Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the AlaX-M *trans*-editing enzyme from *Pyrococcus horikoshii*

Ryuya Fukunaga[a,b] and Shigeyuki Yokoyama[a,b]*

[a]Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, and [b]RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

Correspondence e-mail: yokoyama@biochem.s.u-tokyo.ac.jp

Received 27 November 2006
Accepted 27 December 2006

PDB Reference: AlaX-M, 2e1b, r2e1bsf.

The editing domain of alanyl-tRNA synthetase (AlaRS) contributes to high-fidelity aminoacylation by hydrolyzing (editing) the incorrect products Ser-tRNA^{Ala} and Gly-tRNA^{Ala} (*cis*-editing). The AlaX protein shares sequence homology to the editing domain of AlaRS. There are three types of AlaX proteins, with different numbers of amino-acid residues (AlaX-S, AlaX-M and AlaX-L). In this report, AlaX-M from *Pyrococcus horikoshii* is shown to deacylate Ser-tRNA^{Ala} and Gly-tRNA^{Ala} (*trans*-editing). The crystal structure of *P. horikoshii* AlaX-M has been determined at 2.7 Å resolution. AlaX-M consists of an N-terminal domain (N-domain) and a C-terminal domain (C-domain). A zinc ion is coordinated by the conserved zinc-binding cluster in the C-domain, which is expected to be the enzymatic active site. The glycine-rich motif, consisting of successive conserved glycine residues in the N-domain, forms a loop (the 'glycine-rich loop'). The glycine-rich loop is located near the active site and may be involved in substrate recognition and/or catalysis.

1. Introduction

The genetic code is established by aminoacyl-tRNA synthetases (aaRSs), which attach specific amino acids to the 3′-ends of tRNAs bearing the cognate anticodons for those amino acids (Martinis *et al.*, 1999; Ibba & Soll, 2000). Several aaRSs, including alanyl-tRNA synthetase (AlaRS), have evolved proofreading/editing mechanisms that hydrolyze incorrect products, thereby preventing errors in protein synthesis. The aminoacylation active site of AlaRS misrecognizes the noncognate amino acids serine and glycine as well as the cognate alanine (Swairjo & Schimmel, 2005) and produces Ser-tRNA^{Ala} and Gly-tRNA^{Ala}. The editing domain of AlaRS hydrolyzes (edits) the incorrect products Ser-tRNA^{Ala} and Gly-tRNA^{Ala} (*cis*-editing; Beebe *et al.*, 2003). A defect in the editing activity of AlaRS causes the accumulation of misfolded proteins and cell death in the mouse nervous system (Lee *et al.*, 2006).

In addition to the internal editing domain embedded in AlaRS, there are free-standing proteins, called AlaX proteins, that share sequence homology with the AlaRS editing domain. The AlaX proteins can be divided into three subfamilies according to their amino-acid sequences (referred to as AlaX-L, AlaX-M and AlaX-S; Fig. 1). AlaX-L has the longest amino-acid sequence and possesses domains that are homologous to the entire editing domain and the oligomerization

domain of AlaRS. AlaX-M corresponds to the entire editing domain of AlaRS. AlaX-S is the shortest and shares homology with the C-terminal part of the AlaRS editing domain. The editing domain of threonyl-tRNA synthetase (ThrRS) is appended to the N-terminus of the aminoacylation catalytic domain and hydrolyzes Ser-tRNA^{Thr} (Sankaranarayanan *et al.*, 1999; Dock-Bregeon *et al.*, 2000). The C-terminal parts of the ThrRS editing domain and the AlaRS editing domain share sequence homology. The AlaRS editing domain, AlaX-L, AlaX-M, AlaX-S and the ThrRS editing domain have a completely conserved zinc-binding motif, H*XXX*H and C*XXX*H, which has been shown to be crucial for the editing activities of AlaRS and ThrRS (Dock-Bregeon *et al.*, 2000; Beebe *et al.*, 2003). A conserved 'glycine-rich motif' containing successive glycine residues exists in the homologous N-terminal parts of the AlaRS editing domain, AlaX-L and AlaX-M.

The AlaX-L protein from *Saccharomyces cerevisiae* lacks editing activity (Ahel *et al.*, 2003) and its function remains unknown. The AlaX-M proteins from *Methanosarcina barkeri* and *Sulfolobus solfataricus* have been shown to hydrolyze Ser-tRNA^{Ala} and Gly-tRNA^{Ala} (*trans*-editing), but not Ala-tRNA^{Ala} or Ser-tRNA^{Ser} (Ahel *et al.*, 2003). In contrast, the AlaX-S protein from *Pyrococcus horikoshii* hydrolyzes Ser-tRNA^{Ala}, but not Gly-tRNA^{Ala} (Sokabe *et al.*, 2005). The AlaX-L and AlaX-M proteins are distributed widely among the three kingdoms of life, while AlaX-S is only encoded in some organisms, including *Pyrococcus*. Some organisms have no AlaX proteins. *Pyrococcus* is the only organism that has all three AlaX proteins.

The crystal structures of *P. horikoshii* AlaX-S complexed with serine, of the *Escherichia coli* ThrRS editing domain complexed with several substrate analogues and of the entire *E. coli* ThrRS complexed with tRNA have been determined (Dock-Bregeon *et al.*, 2000, 2004; Sokabe *et al.*, 2005). AlaX-S and the ThrRS editing domain exhibit similar structures and their editing active sites are located in the zinc-binding region. However, the structure and function of the N-terminal parts of the AlaRS editing domain and of AlaX-L and AlaX-M, which contain the conserved glycine-rich motif, are unknown. The mechanism by which the AlaRS editing domain and AlaX-M recognize both Ser-tRNA^{Ala} and Gly-tRNA^{Ala} also remains to be clarified and may be related to the N-terminal part of the enzymes. To address these issues, we determined the crystal structure of AlaX-M.

Here, we report the crystal structure of *P. horikoshii* AlaX-M (PH0108) at 2.7 Å resolution. The structure revealed that the N-terminal part folds into a discrete domain composed of a β-barrel (the N-domain). The glycine-rich motif forms a loop (the glycine-rich loop) which is located near the zinc-binding editing active site within the C-terminal domain (the C-domain). The glycine-rich loop may be involved in substrate recognition and/or catalysis.

2. Materials and methods

2.1. Expression and purification of AlaX-M

The gene fragment encoding *P. horikoshii* AlaX-M (PH0108) was cloned into pET26b (Novagen). The plasmid was transformed into *E. coli* strain BL21(DE3) Codon Plus (Stratagene). For protein overexpression, the cells were grown to an OD_{600} of 0.6–0.8 and expression was induced with 0.5 m*M* isopropyl β-D-thiogalactopyranoside (IPTG) for an additional 4 h at 310 K. The cells were harvested and sonicated in sonication buffer [50 m*M* Tris–HCl buffer pH 8.0 containing 500 m*M* NaCl, 5 m*M* $MgCl_2$, 10 m*M* dithiothreitol (DTT) and 1 m*M* phenylmethylsulfonylfluoride]. The insoluble cell debris was removed by centrifugation at 15 000*g* for 30 min at 277 K. The supernatant was heat-treated at 353 K for 30 min in order to denature the *E. coli* proteins and was then centrifuged at 15 000*g* for 30 min at 277 K. The supernatant was dialyzed against 20 m*M* Tris–HCl buffer pH 8.0 containing 2 m*M* $MgCl_2$ and 1 m*M* DTT and was then loaded onto a Q-Sepharose FF column (GE Healthcare). The enzyme was eluted with a linear gradient of 0–0.5 *M* NaCl. After addition of 3.5 *M* $(NH_4)_2SO_4$ to the fractions to give a final concentration of 1.5 *M*, the mixture was applied onto a Resource PHE column (GE Healthcare) equilibrated with 50 m*M* Tris–HCl buffer pH 8.0 containing 1.5 *M* $(NH_4)_2SO_4$, 5 m*M* $MgCl_2$ and 1 m*M* DTT. AlaX-M was eluted with a linear gradient of 1.5–0 *M* $(NH_4)_2SO_4$. The fractions were dialyzed against 10 m*M* Tris–HCl buffer pH 8.0 containing 50 m*M* NaCl, 5 m*M* $MgCl_2$, 0.2 m*M* zinc acetate and 5 m*M* β-mercaptoethanol and were concentrated to 7–9 mg ml^{-1}. The final purity of the protein, as monitored by SDS–PAGE and UV spectroscopy, was more than 98%. About 5 mg of purified protein was

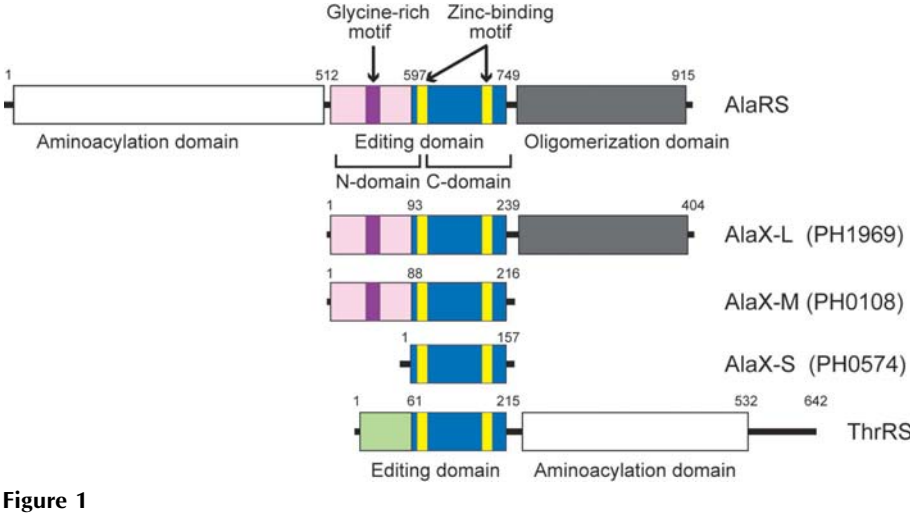

Figure 1
The domain organizations of AlaRS, AlaX-L, AlaX-M, AlaX-S and ThrRS. The amino-acid residue numbers are shown for AlaRS, AlaX-L, AlaX-M and AlaX-S from *P. horikoshii*, and for ThrRS from *E. coli*.

Table 1
Summary of data-collection, phasing and refinement statistics.

Values in parentheses are for the last shell.

		SeMet			
	Native	Peak	Inflection	Remote1	Remote2
Data collection					
Space group	$P6_2$	$P6_2$			
Unit-cell parameters (Å, °)	$a = b = 80.22$, $c = 72.27$, $\alpha = 90$, $\beta = 90$, $\gamma = 120$	$a = b = 81.12$, $c = 71.98$, $\alpha = 90$, $\beta = 90$, $\gamma = 120$			
Wavelength	1.000	0.9793	0.9796	0.9643	0.9953
Resolution (Å)	50–2.7	50–2.9	50–3.0	50–3.0	50–3.1
R_{sym} or R_{merge}†	6.0 (34.0)	7.2 (31.9)	6.1 (30.5)	6.9 (33.0)	6.8 (30.9)
$I/\sigma(I)$	34.9 (2.0)	41.5 (2.9)	36.7 (2.9)	30.6 (2.4)	31.0 (2.5)
Completeness (%)	97.5 (90.2)	98.7 (89.7)	98.6 (88.9)	97.8 (84.0)	97.9 (84.9)
Redundancy	7.8	17.0	8.8	8.6	8.4
Refinement					
Resolution (Å)	50–2.7				
No. reflections	7167				
R_{work}/R_{free}‡	22.7/31.3				
No. of atoms					
Protein	1786				
Zn ion	1				
B factors (Å^2)					
Protein	82.7				
Zn ion	131.5				
R.m.s. deviations					
Bond lengths (Å)	0.0058				
Bond angles (°)	1.20				

† $R_{sym} = \sum|I_{avg} - I_i|/\sum I_i$. ‡ $R_{work} = \sum|F_o - F_c|/\sum F_o$ for reflections belonging to the work set. $R_{free} = \sum|F_o - F_c|/\sum F_o$ for reflections belonging to the test set (10% of the total reflections).

obtained from 1 l cell culture. The protein sample was stored at 193 K until use. For the purification of the selenomethionine-labelled AlaX-M, the transformed methionine-auxotrophic *E. coli* strain B834 Codon Plus (DE3) was grown in minimal medium in which the methionine was substituted with selenomethionine. The labelled protein was purified in the same manner as the native protein.

2.2. Expression and purification of *Archaeoglobus fulgidus* AlaRS

The gene fragment encoding *A. fulgidus* AlaRS was cloned into pET28c, which expresses the *A. fulgidus* AlaRS with a His tag at the N-terminus. The C703A mutation was introduced using the PCR method. The plasmid was transformed into *E. coli* strain BL21(DE3) Codon Plus. For protein over-expression, the cells were grown to an OD_{600} of 0.6–0.8 and expression was induced with 0.5 m*M* IPTG for an additional 4 h at 310 K. The cells were harvested and sonicated in the sonication buffer. Insoluble cell debris was removed by centrifugation at 15 000*g* for 30 min at 277 K. The supernatant was heat-treated at 353 K for 30 min to denature the *E. coli* proteins and was then centrifuged at 15 000*g* for 30 min at 277 K. The wild-type and C703A mutant AlaRS enzymes with a His tag were further purified using a HisTrap column (GE Healthcare). The fractions were dialyzed against 30 m*M* Tris–HCl buffer pH 8.0 containing 150 m*M* KCl, 10 m*M* $MgCl_2$, 0.2 m*M* zinc acetate and 5 m*M* β-mercaptoethanol.

2.3. Gel filtration

A 50 µl aliquot of the purified AlaX-M protein (1.5 mg ml^{-1}) was loaded onto a Superdex200 10/30 gel-filtration column (GE Healthcare) equilibrated with 30 m*M* Tris–HCl buffer pH 8.0 containing 300 m*M* NaCl, 5 m*M* $MgCl_2$, 0.2 m*M* zinc acetate and 5 m*M* β-mercaptoethanol. The flow rate was 0.5 ml min^{-1}. Horse heart cytochrome *c* (12 kDa), bovine erythrocyte carbonic anhydrase (29 kDa) and bovine serum albumin (66 kDa) were used as molecular-weight markers.

2.4. Editing assay

P. horikoshii tRNA^{Ala} was transcribed *in vitro* and purified essentially as described by Fukunaga *et al.* (2005). (^{14}C)-Ser-tRNA^{Ala} and (^{14}C)-Gly-tRNA^{Ala} were prepared by incubating reaction mixtures containing 30 m*M* Tris–HCl buffer pH 7.5, 10 m*M* $MgCl_2$, 3 m*M* ATP, 5 µ*M* *P. horikoshii* tRNA^{Ala}, 80 µ*M* (^{14}C)-labelled amino acid (300 cpm pmol^{-1}; serine or glycine) and 5 µ*M* editing-deficient C703A mutant of *A. fulgidus* AlaRS at 338 K for 10 min. The (^{14}C)-aminoacyl-tRNAs thus produced were purified as described by Giegé *et al.* (1974) and Schmidt & Schimmel (1994). The Ser-tRNA^{Ala} and Gly-tRNA^{Ala} deacylation (editing) reactions were carried out at 328 K in 30 m*M* Tris–HCl buffer pH 7.5 containing 150 m*M* KCl, 10 m*M* $MgCl_2$, 2 µ*M* (^{14}C)-aminoacyl-tRNA and 0, 0.1, 0.5 or 3 µ*M* AlaX-M enzyme. Aliquots were removed at specific time points and were quenched on filter papers (Whatman, 3 mm) equilibrated with 10% trichloroacetic acid (TCA). The filters were washed three times with 5% ice-cold TCA and once with 100% ethanol. The radioactivities of the precipitates were quantitated by scintillation counting.

2.5. Crystallization

The AlaX-M crystals were grown by the hanging-drop vapour-diffusion method at 293 K. The reservoir solution contained 100 m*M* bicine–NaOH buffer pH 9.0, 22% PEG 6000 and 0.1 *M* ammonium acetate. Drops were produced by mixing 1 µl protein solution with 1 µl reservoir solution and were suspended over a 500 µl reservoir. Rod-shaped colourless crystals with dimensions of up to 50 × 50 × 500 µm were obtained within a week. The crystals were harvested in the reservoir solution and were transferred into a cryoprotective solution containing 20%(*v*/*v*) glycerol. The crystals were then mounted in a nylon loop and flash-cooled with liquid nitrogen.

2.6. Data collection

The diffraction data sets were collected on a Quantum-315 CCD detector (Area Detector Systems Corp.) at beamline BL41XU, SPring-8, Harima, Japan at 100 K. The data sets

were integrated and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997).

2.7. Structure determination and refinement

The initial phases for the SeMet AlaX-M structure were solved by the MAD method. The peak-wavelength anomalous scattering data were input into *SnB* (Weeks & Miller, 1999) in order to identify the locations of the Se atoms. Of the five possible Se sites, three were found. Using the identified Se sites, the phases were calculated for the four-wavelength MAD data sets using *MLPHARE* (Collaborative Computational Project, Number 4, 1994). Density modification was performed with *DM* (Collaborative Computational Project, Number 4, 1994). The model structure was built manually using *O* (Jones *et al.*, 1991). Structure refinement was carried out using *CNS* (Brünger *et al.*, 1998). A random 10% of the starting data were set aside for cross-validation. The refinement included several rounds of simulated-annealing refinement and individual *B*-factor refinement. After the model had been built, it was refined against the native data to 2.7 Å resolution. Several rounds of simulated-annealing refinement, energy-minimization refinement and individual *B*-factor refinement were performed. The refinement converged to a final *R* factor of 22.7%, with an $R_{\rm free}$ of 31.3% (Table 1). In the Ramachandran plot, 72.7%, 24.6% and 2.7% of the residues (136, 46 and five amino-acid residues, respectively) are in the most favoured regions, additional allowed regions and generously allowed regions, respectively. No residues are in the disallowed region.

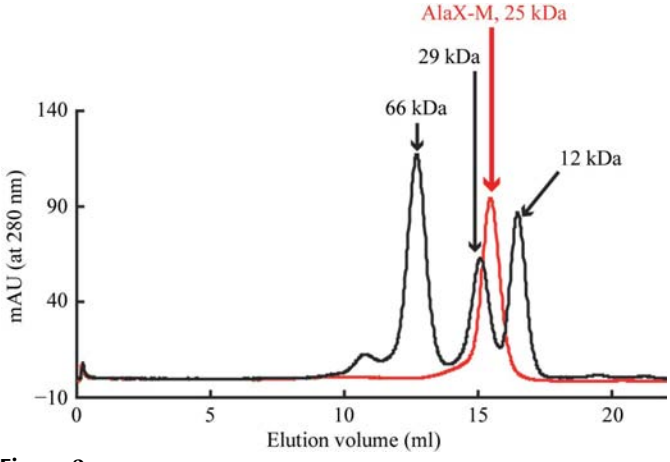

Figure 2
Gel-filtration analysis. The observed and calculated molecular weights of *P. horikoshii* AlaX-M are 25 and 25.3 kDa, respectively, clearly revealing that it exists as a monomer.

3. Results and discussion

3.1. *P. horikoshii* AlaX-M behaves as a monomer

The calculated molecular weight of *P. horikoshii* AlaX-M is 25.3 kDa. The gel-filtration analysis revealed an observed molecular weight of about 25 kDa (Fig. 2). This result clearly shows that *P. horikoshii* AlaX-M behaves as a monomer in solution. In contrast, AlaX-S forms both a monomer and a homodimer and the monomeric form exhibits higher editing activity (Sokabe *et al.*, 2005).

3.2. *P. horikoshii* AlaX-M deacylates Ser-tRNA^{Ala} and Gly-tRNA^{Ala}

The alanine mutation of the conserved cysteine residue (Cys703) in the editing domain of *A. fulgidus* AlaRS abolished the editing activity (data not shown). The C703A mutant of *A. fulgidus* AlaRS efficiently ligated serine and glycine to *P. horikoshii* tRNA^{Ala} and produced Ser-tRNA^{Ala} and Gly-tRNA^{Ala}.

P. horikoshii AlaX-M exhibited deacylation activity against both Ser-tRNA^{Ala} and Gly-tRNA^{Ala} (Fig. 3), consistent with

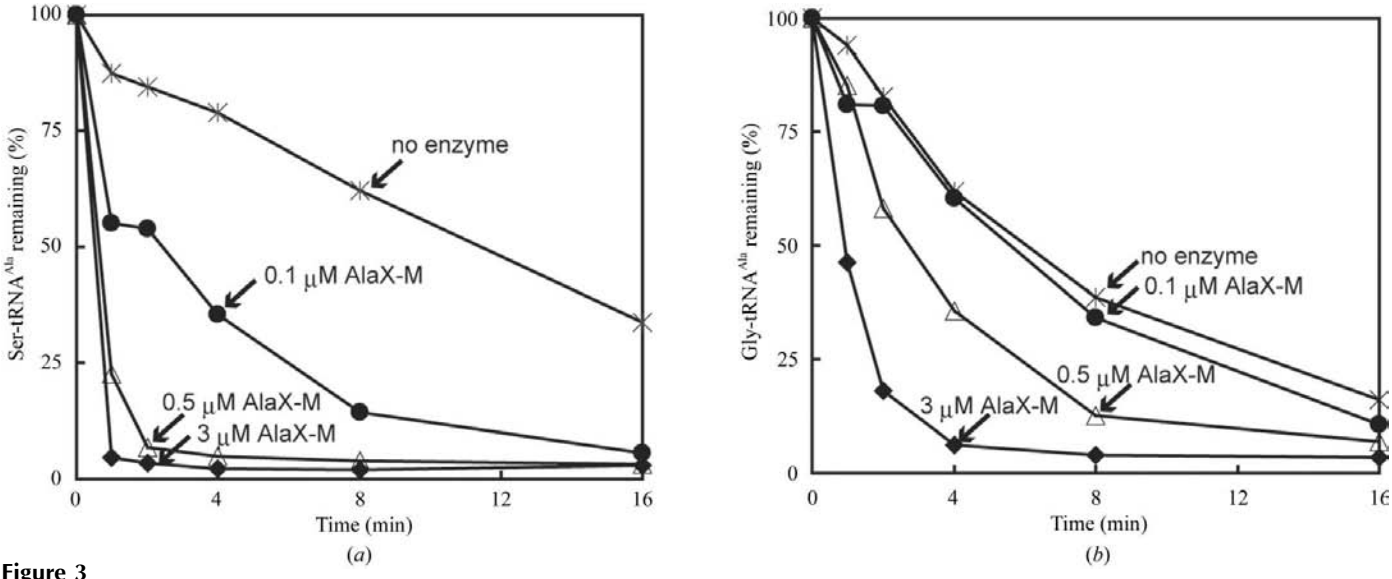

Figure 3
Editing assay. (*a*) Ser-tRNA^{Ala} deacylation by *P. horikoshii* AlaX-M. (*b*) Gly-tRNA^{Ala} deacylation by *P. horikoshii* AlaX-M.

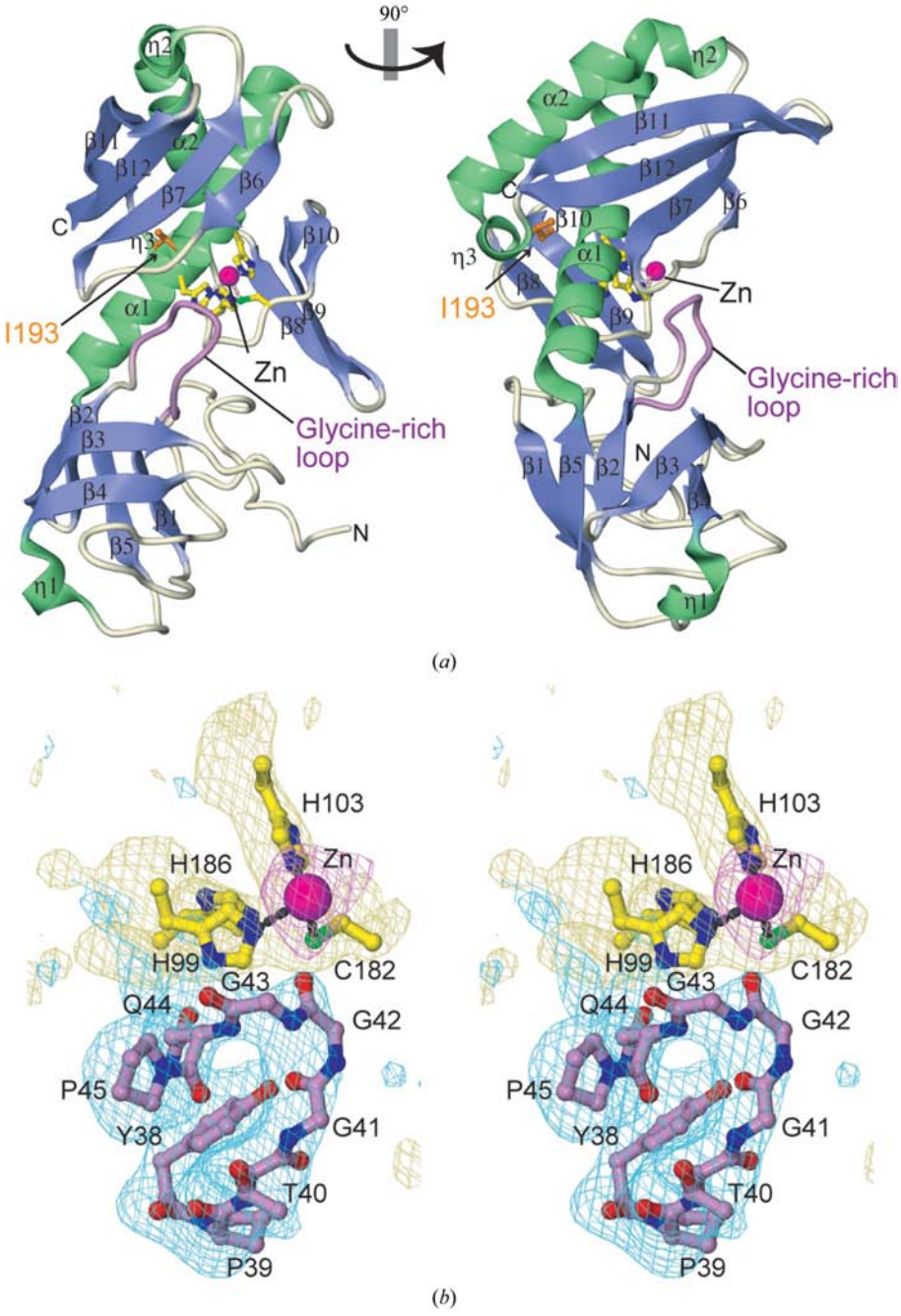

Figure 4
Overall structure of *P. horikoshii* AlaX-M. (*a*) The structure of *P. horikoshii* AlaX-M is shown as a ribbon model. The glycine-rich loop is coloured violet. The zinc ion is depicted as a magenta sphere. The zinc-binding residues (His99, His103, Cys182 and His186) are shown as yellow ball-and-stick models. Ile193 is shown as an orange ball-and-stick model. (*b*) The $|F_o - F_c|$ simulated-annealing OMIT electron-density maps (2.5σ) for the zinc ion, the zinc-binding residues and the glycine-rich loop are shown in magenta, yellow and cyan, respectively (stereoview).

3.3. Overall structure of *P. horikoshii* AlaX-M

The structure of *P. horikoshii* AlaX-M was solved by the SeMet MAD method and was refined to 2.7 Å resolution (Fig. 4). The N-terminal part of AlaX-M (residues 1–87; the N-domain) is a β-barrel consisting of five anti-parallel β-strands arranged in the order $\beta1$–$\beta5$–$\beta2$–$\beta3$–$\beta4$. The C-terminal part of the enzyme (residues 88–216; the C-domain) is composed of a long central α-helix ($\alpha1$) surrounded by two anti-parallel β-sheets and several helices. A zinc ion is bound in the zinc-binding motif on the central helix, involving the conserved residues His99, His103, Cys182 and His186. The glycine-rich motif in the N-domain forms a loop (the glycine-rich loop), which is located near the zinc-binding site in the C-domain.

A structural similarity search using the *DALI* server (Holm & Sander, 1993) revealed a maximum *Z* score of 5.4, corresponding to domain II of the translation-initiation factor IF2/2IF5B (Roll-Mecak *et al.*, 2000). A similarity search using the *VAST* server (Gibrat *et al.*, 1996) generated a maximum VAST score of 9.4, corresponding to domain II of the mitochondrial translation elongation factor Tu (Andersen *et al.*, 2000). However, the functional residues in these enzymes are not conserved in AlaX-M. They do not contain the glycine-rich loop.

3.4. Architecture of the zinc-binding site and the glycine-rich loop

The zinc ion is coordinated by three residues, His99, His103 and Cys182, while His186 does not participate in zinc recognition (Fig. 5*a*). The presence of zinc ions in the crystals was confirmed by X-ray absorption spectroscopy (data not shown). The conserved glycine-rich motif (Tyr39-Pro39-Thr40-Gly41-Gly42-Gly43-Gln44-Pro45) forms the glycine-rich loop. The successive conserved glycine residues (Gly41-Gly42-Gly43) form the tip of the loop and are closest to the zinc-binding site. Some residues within the glycine-rich loop hydrogen bond to residues in the C-domain. The side chain of Tyr38 hydrogen bonds to the main-chain CO of Gly183. The main-chain CO of Gly43 hydrogen bonds to the main-chain NH of Met125. The side chain of Gln44 hydrogen bonds to the side chains of His99 and Thr185 and the main-chain NH of

the results obtained with the AlaX-M enzymes from *M. barkeri* and *S. solfataricus* (Ahel *et al.*, 2003). Deacylation activity against Ser-tRNA^{Ala} was observed at 0.1 μM enzyme concentration. On the other hand, for the editing of Gly-tRNA^{Ala}, a higher enzyme concentration (0.5 μM) was required. Therefore, *P. horikoshii* AlaX-M has *trans*-editing activity against both Ser-tRNA^{Ala} and Gly-tRNA^{Ala}, with higher efficiency against the former.

Thr185. Gln44, His99 and Thr185 are completely conserved, and Tyr38 is mostly conserved among the 'N-domain-bearing proteins' (the AlaRS editing domain, AlaX-L and AlaX-M; Fig. 6), suggesting the importance of the aforementioned hydrogen-bonding interactions between the glycine-rich loop and the C-domain. Notably, His99 is one of the zinc-coordinating residues and therefore these interactions may be important for the zinc-binding and/or editing reaction.

3.5. Interactions between the N- and C-domains

Although the N- and C-domains form discrete domains, they seem to be fixed and do not move flexibly relative to each other, since they interact by hydrogen bonding. The hydrogen-bonding interactions between the two domains are as follows (the first and second residues are in the N- and C-domains, respectively, and S and M represent the side chain and the main chain, respectively): Asp12(S)···Arg92(S), Tyr14(S)···Lys188(M), Tyr14(S)···Glu192(S), Ile36(M)···Arg92(S), Tyr38(S)···Gly183(M), Gly43(M)···Met125(M), Gln44(S)···His99(S), Gln44(S)···Thr185(S), Gln44S(S)···Thr185(M) and Asp47(S)···Arg92(S) (Fig. 5*b*). Some of these residues are conserved among the N-domain-bearing proteins (Fig. 6), indicating that the N- and C-domain structures and the observed interactions between the two domains are shared between the three proteins.

3.6. Hydrophobic core

In the C-domain, a hydrophobic core is formed under the central helix α1, which bears the zinc-binding site (Fig. 5*c*). The hydrophobic residues in the hydrophobic core of the *P. horikoshii* AlaX-M structure are either conserved or replaced by other hydrophobic residues in the AlaRS editing domain, AlaX-S, AlaX-M, AlaX-L and the ThrRS editing domain (Fig. 6). Therefore, the hydrophobic core is important for the architecture of the C-domain and the catalytic site. The mutation of one of the hydrophobic residues, Ala734, constituting the hydrophobic core in mouse AlaRS (corresponding to Ile193 in AlaX-M) to a hydrophilic Glu residue caused protein misfolding and neurodegeneration *in vivo* owing to reduced editing activity towards Ser-tRNA^{Ala} (Lee *et al.*, 2006).

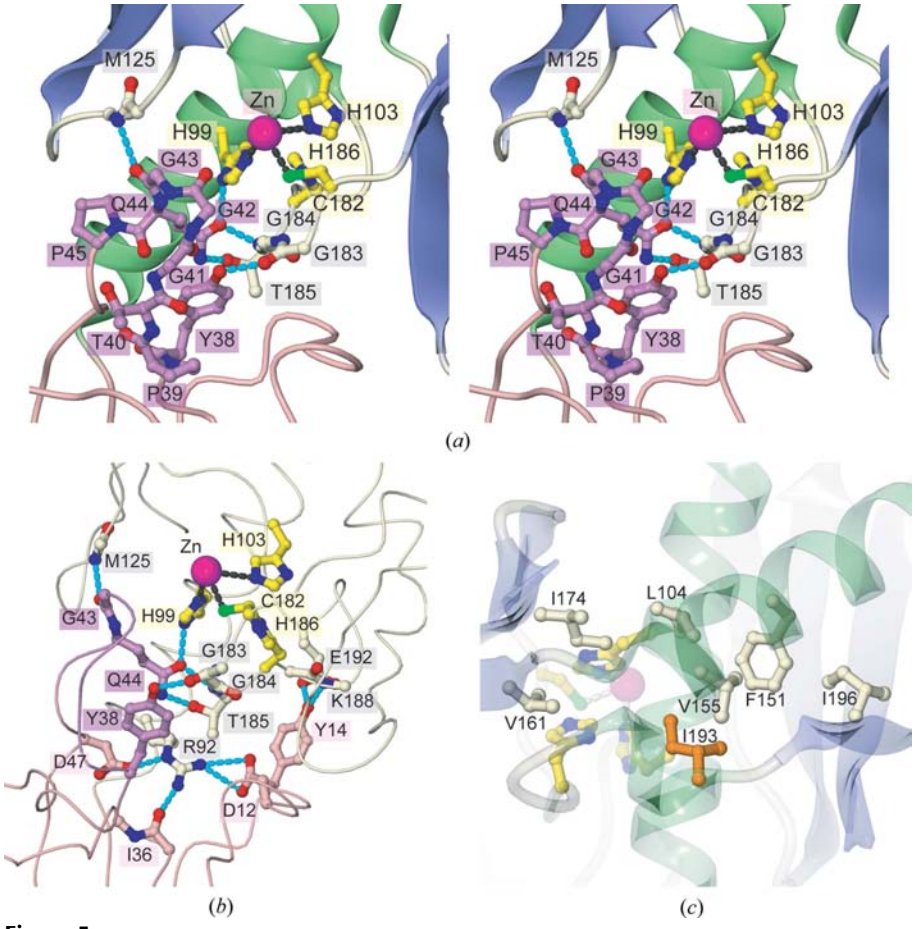

Figure 5
Zinc-binding site and glycine-rich loop. (*a*) Architecture of the zinc-binding site and the glycine-rich loop (stereoview). The residues composing the glycine-rich loop are shown as violet ball-and-stick models. The N-domain is depicted as a pink tube. The C-domain is shown as a ribbon model. The zinc-binding residues and the zinc ion are depicted as yellow ball-and-stick models and a magenta sphere, respectively. (*b*) Interactions between the N- and C-domains, which are shown as pink and white tubes, respectively. The residues participating in the interactions between the N- and C-domains are shown as ball-and-stick models. Residues in the N-domain, the glycine-rich loop and the C-domain are coloured pink, violet and white, respectively. (*c*) The conserved hydrophobic residues composing the hydrophobic core behind the zinc-binding site are shown as ball-and-stick models. Ile193, which corresponds to Ala753 in mouse AlaRS, is coloured orange.

3.7. Structural comparison with AlaX-S and the ThrRS editing domain

The structure of the C-domain of AlaX-M is similar to those of AlaX-S and the ThrRS editing domain, except for the additional inserted helix-loop in AlaX-S and the ThrRS editing domain (Fig. 7; Dock-Bregeon *et al.*, 2000, 2004; Sokabe *et al.*, 2005; Ishijima *et al.*, 2006). In contrast, the structures of the N-domain of AlaX-M and that of the ThrRS editing domain are completely different and AlaX-S lacks an N-domain. The N-domain of the ThrRS editing domain has no loop structure near its catalytic site, unlike the glycine-rich loop in AlaX-M. These findings raise the possibility that the glycine-rich loop contributes to Gly-tRNA deacylation, which the AlaRS editing domain and AlaX-M can perform but not AlaX-S or the ThrRS editing domain. However, AlaX-L also contains the glycine-rich motif (Fig. 6), but does not exhibit editing activity (Ahel *et al.*, 2003). Oligomer formation by the oligomerization domain may inhibit the

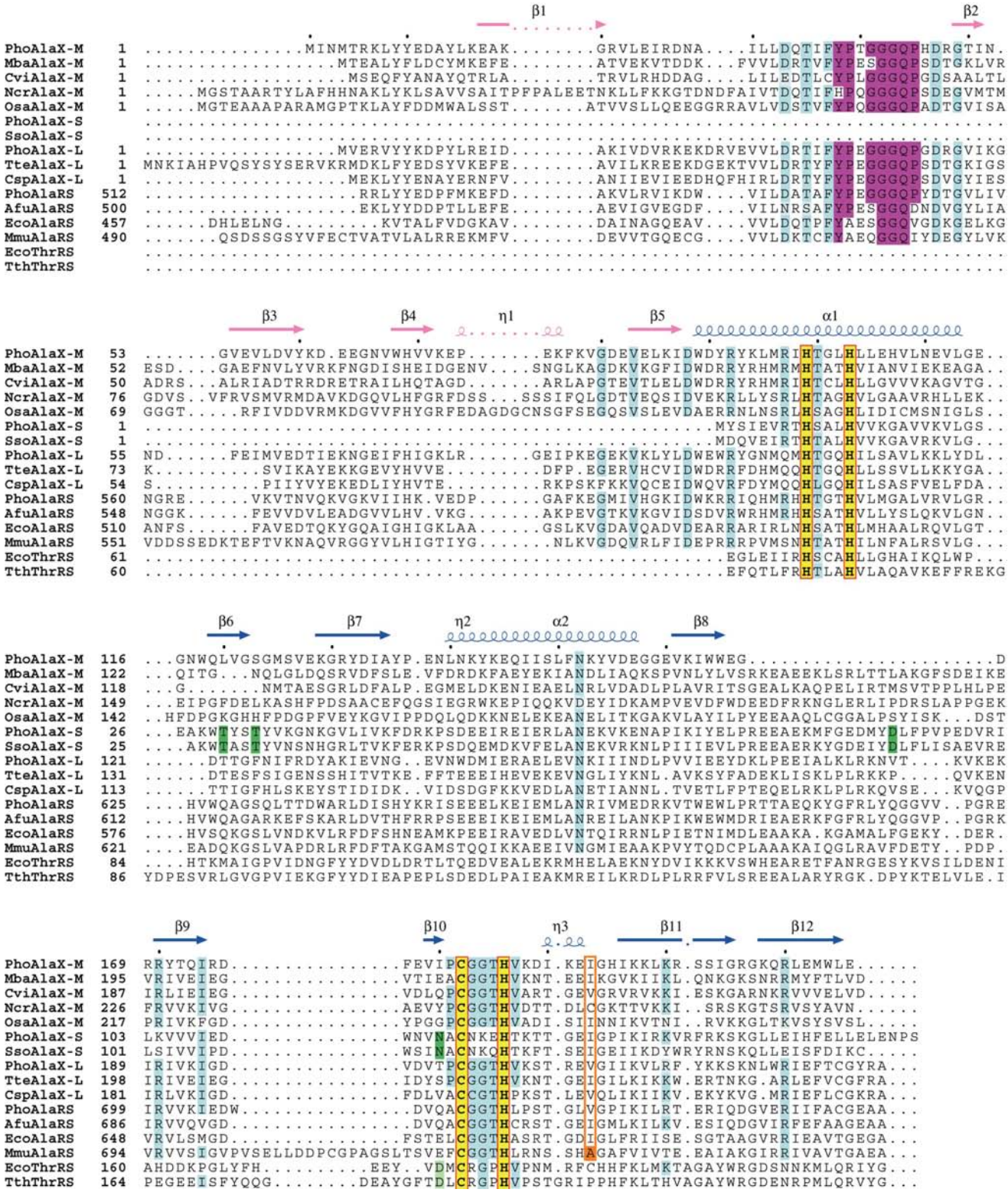

Figure 6
Sequence alignment of AlaX-M, AlaX-S, AlaX-L, the AlaRS editing domain and the ThrRS editing domain. The zinc-binding motifs are boxed in yellow. The glycine-rich motifs are boxed in violet. The residues that are >80% identical among the AlaX-M enzymes are boxed in light blue. Ala753 in mouse AlaRS, the mutation of which causes protein misfolding and neurodegeneration (Lee *et al.*, 2006), is boxed in orange. The residues that recognize the substrate amino-acid side chain are boxed in green and in light green in the AlaX-S enzymes and the ThrRS editing domains, respectively. The secondary structures of *P. horikoshii* AlaX-M are indicated at the tops of the columns in pink and blue for the N- and C-domains, respectively. Each protein was aligned using *ClustalW* and was manually refined. The species aligned are as follows: Pho, *P. horikoshii*; Mba, *M. barkeri*; Cvi, *Chromobacterium violaceum*; Ncr, *Neurospora crassa*; Osa, *Oryza sativa*; Sso, *S. solfataricus*; Tte, *Thermoanaerobacter tengcongensis*; Csp, *Clostridium* sp.; Afu, *A. fulgidus*; Eco, *E. coli*; Mmu, *Mus muscus*; Tth; *Thermus thermophilus*.

editing function in the absence of the aminoacylation domain.

The loop between the β8 and β9 sheets of AlaX-M, AlaX-S and the ThrRS editing domain contains an insertion composed of a loop and two helices (the inserted helix-loop). In AlaX-S, the loop region is long and the two short helices are perpendicular. In the ThrRS editing domain, the two helices are longer and antiparallel to each other and the loop is shorter. AlaX-S forms both a monomer and dimer (Sokabe *et al.*, 2005). In the dimeric structure, the inserted helix-loop is located at the dimer interface and contributes to dimerization. The conserved hydrophobic residues in the inserted helix-loop (Leu93 and Phe94) make hydrophobic interactions that stabilize the dimeric form of AlaX-S. The reason why *P. horikoshii* AlaX-M only exists as a monomer may be explained by its lack of the inserted helix-loop. Some of the other AlaX-M enzymes also have insertions in the corresponding region (Fig. 6). However, the low sequence conservation within the insertions in these AlaX-M enzymes suggests that the insertions do not cause the dimerization of the AlaX-M enzymes and thus all the AlaX-M proteins probably exist in the monomeric form.

3.8. Differences in the zinc-binding patterns

P. horikoshii AlaX-M coordinates the zinc ion using three residues, His99, His103 and Cys182, while His186 does not directly interact with the zinc ion (Fig. 8*a*). In contrast, in AlaX-S the zinc ion is coordinated by four residues, His9, His13, Cys116 and His120, regardless of the presence or absence of L-serine (Fig. 8*b*; Sokabe *et al.*, 2005). Thus, the positions of the zinc ions differ markedly between the two enzymes (Fig. 8*c*). The position corresponding to the location of the zinc ion in the AlaX-M structure contains the O atom of the α-carboxyl group of L-serine in the serine-bound AlaX-S structure. The difference in the locations of the zinc ions is attributable to differences in the zinc-binding site architecture. The side-chain rings of the His99 and His186 residues are rotated by 90 and 180° in AlaX-M compared with those of the corresponding His9 and His120 residues in AlaX-S, respectively. The side chain of Thr100 hydrogen bonds to the rotated His186 side chain as well as to the main-chain NH of Val187 in AlaX-M. In contrast, in AlaX-S the side chain of Ser10 (corresponding to Thr100 in AlaX-M) hydrogen bonds to the main-chain CO of Val6 as well as to the main-chain NH of Thr121 (corresponding to Val187 in AlaX-M). A zinc ion is not bound in the zinc-binding cluster of the ThrRS editing domain complexed with substrate analogues (Fig. 7; Dock-Bregeon *et al.*, 2004). Even after AlaX-S was treated with EDTA to remove the zinc ions by chelation, the AlaX-S enzyme hydrolyzed Ser-tRNA^{Ala} (Sokabe *et al.*, 2005). The authors argued that the zinc ion is not required for the editing activity of AlaX-S. The role of the zinc ion in AlaX-M, if any, should be clarified in the future.

It is not known whether the AlaRS editing domain coordinates a zinc ion in the conserved zinc-binding cluster and, if so, how it accomplishes this. The alanine mutation of the conserved cysteine residue in the zinc-binding cluster of *A. fulgidus* AlaRS abolished the editing activity (data not shown). A possible reason for the absence of activity is that the zinc ion may be essential for the editing reaction by AlaRS and the C703A mutation might abolish the zinc-binding activity. It is also possible that the conserved cysteine residue

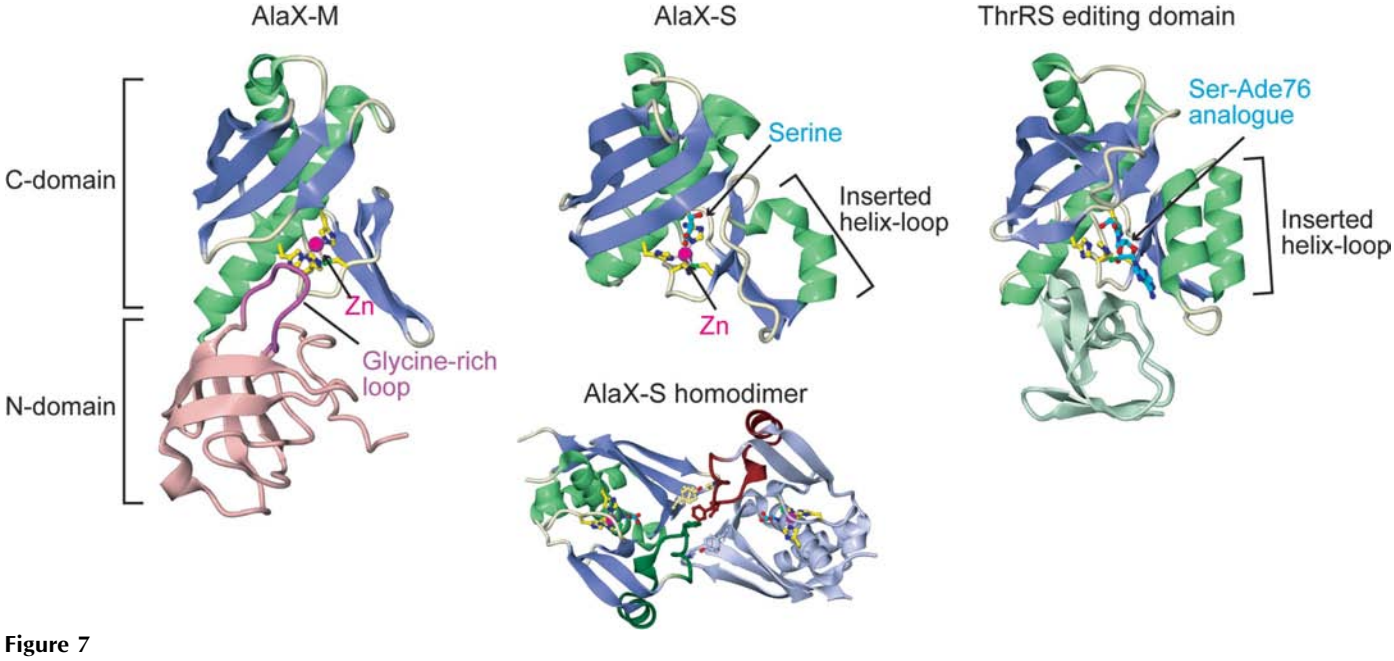

Figure 7
Structure comparison. Structures of *P. horikoshii* AlaX-M, *P. horikoshii* AlaX-S and the *E. coli* ThrRS editing domain are shown as ribbon models. The N-domains of AlaX-M and the ThrRS editing domain are coloured pink and light green, respectively. The zinc-binding residues are shown as yellow ball-and-stick models. The zinc ions are depicted as magenta spheres. The bound ligands are indicated as cyan ball-and-stick models. For AlaX-S, the homodimeric structure is also shown. The inserted helix-loops are coloured dark green and dark red in subunits *A* and *B*, respectively. The other region in subunit *B* is coloured light blue. The residues contributing to the dimerization are shown as ball-and-stick models.

may play an essential role in the substrate binding and/or catalysis.

3.9. tRNA-docking model

We attempted to construct a tRNA-docking model using AlaX-M. Firstly, we constructed the editing-state model of ThrRS in which the tRNA 3′-terminus enters the editing catalytic site, using the ThrRS–tRNA^{Thr} complex structure of the aminoacylation state (Sankaranarayanan *et al.*, 1999), by changing the conformation of the tRNA 3′-terminus (not shown). Next, by superposing the C-domain of AlaX-M onto that of the ThrRS editing domain, we constructed the AlaX-M–tRNA complex model (Fig. 9). The modelled tRNA fits quite well into the positively charged patch in the C-domain of AlaX-M (Fig. 9*a*). The main part of the N-domain is negatively charged, indicating that it does not interact with tRNA, which is consistent with the constructed model. In contrast, the glycine-rich loop is neutral and may participate in tRNA recognition (Fig. 9*a*). In our model, the glycine-rich loop is located in the vicinity of the terminal Ade76.

AlaX-M should discriminate the substrates Ser-tRNA^{Ala} and Gly-tRNA^{Ala} from the correct products Ser-tRNA^{Ser} and Gly-tRNA^{Gly} (Ahel *et al.*, 2003). Therefore, it should have specificity for tRNA^{Ala}. The G3–U70 wobble base pair in the acceptor arm is unique to tRNA^{Ala} and is used as an identity element in the alanylation reaction by AlaRS (Hou & Schimmel, 1988). It is possible that AlaX-M recognizes the G3–U70 wobble base pair in tRNA^{Ala}. AlaX-M does not seem to be large enough to contact the tRNA anticodon arm. In our model, the loop between β11 and β12 can contact the G3–U70 wobble base pair, as suggested for AlaX-S (Sokabe *et al.*, 2005; Fig. 9*b*).

3.10. Aminoacyl moiety recognition

We compared the substrate recognition sites in AlaX-M, AlaX-S and the ThrRS editing domain (Fig. 10). In AlaX-S, L-serine is bound in the presence of the zinc ion in the zinc-binding site (Fig. 10*b*; Sokabe *et al.*, 2005). In the ThrRS editing domain, an analogue of Ser-Ade76 is bound in the absence of a zinc ion (Fig. 10*c*). The residues in the zinc-binding clusters, His73, His77 and Cys182, recognize the Ser-Ade76 analogue. The substrate-recognition patterns differ between AlaX-S and the ThrRS editing domain. The residues recognizing serine in AlaX-S are conserved only among the AlaX-S enzymes (Fig. 6). Some of the residues recognizing the Ser-Ade76 analogue in the ThrRS editing domain are conserved only among ThrRSs (Fig. 6). In AlaX-M, the unique glycine-rich loop is located near the zinc-binding site. Therefore, the methods of substrate-recognition may be distinct between the AlaRS editing domain, AlaX-M, AlaX-S and the ThrRS editing domain.

We tried to determine the cocrystal structures of AlaX-M complexed with L-serine, glycine, a Ser-AMP analogue and a Gly-AMP analogue. We also tried soaking these ligands with the apo-form crystals. However, none of the ligands were observed in the structures determined. In AlaX-M, it is unclear whether the substrates are bound in the presence or absence (or both) of the zinc ion. It is also possible that the substrate-specificity (Ser-tRNA^{Ala} or Gly-tRNA^{Ala}) depends on the presence/absence of the zinc ion. It will be interesting to determine the cocrystal structures of AlaX-M complexed with analogues of Ser-Ade76 and Gly-Ade76 and to elucidate the roles of the glycine-rich loop and the zinc ion, if any.

The C703A mutant of *A. fulgidus* AlaRS did not exhibit editing activity. The recognition of the substrate α-amino group by the completely conserved Asp residue and the abolition of the editing activity by the replacement of the Asp residue by alanine are seen in the class I editing aaRSs (isoleucyl-tRNA, valyl-tRNA and leucyl-tRNA synthetase; Lincecum *et al.*, 2003; Fukunaga & Yokoyama, 2005, 2006). It is possible that the conserved cysteine residue (Cys182 in

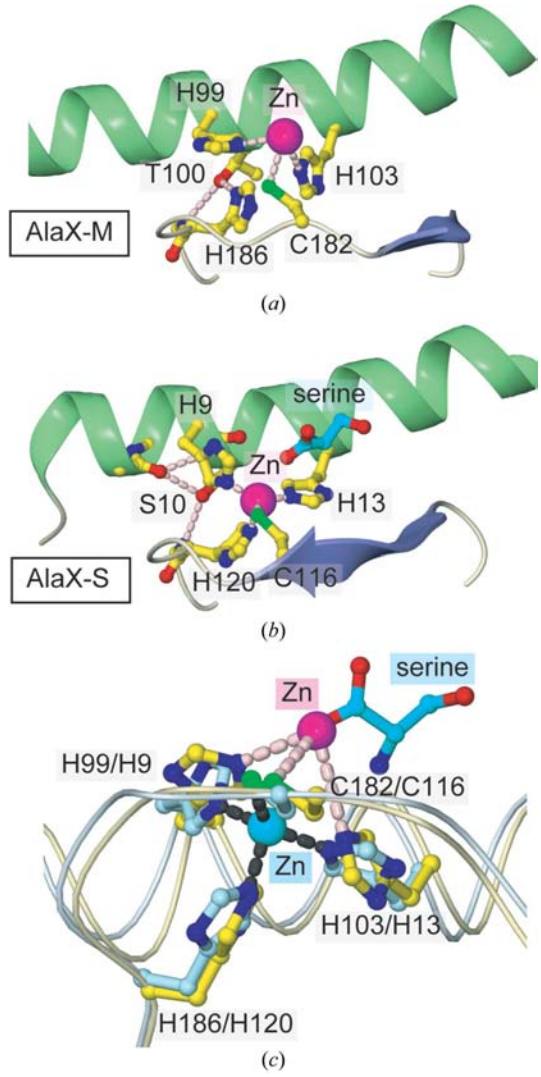

Figure 8
Different zinc-binding patterns. (*a*) The zinc-binding site in AlaX-M. (*b*) The zinc-binding site in AlaX-S. The bound serine is shown as a cyan ball-and-stick model. (*c*) Superposition of the zinc-binding sites in AlaX-M (yellow) and AlaX-S (light blue). The zinc ions in the AlaX-M and AlaX-S structures are coloured magenta and cyan, respectively.

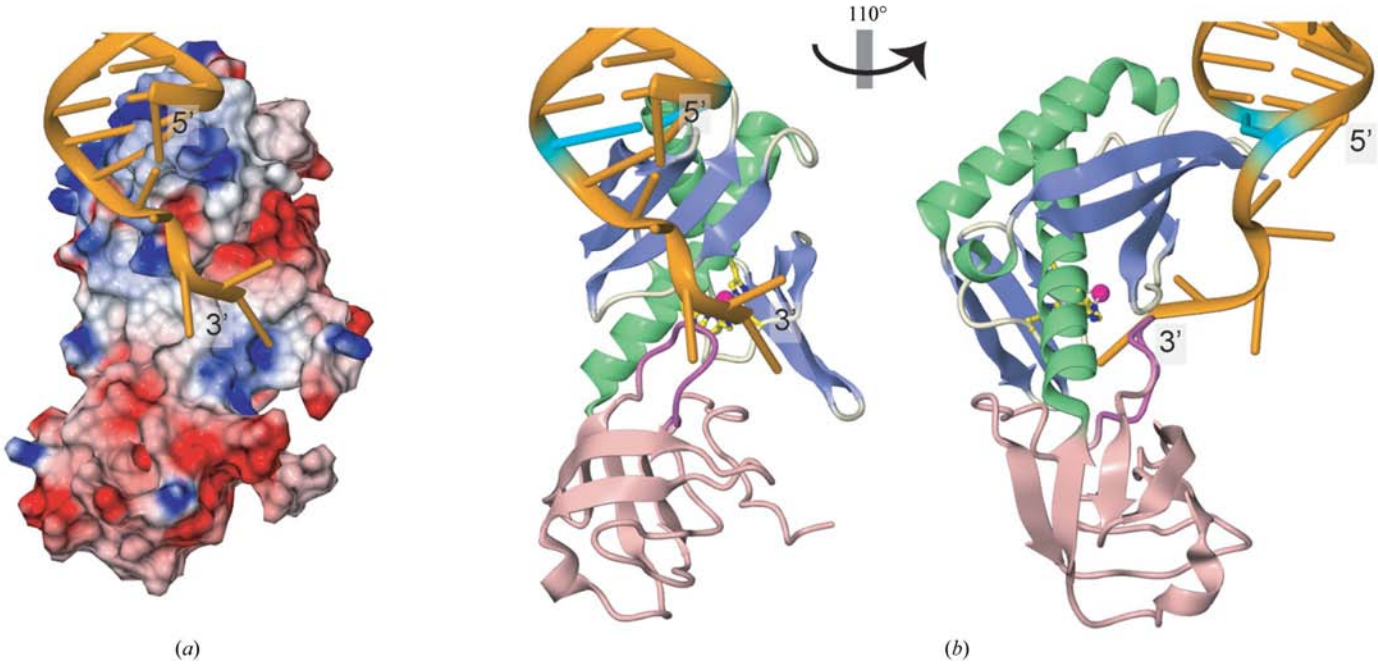

Figure 9
tRNA-binding model. (*a*) tRNA-binding model (orange) on a surface electropotential model of AlaX-M. Only the tRNA acceptor arm is shown. (*b*) The N-domain is coloured pink and the glycine-rich loop is coloured violet. The zinc-binding residues and the zinc ion are shown as yellow ball-and-stick models and as a magenta sphere, respectively. The tRNA model is shown in orange. The third base pair in the acceptor stem, corresponding to the G3–U70 wobble base pair in $tRNA^{Ala}$, is coloured cyan.

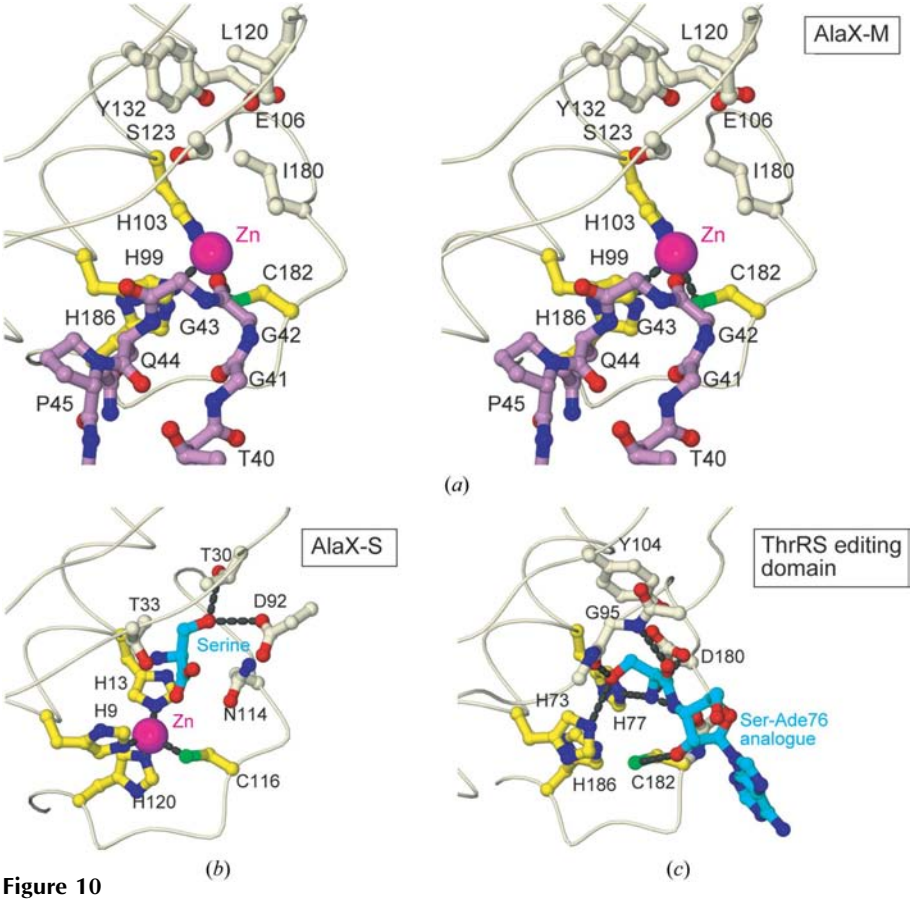

Figure 10
Comparison of substrate-recognition sites. (*a*) The putative substrate-recognition site in AlaX-M (stereoview). The zinc-binding residues are coloured yellow. The glycine-rich loop is coloured violet. (*b*) Serine recognition in AlaX-S. (*c*) The recognition of a Ser-Ade76 analogue in the ThrRS editing domain.

P. horikoshii AlaX-M) or the zinc ion bound by the cysteine residue recognizes the α-amino group of the substrates.

4. Conclusions

The crystal structure of the *trans*-editing enzyme *P. horikoshii* AlaX-M has been solved at 2.7 Å resolution. It consists of N- and C-domains. The glycine-rich motif in the N-domain forms a loop (the glycine-rich loop) and is located near the zinc-binding catalytic site in the C-domain. The glycine-rich loop may be involved in substrate (Ser-$tRNA^{Ala}$ and Gly-$tRNA^{Ala}$) recognition. A reasonable docking model for tRNA has been constructed. Since AlaX-M shares high sequence homology with the editing domain of AlaRS, the structure also provides insight into the editing mechanism of AlaRS. The structures of the N- and C-domains, the hydrogen-bonding interactions between the two domains and the intriguing occupation and conformation of the glycine-rich loop must be similar to those of the AlaRS editing domain. The determination of the structure of the complex of AlaX-M with $tRNA^{Ala}$ will reveal the

tRNA-specificity mechanism. In addition, the structures of AlaX-M complexed with analogues of Ser-Ade76 and Gly-Ade76 will clarify the precise mechanism of the dual specificity.

We thank Drs S. Sekine, T. Ito (University of Tokyo), T. Yanagisawa, T. Sengoku, R. Ishii (RIKEN), M. Kawamoto, N. Shimizu and H. Sakai (JASRI) for their help in data collection at SPring-8. This work was supported by Grants-in-Aid for Scientific Research in Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, the RIKEN Structural Genomics/Proteomics Initiative (RSGI) of the National Project on Protein Structural and Functional Analyses, MEXT. RF was supported by Research Fellowships from the Japan Society for the Promotion of Science for Young Scientists.

References

Ahel, I., Korencic, D., Ibba, M. & Soll, D. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 15422–15427.
Andersen, G. R., Thirup, S., Spremulli, L. L. & Nyborg, J. (2000). *J. Mol. Biol.* **297**, 421–436.
Beebe, K., Ribas De Pouplana, L. & Schimmel, P. (2003). *EMBO J.* **22**, 668–675.
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.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dock-Bregeon, A.-C., Rees, B., Torres-Larios, A., Bey, G., Caillet, J. & Moras, D. (2004). *Mol. Cell*, **16**, 375–386.
Dock-Bregeon, A.-C., Sankaranarayanan, R., Romby, P., Caillet, J., Springer, M., Rees, B., Francklyn, C. S., Ehresmann, C. & Moras, D. (2000). *Cell*, **103**, 877–884.
Fukunaga, R., Ishitani, R., Nureki, O. & Yokoyama, S. (2005). *Acta Cryst.* F**61**, 30–32.
Fukunaga, R. & Yokoyama, S. (2005). *J. Biol. Chem.* **280**, 29937–29945.
Fukunaga, R. & Yokoyama, S. (2006). *J. Mol. Biol.* **359**, 901–912.
Gibrat, J. F., Madej, T. & Bryant, S. H. (1996). *Curr. Opin. Struct. Biol.* **6**, 377–385.
Giegé, R., Kern, D., Ebel, J.-P., Grosjean, H., de Henau, S. & Chantrenne, H. (1974). *Eur. J. Biochem.* **45**, 351–362.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Hou, Y. M. & Schimmel, P. (1988). *Nature (London)*, **333**, 140–145.
Ibba, M. & Soll, D. (2000). *Annu. Rev. Biochem.* **69**, 617–650.
Ishijima, J., Uchida, Y., Kuroishi, C., Tuzuki, C., Takahashi, N., Okazaki, N., Yutani, K. & Miyano, M. (2006). *Proteins*, **62**, 1133–1137.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Lee, J. W., Beebe, K., Nangle, L. A., Jang, J., Longo-Guess, C. M., Cook, S. A., Davisson, M. T., Sundberg, J. P., Schimmel, P. & Ackerman, S. L. (2006). *Nature (London)*, **443**, 50–55.
Lincecum, T. L., Tukalo, M., Yaremchuk, A., Mursinna, R. S., Williams, A. M., Sproat, B. S., Van Den Eynde, W., Link, A., Van Calenbergh, S., Grotli, M., Martinis, S. A. & Cusack, S. (2003). *Mol. Cell*, **11**, 951–963.
Martinis, S. A., Plateau, P., Cavarelli, J. & Florentz, C. (1999). *Biochimie*, **81**, 683–700.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Roll-Mecak, A., Cao, C., Dever, T. E. & Burley, S. K. (2000). *Cell*, **103**, 781–792.
Sankaranarayanan, R., Dock-Bregeon, A.-C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B. & Moras, D. (1999). *Cell*, **97**, 371–381.
Schmidt, E. & Schimmel, P. (1994). *Science*, **264**, 265–267.
Sokabe, M., Okada, A., Yao, M., Nakashima, T. & Tanaka, I. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 11669–11674.
Swairjo, M. A. & Schimmel, P. R. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 988–993.
Weeks, C. M. & Miller, R. (1999). *Acta Cryst.* D**55**, 492–500.