_8 \) and \( \beta_9 \)) and a loop containing residues 81–88 linking \( \beta_4 \) to \( \beta_5 \), 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 \( C^\alpha \) 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. 1b).

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 \( \beta_8 \) and \( \beta_9 \)) 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.
The \textit{T. thermophilus} 2-H 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 a4 corresponding to these positions. PH0099 possesses two \(\alpha\)-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., 2000) and CNPase (Sakamoto et al., 2005; Kozlov et al., 2003). Compared with \textit{T. thermophilus} 2-H 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 \(\beta_4\), \(\beta_5\) and the C-terminus of a4 in PH0099 is mostly negatively charged (Fig. 3a), which implies that it is not involved in binding the RNA precursor. The corresponding region (\(\beta_4\), \(\beta_5\) and loop 138–148) in \textit{T. thermophilus} 2-H RNA ligase is positively charged (Fig. 3b) and hence a larger positively charged region is formed compared with that of PH0099. In the case of \textit{T. thermophilus} 2-H 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 \(\alpha\)-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 \(\beta\)-sheets. At the predicted active sites containing the two motifs, the structures of the \textit{T. thermophilus} 2-H 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 His30 in \textit{T. thermophilus} 2-H 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 \(\beta\)-strands except for Thr132 in \textit{T. thermophilus} 2-H 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
structural genomics communications
