Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the nondiscriminating aspartyl-tRNA synthetase from the crenarchaeon *Sulfolobus tokodaii* strain 7 reveals the recognition mechanism for two different tRNA anticodons

Yoshiteru Sato,[a] Yohei Maeda,[a] Satoru Shimizu,[a] Md Tofazzal Hossain,[a] Souichirou Ubukata,[b] Kaoru Suzuki,[b] Takeshi Sekiguchi[b] and Akio Takénaka[a]*

[a]Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226-8501, Japan, and [b]College of Science and Engineering, Iwaki-Meisei University, Chuou-dai-Iino, Iwaki, Fukushima 970-8551, Japan

Correspondence e-mail: atakenak@bio.titech.ac.jp

Received 16 April 2007
Accepted 3 August 2007

PDB Reference: aspartyl-tRNA synthetase, 1wyd, r1wydsf.

In protein synthesis, 20 types of aminoacyl-tRNA synthetase (aaRS) are generally required in order to distinguish between the 20 types of amino acid so that each achieves strict recognition of the cognate amino acid and the cognate tRNA. In the crenarchaeon *Sulfolobus tokodaii* strain 7 (*St*), however, asparaginyl-tRNA synthetase (AsnRS) is missing. It is believed that AspRS instead produces Asp-tRNA^{Asn} in addition to Asp-tRNA^{Asp}. In order to reveal the recognition mechanism for the two anticodons, GUC for aspartate and GUU for asparagine, the crystal structure of *St*-AspRS (nondiscriminating type) has been determined at 2.3 Å resolution as the first example of the nondiscriminating type of AspRS from crenarchaea. A structural comparison with structures of discriminating AspRSs indicates that the structures are similar to each other overall and that the catalytic domain is highly conserved as expected. In the N-terminal domain, however, the binding site for the third anticodon nucleotide is modified to accept two pyrimidine bases, C and U, but not purine bases. The C base can bind to form a hydrogen bond to the surrounding main-chain amide group in the discriminating AspRS, while in the nondiscriminating AspRS the corresponding amino-acid residue is replaced by proline, which has no amide H atom for hydrogen-bond formation, thus allowing the U base to be accommodated in this site. In addition, the residues that cover the base plane are missing in the nondiscriminating AspRS. These amino-acid changes make it possible for both C and U to be accepted by the nondiscriminating AspRS. It is speculated that this type of nondiscriminating AspRS has been introduced into *Thermus thermophilus* through horizontal gene transfer.

1. Introduction

In general, aminoacylation of tRNA is performed by aminoacyl-tRNA synthetase (aaRS[1]) through the aminoacylation of AMP and transfer of the aminoacyl group to tRNA. In order to ensure translation accuracy, 20 types of aaRSs are required to distinguish between the 20 types of amino acid. In order to attach a specific amino acid (aa) to the CCA terminus of the corresponding tRNA (tRNA^{aa}), each aaRS must recognize the amino acid and the anticodon and/or other characteristic parts of the tRNA^{aa} (Cavarelli & Moras, 1993; Cusack, 1997; Beuning & Musier-Forsyth, 1999). Based on their distinct catalytic domains, aaRSs are divided into two classes (I and II), with ten enzymes in each (Eriani *et al.*, 1990; Cusack *et al.*, 1990). For instance, AspRS (aspartyl-tRNA^{Asp} synthetase)

[1] RS and the variable aa represent aminoacyl-tRNA synthetase and amino-acid three-letter code, respectively.

and AsnRS belong to class II, while GluRS and GlnRS belong to class I.

Recent genomic analyses have revealed that some organisms constitute an exception in which the correspondence between amino acid and aaRS is not complete (Kawarabayasi *et al.*, 1999, 2001). In the crenarchaea *Sulfolobus tokodaii* strain 7 (*St*) and *Aeropyrum pernix* K1 (*Ap*), the genes for AsnRS (class II) and GlnRS (class I) are missing from their genomes (http://www.bio.nite.go.jp/dogan). In such organisms, it is believed that AspRS (class II) catalyzes the reaction producing Asp-tRNA^{Asn} as well that producing Asp-tRNA^{Asp} (Becker & Kern, 1998; Curnow *et al.*, 1998). Subsequently, the aspartyl group attached to tRNA^{Asn} is converted to the asparaginyl group by another enzyme, GatCAB (Curnow *et al.*, 1998). In this case, AspRS only catalyzes the aspartylation of AMP, but it must accept two different anticodons, GUC and GUU for Asp and Asn, respectively, by excluding the other anticodons (Becker & Kern, 1998). Gln-tRNA^{Gln} is produced in a similar way in class I (Curnow *et al.*, 1998). To distinguish them from the normal or original aaRSs (the discriminating type), they are referred to as the nondiscriminating type (Curnow *et al.*, 1998).

Some bacteria such as *Thermus thermophilus* (*Tt*) and *Deinococcus radiodurans* (*Dr*) were found to contain genes for both the discriminating type and the nondiscriminating type (Becker & Kern, 1998; Curnow *et al.*, 1998). However, *St* and *Ap* have only the nondiscriminating type, suggesting that they are the true ancestor in which the nondiscriminating system originated (Shiba *et al.*, 1998). To establish the recognition mechanism for the two anticodons as well as their evolutionary relationship, we initiated X-ray analyses of the nondiscriminating AspRS and GluRS from *St* and *Ap*. This paper will describe and discuss the structural details of *St*-AspRS and compare them with those of the discriminating AspRS.

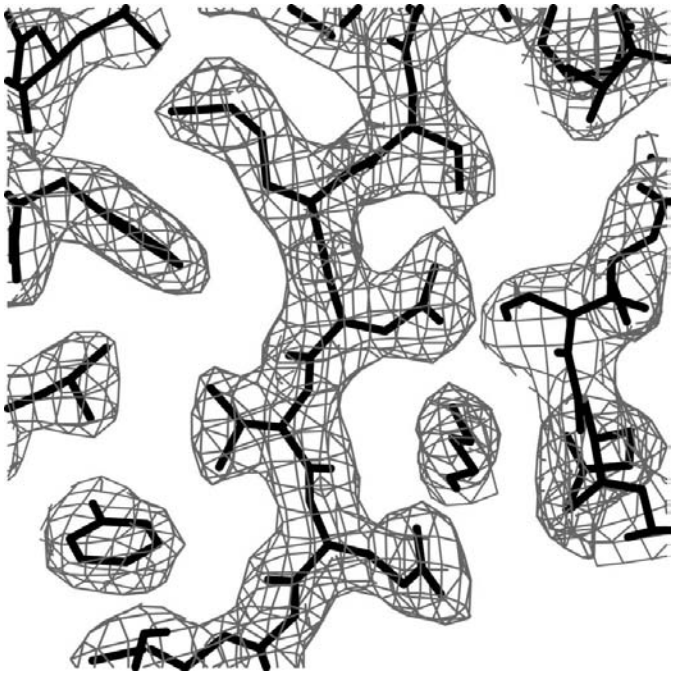

Figure 1
A $2F_o - F_c$ map around the Asp227 residue, contoured at $\rho = 1.2\sigma$.

Table 1
Statistics of data processing and structure refinement.

Values in parentheses are for the outer shell.

Data processing	
Resolution (Å)	38.0–2.3
Observed reflections	394690
Unique reflections	54522
Completeness (%)	99.2 (99.2)
R_{merge}† (%)	8.6 (26.2)
$I/\sigma(I)$	5.8 (1.5)
Average multiplicity	7.2
Space group	$P2_12_12$
Unit-cell parameters (Å)	
a	116.0
b	139.3
c	75.3
Z‡	2
Structure refinement	
Resolution range (Å)	30–2.3
Reflections used	48969
R factor§ (%)	22.9
R_{free}¶ (%)	26.5
No. of protein atoms	6789
No. of HEPES molecules	2
No. of sulfate ions	2
No. of chloride ions	2
No. of water molecules	220
R.m.s. deviations	
Bond lengths (Å)	0.02
Bond angles (°)	1.5
Between subunits (Å)	0.4

† $R_{merge} = 100 \times \sum_{hj} |I_{hj} - \langle I_h \rangle| / \sum_{hj} I_{hj}$, where I_{hj} is the *j*th measurement of the intensity of reflection **h** and $\langle I_h \rangle$ is its mean value. ‡ *Z* is the number of subunits in the asymmetric unit. § *R* factor $= 100 \times \sum ||F_o| - |F_c|| / \sum |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure-factor amplitudes, respectively. ¶ Calculated using a random set containing 10% of observations that were not included throughout refinement (Brünger, 1992).

2. Methods and materials

The *AspRS* gene of *S. tokodaii* strain 7 was overexpressed in *Escherichia coli* BL21(DE3) using plasmid pET-11a and the protein was purified according to the method described previously (Suzuki *et al.*, 2007). Crystals suitable for X-ray experiments were obtained under the previously described conditions. X-ray data were collected at 100 K using synchrotron radiation ($\lambda = 1.00$ Å) at beamline NW12A of PF (Tsukuba, Japan). Diffraction data were recorded on a 210 × 210 mm CCD detector array (ADSC Q210) positioned 200 mm from the crystal. Each frame was taken with 1° oscillation and 1 s exposure with an ω range of 0–180°. Diffraction patterns were processed at 2.3 Å resolution using the program *CrystalClear* (Rigaku/MSC) and intensities were converted to amplitudes using programs from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). Detailed data-processing statistics are given in Table 1.

Initial phases were derived by molecular replacement using the *Tk*-AspRS structure from *Thermococcus kodakaraensis* (PDB code 1b8a; Schmitt *et al.*, 1998) as a search model in *AMoRe* (Navaza, 1994). The molecular structure was built into the electron-density map using the program *QUANTA*

(Accelrys Inc.). After rigid-body refinement, the atomic parameters were refined with the programs *REFMAC* (Murshudov *et al.*, 1997) in *CCP4i* (Potterton *et al.*, 2003) and *CNS* (Brünger *et al.*, 1998) through a combination of rigid-body, simulated-annealing, crystallographic conjugate-gradient minimization, *B*-factor and maximum-likelihood refinement. The structure was revised by interpreting OMIT maps at every residue. 220 water molecules, two chloride ions, two HEPES molecules and two sulfate ions that were found in $|F_o| - |F_c|$ maps ($>2.5\sigma$) were included in the final refinement. The final *R* factor was 22.9% for 30–2.3 Å resolution data (R_{free} = 26.5 for 10% of the observed data). Statistical data for the structure determination are given in Table 1. Fig. 1 shows a $2F_o - F_c$ map drawn using the program *O* (Jones *et al.*, 1991). For structural comparisons, the program *LSQMAN* (Kleywegt & Jones, 1994) was used to superimpose corresponding C^α atoms. Figs. 2, 4, 5 and 6 were produced using *PyMOL* (http://pymol.sourceforge.net).

3. Results and discussion

3.1. Overall structure

St-AspRS is a homodimeric enzyme, with each subunit comprising 429 residues (see Fig. 2). The two subunits are crystallographically independent, but their structures are similar to each other, with an r.m.s.d. of 0.38 Å between the corresponding C^α atoms. Therefore, only one of the two subunits will be discussed below. Comparisons of the structure with those reported previously (Ruff *et al.*, 1991; Schmitt *et al.*, 1998; Charron *et al.*, 2003) make it possible to describe the following structural features. Fig. 3 shows the topology of the secondary structure. The subunit consists of an N-terminal anticodon-binding domain (residues 1–96) and a C-terminal catalytic domain (residues 128–429). These two domains are linked together by the hinge domain (residues 97–127).

The anticodon-binding domain adopts a standard OB-fold (Murzin, 1993) formed by a five-stranded β-barrel (S1–S5) with an α-helix (Hα, residues 57–60) between strands S3 and S4. The L1 loop (residues 163–171) between S4 and S5 protrudes, suggesting that it can make contact with the anticodon loop of tRNA. The catalytic domain is easily identified from the characteristic class II motifs 1 (residues 145–160), 2 (residues 198–223) and 3 (residues 389–422) (Eriani *et al.*, 1990). As shown in Fig. 4, this domain is composed of a six-stranded antiparallel β-sheet typical of class II aaRSs (Cusack *et al.*, 1990; Eriani *et al.*, 1995). The flipping loop (residues 163–171) is essential for the binding of ATP and for catalytic action (Schmitt *et al.*, 1998). The catalytic site contains space for the specific binding of aspartic acid to generate aspartyl-AMP in the first reaction; the aspartyl group

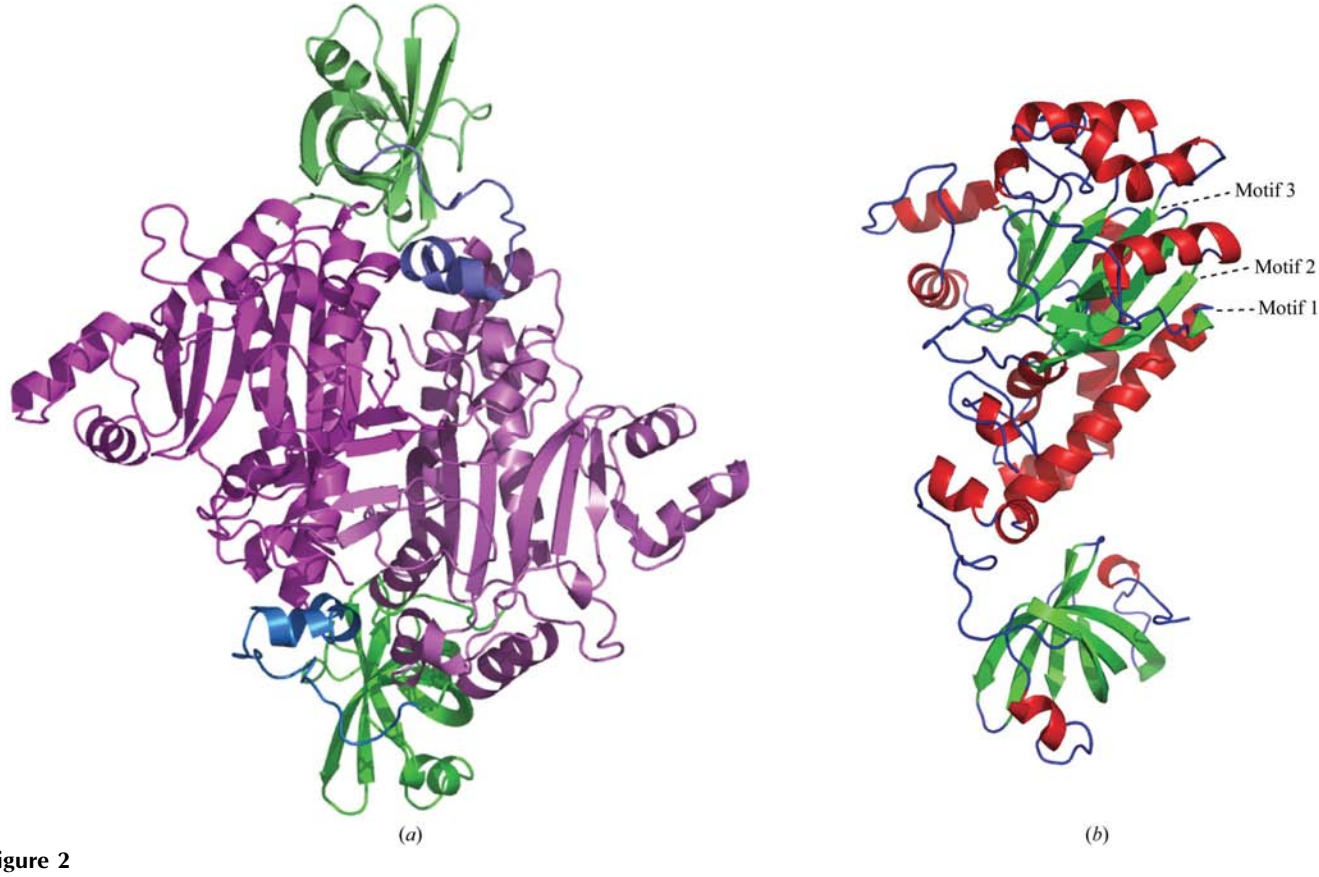

Figure 2
Ribbon drawings of dimeric *St*-AspRS (*a*) and its subunit structure (*b*). In (*a*), the N-terminal, hinge and C-terminal domains are coloured green, blue and purple, respectively. In (*b*), β-strands, helices and loops are coloured light green, orange and blue, respectively. The catalytic domain is found in the C-terminal domain containing the three motifs 1, 2 and 3 characteristic of the class II aaRS. The N-terminal domain adopts the OB-folding suggestive of the anticodon-binding domain.

is transferred to the CCA terminus of the cognate $tRNA^{Asp}$ in the second reaction. In the present crystal, however, this space is occupied by several water molecules and a HEPES molecule, since the protein was crystallized without substrates.

3.2. Comparison with a discriminating AspRS

To characterize the differences of the structure from the discriminating AspRSs, the *St*-AspRS structure is compared with those reported previously, in particular *Tk*-AspRS from the archaeon *T. kodakaraensis* KOD1 (Schmitt *et al.*, 1998). The enzymes are both dimeric and are similar to each other. When superimposed on each other, the r.m.s.d.s of corresponding C^{α} atoms are 1.81, 0.94 and 1.20 Å between the subunits, between the anticodon-binding domains and between the catalytic domains, respectively. These values indicate that the relative orientation of the two domains differs slightly from that in *Tk*-AspRS at the flexible hinge domain responsible for tRNA binding.

In the catalytic domain, the essential amino-acid residues, with the exception of motifs 1, 2 and 3, are highly conserved even in the nondiscriminating AspRS. When the active-site residues are superimposed between *St*-AspRS and *Tk*-AspRS, these residues fit well to each other, as shown in Fig. 4. Therefore, the binding geometries of ATP, aspartate and aspartyl-AMP in *Tk*-AspRS (Schmitt *et al.*, 1998) are expected to be applicable to the present nondiscriminating enzyme. The residues Glu167, Gln189, Lys192, Arg210, Glu212, Asp227, Glu352, Ser355, Arg359, Arg403 are highly conserved in the active site of both enzymes. It is surprising to see that Gln189, Lys192, Ser355 and Arg359 maintain their side-chain conformations to recognize aspartate. The side-chain conformation of Glu352 is also maintained to bind Mg^{2+} ions for ATP binding. The remaining residues are likely to become reoriented when ATP binds. These features of the nondiscriminating enzyme are reasonable as this domain only acts in aspartylation.

The OB-fold of the anticodon-binding domain is also similar to that of *Tk*-AspRS. However, this domain must differ in its anticodon recognition because *St*-AspRS accepts the two different tRNAs for aspartate and asparagine, whereas *Tk*-AspRS only

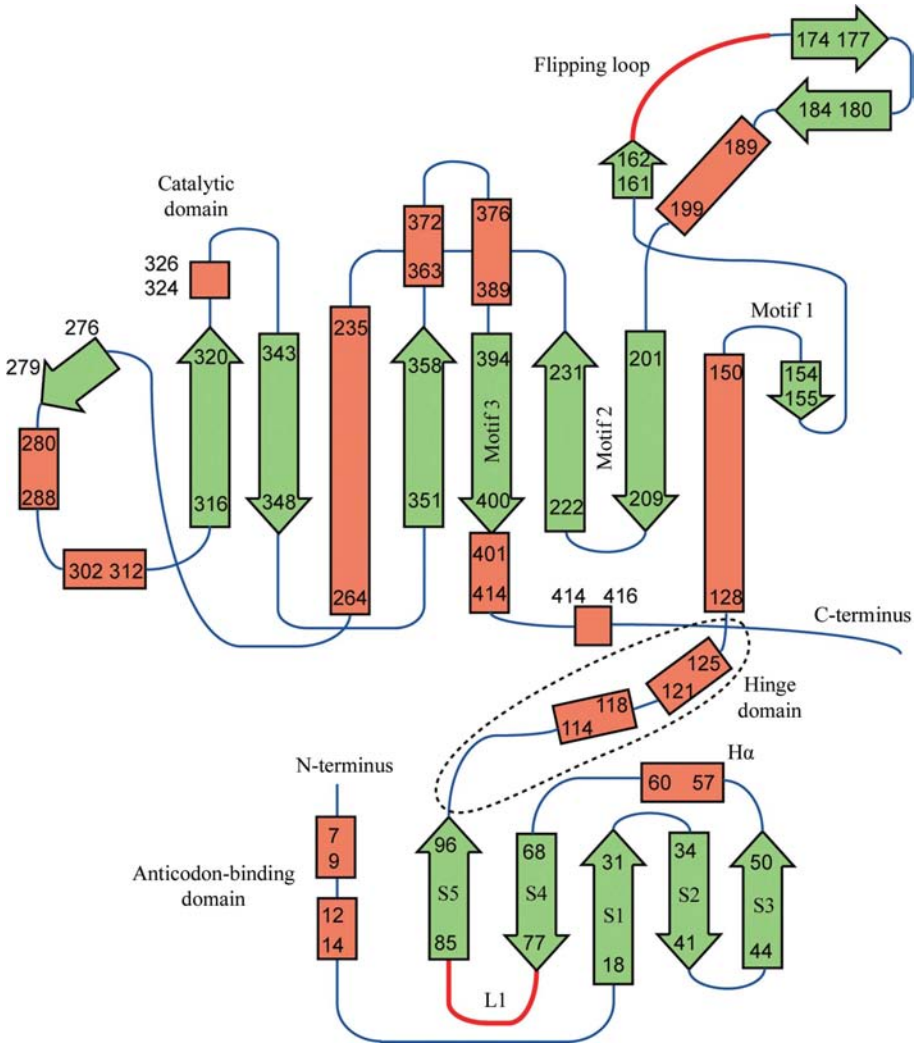

Figure 3
A topological diagram of the secondary structure of *St*-AspRS. The secondary structure was assigned using the program *PROCHECK* (Laskowski *et al.*, 1993). β-Strands, helices and loops are drawn as light green arrows, as orange rods and as blue lines, respectively. The N-terminal and C-terminal domains are linked by a hinge domain.

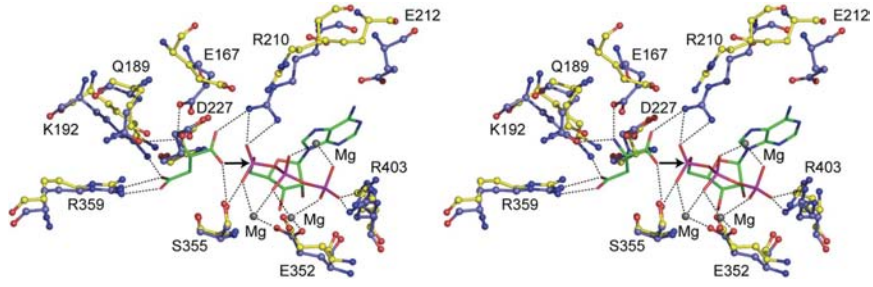

Figure 4
Stereoview of the catalytic sites superimposed between *St*-AspRS (yellow) and *Tk*-AspRS (blue). The bound aspartate and ATP are coloured green and red. The grey spheres indicate bound magnesium cations. The hydrogen-bonding interactions in *Tk*-AspRS are shown as dotted lines. The residues from St are labelled.

accepts $tRNA^{Asp}$. In the *St* genome only one codon is found for aspartate and one for asparagine, the anticodons being GUC and GUU, respectively (Kawarabayasi *et al.*, 2001). The difference between the anticodons occurs at the third nucleotide. Therefore, the nondiscriminating AspRS must accept C and U but not G and A at the third residue. In order to discuss the recognition mechanism, tRNA-binding models of *St*-AspRS and *Tk*-AspRS have been constructed based on the structure of *Sc*-AspRS in complex with *Sc*-$tRNA^{Asp}$ (PDB code 1asy; Ruff *et al.*, 1991).

3.3. Anticodon recognition

Fig. 5 shows the tRNA-binding models of *St*-AspRS and *Tk*-AspRS, together with the X-ray structure of *Sc*-AspRS bound to the cognate tRNA. In the discriminating *Sc*-AspRS and *Tk*-AspRS the third anticodon cytosine residue (C36) is covered by the L1 loop (Fig. 5*b*). In the nondiscriminating *St*-AspRS, however, one side of the cytosine base appears to be exposed to the outside. The detailed structures suggest that in the discriminating AspRS, the C36 residue of tRNA is recognized through hydrogen bonds between N4 of the cytosine base and the main-chain carbonyl O atom of Lys83 and between N3 of the cytosine base and the main-chain amide group of Lys85, similar to the case in *Sc*-AspRS. On the other hand, in the nondiscriminating *St*-AspRS (Fig. 5*c*) the C36 base cannot form such hydrogen bonds, since the nearest residue is Pro82, which has no amide proton. Furthermore, the Pro82 and Ala81 residues in the L1 loop shift away from the third residue of the anticodon to make a space to accept a uracil base as well as a cytosine base. Similar proline replacements at this site are conserved in other nondiscriminating AspRSs (Charron *et al.*, 2003; Feng *et al.*, 2005). These local changes might make it possible for the normal (discriminating) AspRS to accept the two different anticodons of $tRNA^{Asp}$ and $tRNA^{Asn}$. In order to confirm this hypothesis, it is necessary to analyze the crystal structures of *St*-AspRS in complex with $tRNA^{Asp}$ and $tRNA^{Asn}$.

Thermus thermophilus (*Tt*) is an unusual organism in that it contains two genes for AspRS in the genome, AspRS-1 being the discriminating type and AspRS-2 the nondiscriminating type (Charron *et al.*, 2003). The structure of *Tt*-AspRS-2 is similar to that of *St*-AspRS and the residue corresponding to Pro72 is also conserved. From an L1 loop-swapping experiment based on a mutual comparison between the AspRS-1 and AspRS-2 structures, Feng *et al.* (2005) proved that the proline residue in question is a crucial residue for the nondiscriminating function.

Suzuki *et al.* (2007) showed that the sequence identity between *St*-AspRS and *Tt*-AspRS-2 is higher than that between *Tt*-AspRS-1 and *Tt*-AspRS-2. This suggests that the nondiscriminating AspRS of organisms that employ two

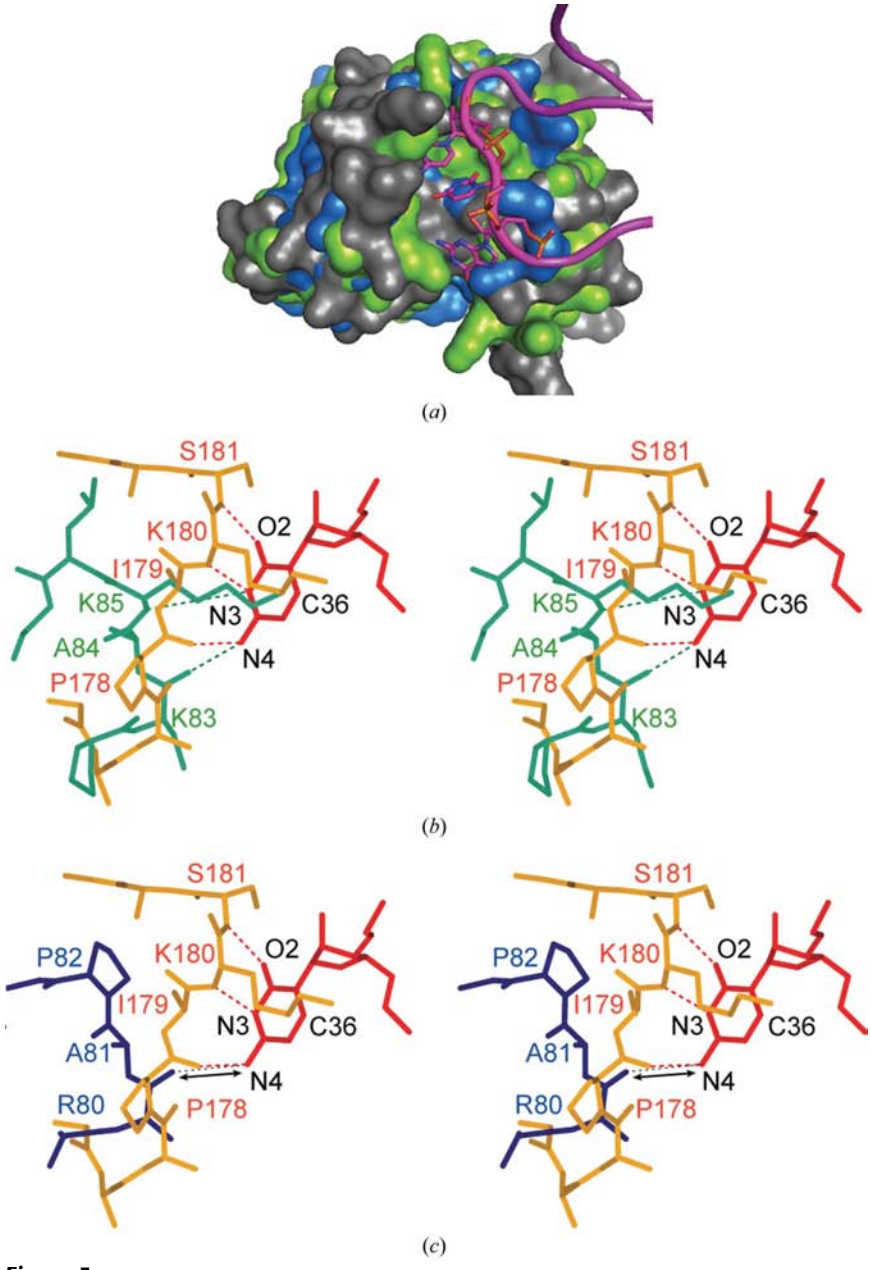

Figure 5
The anticodon loop of tRNA bound in the N-terminal domains. (*a*) Structural models of *St*-AspRS (blue) and *Tk*-AspRS (green) constructed by superimposition of the N-terminal domains onto the observed structure of *Sc*-AspRS in complex with the cognate $tRNA^{Asp}$ (grey). The detailed structures of *Tk*-AspRS (green) and *St*-AspRS (blue) are superimposed on that of *Sc*-AspRS (yellow) in (*b*) and (*c*), respectively. The broken lines indicate hydrogen-bonding interactions in *Sc*-AspRS and those expected in the other AspRSs. The residues of *St*, *Tk* and *Sc* are coloured blue, green and orange, respectively. The r.m.s.d.s of corresponding C^{α} atoms are 1.23 Å between *St*-AspRS and *Sc*-AspRS and 1.38 Å between *Tk*-AspRS and *Sc*-AspRS.

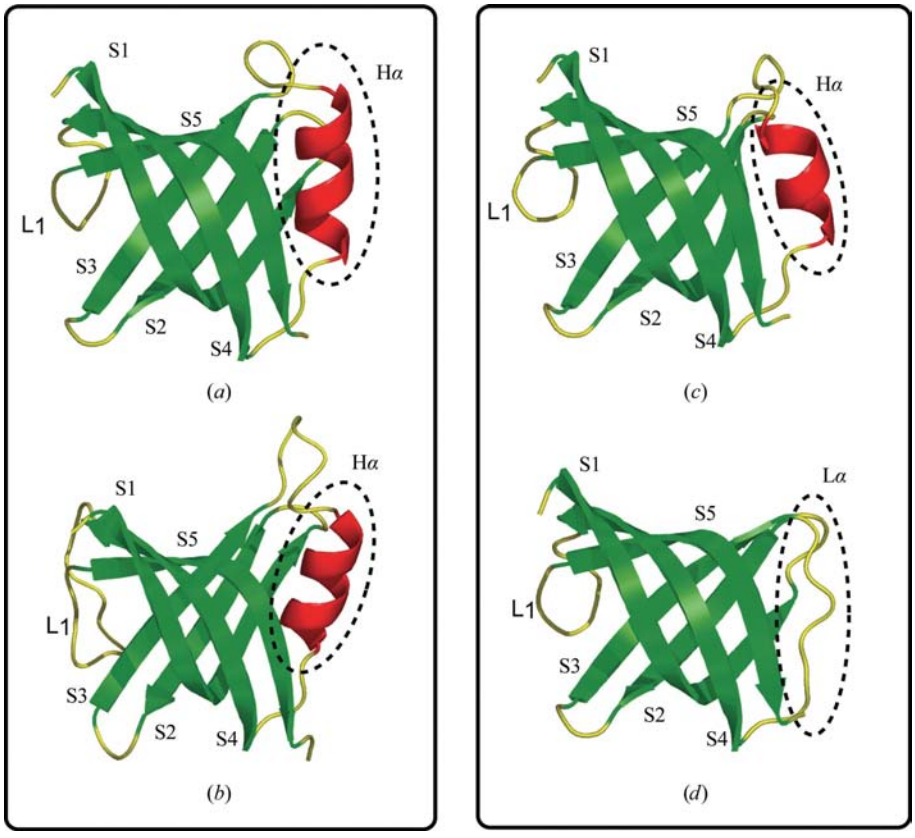

Figure 6
The Hα helices behind the anticodon-binding sites found in the discriminating *Tk*-AspRS and *Sc*-AspRS (left column, *a* and *b*, respectively) and in the nondiscriminating *St*-AspRS (right column *c*). The corresponding part is rewound in *Tt*-AspRS. Some amino-acid residues at the N-terminal ends are omitted for clarity: 14 in *Tk*-AspRS, 104 in *Sc*-AspRS, 14 in *St*-AspRS and 11 in *Tt*-AspRS (*d*).

AspRS systems came from crenarchaeota through horizontal gene transfer. However, there is a significant difference between the two structures in the anticodon-binding domain. In all AspRS structures solved to date (discriminating and nondiscriminating), an Hα/Lα motif exists on the opposite side of the anticodon-recognition site (see Fig. 6), with the exception of *Tt*-AspRS-2. This suggests that the Hα helix in *Tt*-AspRS-2 is rewound into a loop. Such a structural change may, however, be irrelevant to the evolution of the nondiscriminating function.

We thank N. Igarashi and S. Wakatsuki for facilities and help during data collection, E. C. M. Juan for proofreading the original manuscript and D. Moras for giving us valuable comments. This work was supported in part by the Grants-in-Aid for Protein3000 Research Program from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

Becker, H. D. & Kern, D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 12832–12837.
Beuning, P. J. & Musier-Forsyth, K. (1999). *Biopolymers*, **52**, 1–28.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
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.
Cavarelli, J. & Moras, D. (1993). *FASEB J.* **7**, 79–86.
Charron, C., Roy, H., Blaise, M., Giegé, R. & Kern, D. (2003). *EMBO J.* **22**, 1632–1643.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Curnow, A. W., Tumbula, D. L., Pelaschier, J. T., Min, B. & Söll, D. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 12838–12843.
Cusack, S. (1997). *Curr. Opin. Struct. Biol.* **7**, 881–889.
Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N. & Leberman, R. (1990). *Nature (London)*, **347**, 249–255.
Eriani, G., Cavarelli, J., Martin, F., Ador, L., Rees, B., Thierry, J. C., Gangloff, J. & Moras, D. (1995). *J. Mol. Evol.* **40**, 499–508.
Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990). *Nature (London)*, **347**, 203–206.
Feng, L., Yuan, J., Toogood, H., Tumbula, D. L. & Söll, D. (2005). *J. Biol. Chem.* **280**, 20638–20641.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kawarabayasi, Y. *et al.* (1999). *DNA Res.* **6**, 83–101.
Kawarabayasi, Y. *et al.* (2001). *DNA Res.* **8**, 123–140.
Kleywegt, G. J. & Jones, T. A. (1994). *CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 9–14.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murzin, A. G. (1993). *EMBO J.* **12**, 861–867.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.
Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J. C. & Moras, D. (1991). *Science*, **252**, 1682–1689.
Schmitt, E., Moulinier, L., Fujiwara, S., Imanaka, T., Thierry, J.-C. & Moras, D. (1998). *EMBO J.* **17**, 5227–5237.
Shiba, K., Motegi, H., Yoshida, M. & Noda, T. (1998). *Nucleic Acids Res.* **26**, 5045–5051.
Suzuki, K., Sato, Y., Maeda, Y., Shimizu, S., Hossain, M. T., Ubukata, S., Sekiguchi, T. & Takénaka, A. (2007). *Acta Cryst.* F**63**, 608–612.