Sugiura, I. et al. The 2.0 Å crystal structure of Thermus thermophilus methionyl-tRNA synthetase reveals two RNA-binding modules. Structure Fold Des. 8, 197-208

ARTICLE *in* STRUCTURE · MARCH 2000

Impact Factor: 5.62 · DOI: 10.1016/S0969-2126(00)00095-2 · Source: PubMed

The 2.0 Å crystal structure of *Thermus thermophilus* methionyl-tRNA synthetase reveals two RNA-binding modules

Ikuko Sugiura[1], Osamu Nureki[2], Yoshiko Ugaji-Yoshikawa[1], Sachiko Kuwabara[1], Atsushi Shimada[2], Masaru Tateno[3], Bernard Lorber[4], Richard Giegé[4], Dino Moras[5], Shigeyuki Yokoyama[2,3]* and Michiko Konno[1]*

Background: The 20 aminoacyl-tRNA synthetases are divided into two classes, I and II. The 10 class I synthetases are considered to have in common the catalytic domain structure based on the Rossmann fold, which is totally different from the class II catalytic domain structure. The class I synthetases are further divided into three subclasses, a, b and c, according to sequence homology. No conserved structural features for tRNA recognition by class I synthetases have been established.

Results: We determined the crystal structure of the class Ia methionyl-tRNA synthetase (MetRS) at 2.0 Å resolution, using MetRS from an extreme thermophile, *Thermus thermophilus* HB8. The *T. thermophilus* MetRS structure is in full agreement with the biochemical and genetic data from *Escherichia coli* MetRS. The conserved 'anticodon-binding' residues are spatially clustered on an α-helix-bundle domain. The Rossmann-fold and anticodon-binding domains are connected by a β–α–α–β–α topology ('SC fold') domain that contains the class I specific KMSKS motif.

Conclusions: The α-helix-bundle domain identified in the MetRS structure is the signature of the class Ia enzymes, as it was also identified in the class Ia structures of the isoleucyl- and arginyl-tRNA synthetases. The β–α–α–β–α topology domain, which can now be identified in all known structures of the class Ia and Ib synthetases, is likely to dock with the inner side of the L-shaped tRNA, thereby positioning the anticodon stem.

Addresses: [1]Department of Chemistry, Faculty of Science, Ochanomizu University, Otsuka, Bunkyo-Ku, Tokyo 112-8610, Japan, [2]The Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Hongo, Bunkyo-Ku, Tokyo 113-0033, Japan, [3]The Institute of Physical and Chemical Research (RIKEN), 2-1 Hirosawa, Wako, Saitama 351-0198, Japan, [4]Institut de Biologie Moléculaire et Cellulaire du CNRS, F-67084 Strasbourg, Cedex, France and [5]Institut de Genetique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP BP 163, F-67404 Illkirch Cedex, France.

*Corresponding authors.
E-mail: yokoyama@biochem.s.u-tokyo.ac.jp
konno@cc.ocha.ac.jp

Key words: methionyl-tRNA synthetase, RNA-binding molecule, Rossmann fold, *Thermus thermophilus*, tRNA

Received: **6 September 1999**
Revisions requested: **6 October 1999**
Revisions received: **1 November 1999**
Accepted: **9 November 1999**

Published: **1 February 2000**

Structure 2000, **8**:197–208

0969-2126/00/$ – see front matter

Introduction

Aminoacyl-tRNA synthetases catalyze the aminoacylation reaction, which is considered to proceed in two steps: first the amino acid and ATP form an aminoacyl-AMP, and then the aminoacyl moiety is transferred to the 3′-terminal adenosine of the tRNA. The 20 synthetases are divided into two classes, I and II (each consisting of 10 members), which must have evolved from two distinct ancestors [1]. The class I enzymes have two motifs, with consensus sequences of His–Ile–Gly–His (HIGH; in single-letter amino acid code) and Lys–Met–Ser–Lys–Ser (KMSKS) [1]. The 10 class I enzymes are further divided into three subclasses (classes Ia, Ib and Ic) on the basis of sequence homology [1–3]. Class Ia consists of the methionyl-, isoleucyl, valyl-, leucyl, cysteinyl- and arginyl-tRNA synthetases (MetRS, IleRS, ValRS, LeuRS, CysRS and ArgRS, respectively); class Ib consists of the glutaminyl- and glutamyl-tRNA synthetases (GlnRS and GluRS, respectively); and class Ic consists of the tyrosyl- and tryptophanyl-tRNA synthetases (TyrRS and TrpRS, respectively).

The crystal structures have been reported for *Bacillus stearothermophilus* TyrRS [4] and TrpRS [5], *Escherichia coli* GlnRS [6] and MetRS [7], *Thermus thermophilus* GluRS [8] and IleRS [9], *Staphylococcus aureus* IleRS [10] and *Saccharomyces cerevisiae* ArgRS [3]. The catalytic domains of the class I synthetases are characterized by a dinucleotide-binding or Rossmann fold, which consists of five parallel β strands that are connected by α helices. TyrRS and TrpRS in class Ic are dimeric, which is essential to their enzymatic activities: one cognate tRNA molecule interacts with the N-terminal region of one subunit and the C-terminal region of the other subunit in the α_2 dimer [4,5]. On the other hand, the class Ia and class Ib enzymes are monomeric [3,6,8,9], except for MetRS, which is normally in a homodimeric form. The monomeric *E. coli* and *T. thermophilus* MetRS fragments lacking the dimerization

domain are nearly fully active [11–13]. The tRNA-binding mechanisms of the class Ia/Ib (monomeric) enzymes must, therefore, be distinct from those of the class Ic (dimeric) enzymes. Only two tRNA-complexed class I enzyme structures have been reported [6,10], whereas a number of tRNA-bound structures have been determined for the class II enzymes [14–18]. In the complex of *E. coli* GlnRS (class Ib) and $tRNA^{Gln}$, the CCA and anticodon arms of $tRNA^{Gln}$ interact with the N-terminal Rossmann-fold domain and the C-terminal β-barrel domains, respectively [6]. The other class 1 complex is the recently reported structure of a complex of *S. aureus* IleRS and $tRNA^{Ile}$, which is not a normal complex but an 'editing complex' with the CCA terminus of the tRNA to be bound with the editing active site 34 Å away from the aminoacylation active site of IleRS [10]. As no other tRNA–enzyme complex structure is available, it remains to be elucidated which RNA-binding mechanisms are conserved for the other class Ia/Ib enzymes. For example, the anticodon-binding domains of GlnRS are completely replaced by other domains in the GluRS structure [8] of the same subclass (class Ib), suggesting that the class I enzymes have a variety of anticodon-recognition mechanisms.

Although a crystal structure of *E. coli* MetRS in a monomeric form was reported on the basis of diffraction data at 2.5 Å resolution [7], the polypeptide-chain tracing cannot adequately explain the large amount of data accumulated by biochemical and mutagenesis studies of tRNA recognition and zinc binding by the *E. coli* MetRS [19–24]. In this study, we determined the crystal structure of *T. thermophilus* MetRS in the monomeric form (M_r 58,128) at 2.0 Å resolution, which is consistent with the tRNA recognition and zinc-binding data of the *E. coli* and *T. thermophilus* MetRS enzymes [19–26]. This allowed us to make a systematic comparison between the structures of all three subclasses. Thus, we can now unambiguously define two RNA-binding modules: one RNA-binding domain is conserved throughout the class Ia/Ib enzymes, and the other domain is conserved in the class Ia enzymes.

Results and discussion

Structure determination

First, we overproduced and purified the dimeric, full-length *T. thermophilus* MetRS (616 amino acid residues) as described [27], and tried to crystallize it. Two months after setting up the hanging drops, crystals appeared, which were resolved and evaluated using sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS–PAGE). Thus, the molecular mass of the crystallized protein was estimated to be approximately 58 kDa, which is quite consistent with the value of the T1–T2 fragment obtained by limited digestion with trypsin [11]. The T1–T2 tryptic fragment was found to be monomeric and nearly as active as the full-length protein in the aminoacylation reaction [11]. We therefore constructed an overexpression system of a fragment (amino acid residues 1–502, M_r 58,128) corresponding to the T1–T2 tryptic fragment. The recombinant monomeric form, lacking the dimerization domain, of *T. thermophilus* MetRS was overexpressed in *E. coli* strain JM109(DE3), and was purified by heat treatment (at 70°C for 30 minutes) and sequential chromatography on DEAE-Sephacel and Phenyl Superose fast protein liquid chromatography (FPLC) columns.

The structure was determined using multiple isomorphous replacement, and the selenomethionylated protein contributed mainly to the phase determination. The atomic model, including residues 1–500 and 136 water molecules, was refined for 31,858 reflections at 6.0–2.0 Å resolution. The last two residues at the C-terminal end could not be observed in the electron-density map. The final crystallographic R factor was 21.0%, and the R_{free} was 29.1% (Table 1). The structure of the monomeric *T. thermophilus* MetRS is an elongated ellipsoid, with dimensions of $50 \times 35 \times 90$ Å (Figure 1a). The structure is inconsistent with the previous chain-tracing of the crystal structure of *E. coli* MetRS [7], except for approximately 120 amino acid residues from the N terminus to the end of the β4 strand (Figures 1a,b). In contrast, a new electron-density map of *E. coli* MetRS has successfully been chain-traced using the present *T. thermophilus* MetRS structure as a guide, indicating that the *E. coli* and *T. thermophilus* MetRS structures are in fact very similar to each other [28].

The Rossmann-fold domain

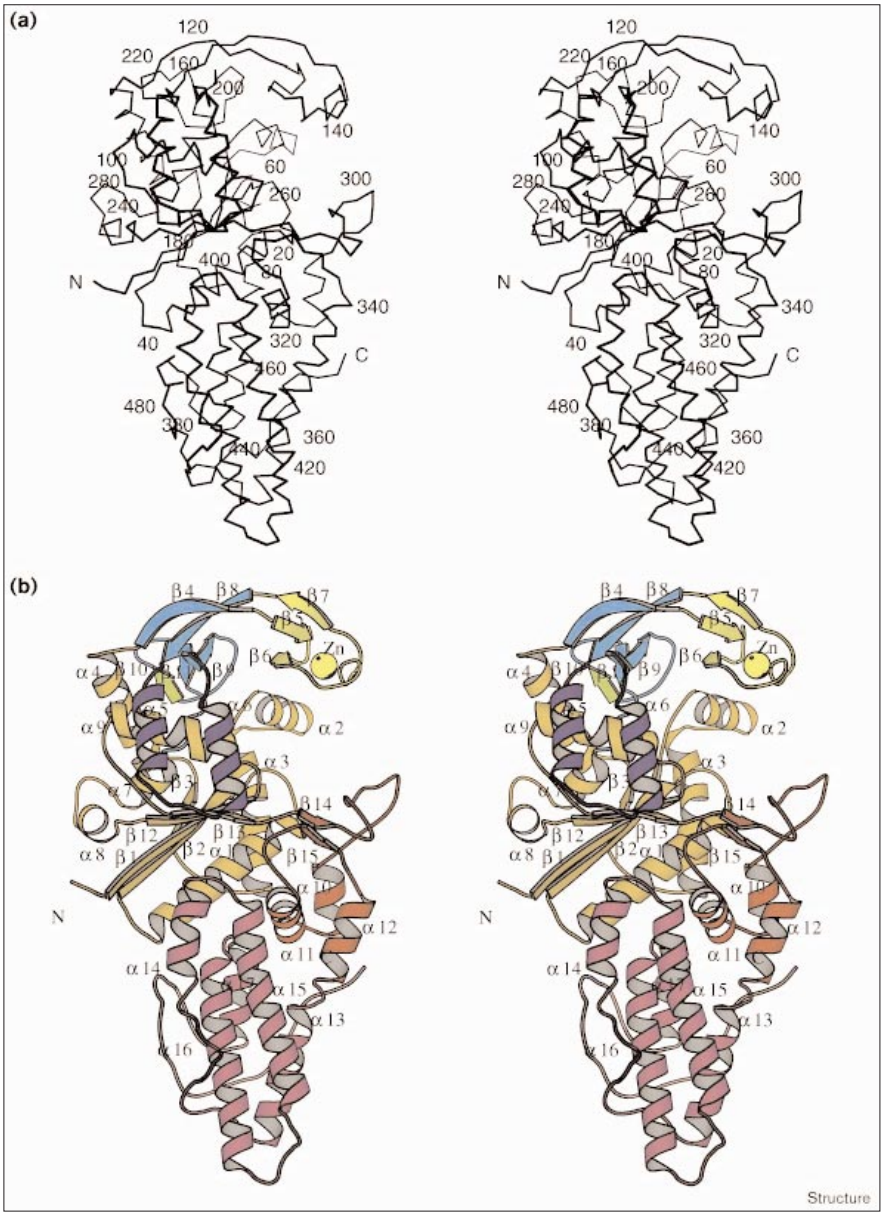
The Rossmann-fold domain of *T. thermophilus* MetRS is formed by two polypeptide segments (residues 1–113 and 232–287; colored in orange in Figures 1b and 2) separated by an insertion (residues 114–231; described later in detail). The C terminus of the MetRS Rossmann-fold domain was determined to be Gly287, which might require the corresponding redefinition of the domain boundaries for the reported class Ia/Ib enzyme structures [3,6,8,9], as follows. In the present MetRS structure, Gly287 is the last residue of the fifth β strand (β13) and forms a hydrogen bond between its backbone N–H group and the C=O group of Ile257 in the fourth strand (β12) in the Rossmann-fold β sheet. The canonical β-strand conformation of β13 ends at Gly287, as the C=O group of Gly287

Table 1

Refinement statistics.

Resolution range (Å)	6.0–2.0
Number of reflections in refinement ($I > 2\sigma$)	31858
Number of protein atoms	4098
Number of water molecules	136
R factor (%)	21.0
R_{free} (%)	29.1
Rms deviation from ideal geometry	
Bond lengths (Å)	0.006
Bond angles (°)	1.3

Figure 1

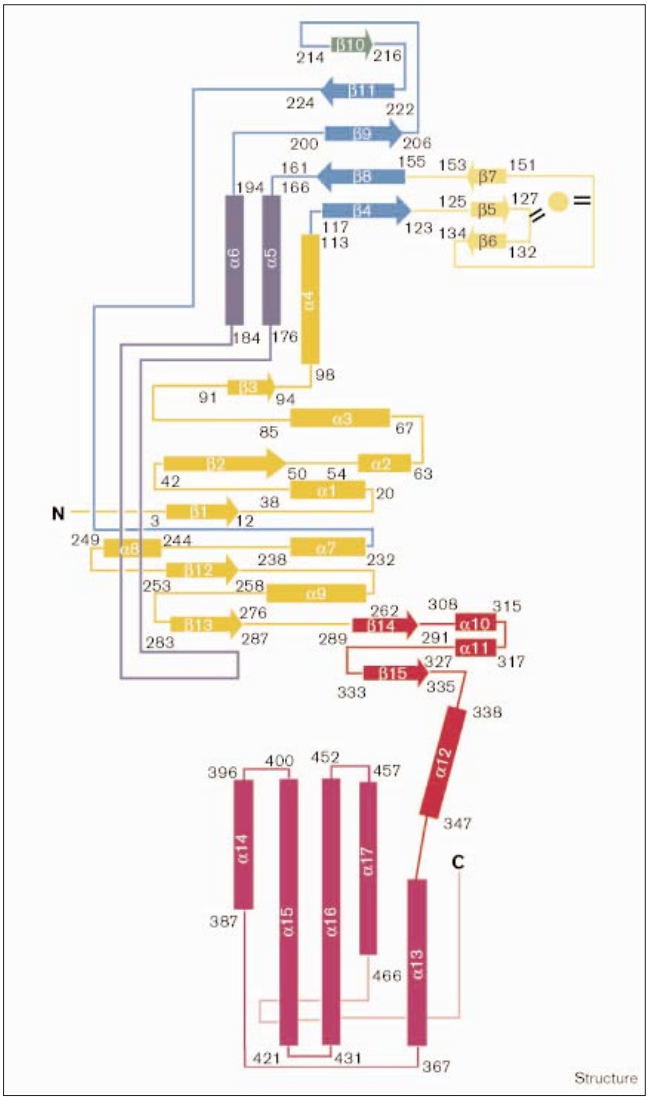


The overall structure of *T. thermophilus* MetRS. **(a)** Stereoview of the Cα trace of the *T. thermophilus* MetRS structure with every 20th residue labeled. **(b)** Stereoview of a schematic drawing of the structure of *T. thermophilus* MetRS. The Rossmann-fold domain (shown in orange) characterizes the class I aminoacyl-tRNA synthetases. The connective polypeptide (CP) domain consists of the antiparallel β sheet core (cyan), a MetRS-specific β-strand insertion (green), the β ribbon with the zinc finger (yellow), and the antiparallel α_2 helical structure (blue). The stem contact (SC) fold domain (red) is identified here as an RNA-binding domain. The α-helix-bundle domain (magenta) is the putative anticodon-binding domain. The C-terminal residues 467–500 (pink) in the present structure adopt a random coil, and are to be followed by a dimerization domain in the full-length protein. The secondary structure elements are numerically labeled starting from the N terminus, as the β strands (β1–β15) and the α helices (α1–α17). The bound Zn^{2+} ion is represented as a yellow ball.

forms a hydrogen bond with the backbone N–H group of Lys259 in the lysine–aspartic-acid–isoleucine loop connecting β12 to α9. Then, the Rossmann-fold domain is connected to another domain through Gly288, which is not involved in any hydrogen-bonding interaction. In the previous reports on the class Ia/Ib enzyme structures [3,6,8,9], the fifth β strand of the Rossmann fold was considered to continue over the end of the hydrogen-bonding interaction with the fourth β strand. In all of these structures, however, we noticed that the canonical β-strand conformation of the fifth β strand of the Rossmann fold is terminated at the same position as in *T. thermophilus* MetRS. The 'last' Rossmann-fold residue forms a hydrogen bond with the fourth β strand, and is usually followed by a glycine or a proline residue [3,6,8,9]. This glycine/proline linker is followed by another β strand, which had been interpreted to be an extension of the fifth β strand of the Rossmann fold, but is now proposed to be the first β strand of the next domain (the 'SC-fold' domain, see later).

As a result of systematic comparisons among all three subclasses (classes Ia, Ib and Ic) we propose that the α helix just following the N-terminal half of the Rossmann fold (α4 in the present MetRS structure) can be included in the core of the Rossmann-fold domain because the α helix is common to all seven synthetase structures. In contrast

Figure 2



Topology diagram of *T. thermophilus* MetRS; the α helices are represented by rectangles and the β strands by arrows. The numbers of the secondary structure elements and the color definitions for the domains are the same as those in Figure 1b. The bound Zn^{2+} ion and its ligands are represented as a yellow circle and black bars, respectively.

to the '$\beta_3\alpha_3$ plus $\beta_2\alpha_2$' topology of the Rossmann-fold domains in classes Ib and Ic [4,5,6,8], an α helix (α2) is inserted between β2 and α3 in the N-terminal half and α8 is inserted between α7 and β12 in the C-terminal half of the Rossmann fold (Figures 1b and 2). Similar insertions, consisting of three and one α helices, respectively, occur in the structures of *T. thermophilus* IleRS [9] and *S. cerevisiae* ArgRS [3] (Figure 3). We indicate, therefore, that these insertions are class Ia specific, although the strict numbers of α helices are not the same.

The twist in the five-stranded parallel β sheet creates a deep catalytic cleft. The HIGH motif of *T. thermophilus* MetRS, the actual sequence of which is His19–Leu20–Gly21–His22, is located in a loop between β1 and α1, with the two histidine residues stacked upon each other (Figure 3). This feature is conserved among all three class Ia synthetases (MetRS, IleRS and ArgRS), the two class Ib synthetases (GlnRS and GluRS), and the two class Ic synthetases (TyrRS and TrpRS) [4–6,8,9] (Figure 3). On the basis of the present domain-boundary assignments, the other catalytic motif, KMSKS, of the class I enzyme is not located in the Rossmann-fold domain, but is in the next domain (the 'SC-fold' domain; see later).

The 'connective polypeptide' domain

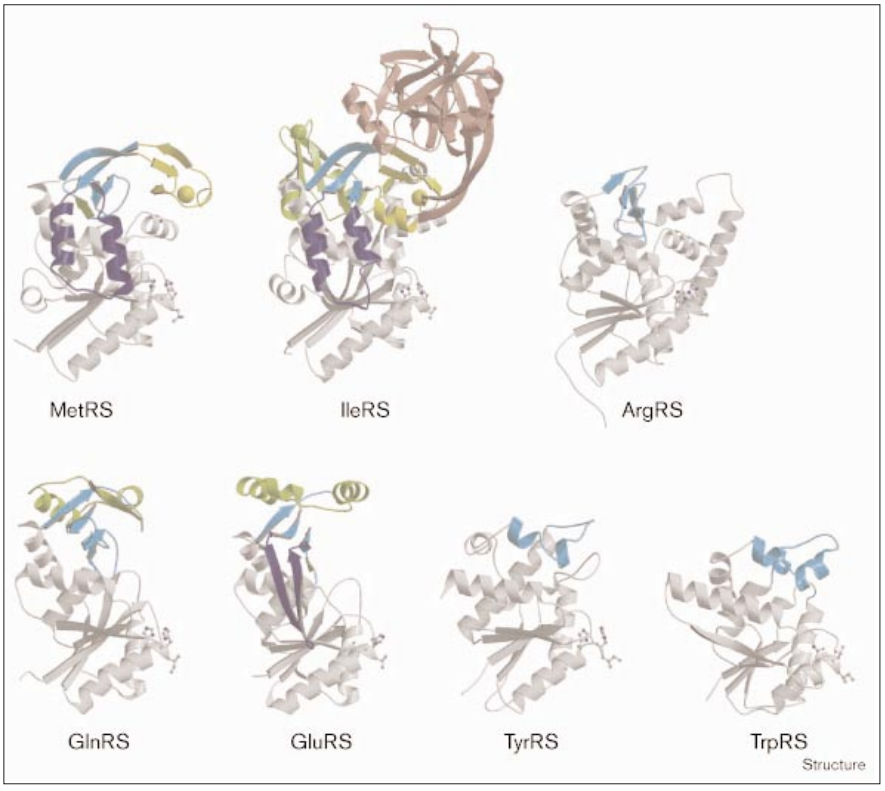
The present *T. thermophilus* MetRS structure has a domain (one continuous polypeptide segment consisting of residues 114–231) inserted between the N- and C-terminal halves of the Rossmann fold. Hereafter we denote this inserted domain as the 'connective polypeptide' domain (or CP domain). In the previous incorrect structure of *E. coli* MetRS [7], this polypeptide was divided into two segments, CP1 and CP2, by the presence of a secondary structure element of the Rossmann fold in the middle. The present structure, however, unambiguously shows that the inserted segment is not divided in that manner, but is folded continuously and independently of the Rossmann fold.

The CP domain of *T. thermophilus* MetRS consists mainly of an antiparallel β sheet (β4–β11, α5 and α6). The other class I enzyme structures also have an insertion (the CP domain) between the N- and C-terminal halves of the Rossmann fold. Based on a full comparison of these CP-domain structures, we define here the 'CP-domain core' as a four-stranded antiparallel β-sheet structure, which is conserved in all of the class Ia enzymes (MetRS, IleRS and ArgRS) and the class Ib enzymes (GlnRS and GluRS) (Figure 3). The CP-domain core of *T. thermophilus* MetRS consists of β4, β8, β9 and β11 (colored in cyan in Figures 1b and 2). The central, third strand (β9) of the core antiparallel β sheet significantly kinks in the middle, which results in hydrophobic interactions inside. The yeast ArgRS exhibits the simplest CP domain, which has only the four-stranded CP-domain core. The corresponding domain of the *T. thermophilus* arginyl-tRNA synthetase also consists of four strands (AS *et al.*, unpublished observations). In contrast, other CP domains have rather large insertions between the core β strands, as described later (Figure 3). On the other hand, the CP-domain structures of the class Ic enzymes (TyrRS and TrpRS), consisting of counterclockwise-arranged α helices that form a dimer interface [4,5], are completely different from the class Ia/Ib CP-domain structures (Figure 3).

Between the first and second β strands (β4 and β8) of the CP-domain core, an antiparallel β sheet (β5, β6, and β7, colored yellow in Figures 1b and 2) is inserted. At the tip of this β-sheet insertion, one Zn^{2+} ion is coordinated by an

Figure 3

Schematic representations of the structures of the CP domains intervening into the Rossmann folds of the class I synthetases (MetRS, IleRS, ArgRS, GlnRS, GluRS, TyrRS and TrpRS) viewed from the same direction as in Figure 1b. The class I specific Rossmann fold domains are also shown in gray. The class Ia/Ib specific antiparallel four-stranded β sheet ('the CP-domain core') is shown in cyan, and the inserts between the first and second β strands, between the second and third β strands, and between the third and fourth β strands of the antiparallel β sheet are shown in yellow, blue and green, respectively. In TyrRS and TrpRS the cyan antiparrallel β sheet is replaced by two short helices. A globular β-barrel structure intervening between the zinc finger in IleRS is shown in pink.



appropriately oriented nitrogen and three sulfur atoms in the zinc finger motif, Cys127–X_2–Cys130–X_{13}–Cys144–X_2–His147 (where X is any amino acid). This zinc finger structure is essential for the aminoacylation activity of *T. thermophilus* MetRS [26]. The peptide conformation of this zinc-binding site of *T. thermophilus* MetRS is nearly the same as the solution structure of the zinc-binding fragment of *E. coli* MetRS [25], although the positions of these zinc-binding sites in the primary structures are not the same in the two enzymes [28]. This antiparallel β-sheet insertion with a zinc-binding site is well conserved in the IleRS structure of the same subclass, class Ia [9]. As shown in Figure 3, IleRS has a zinc finger, Cys181–X_2–Cys184–X_{204}–Cys389–X_2–Cys392, at the tip of the β sheet insertion (yellow), which is essential for aminoacylation [29–31]. In IleRS, however, a 189-residue insert, folded into a globular β barrel structure (shown in pink in Figure 3), intervenes between the second and third cysteine residues of the zinc finger. This β-barrel domain has been shown to function in the hydrolytic editing of Val–AMP and Val–$tRNA^{Ile}$ [9,32]. In contrast, MetRS has no such editing activity. Furthermore, the class Ib structures of GlnRS and GluRS have polypeptide insertions with α–β–α and α–α–α topologies, respectively, between the first and second β strands of the CP-domain core. For this α–α–α segment in a structural model of *E. coli* GluRS, a zinc-finger-like structure is formed around the first α helix, and actually coordinates a functionally important Zn^{2+} ion [33]. Intriguingly, this Zn^{2+} ion in the *E. coli* GluRS structural model is located at a position quite similar to those of the Zn^{2+} ions in the *T. thermophilus* MetRS and IleRS structures.

Between the second and third β strands (β8 and β9), two antiparallel α helices (α5 and α6), which run perpendicular to the β sheet of the Rossmann fold, are inserted (shown in blue in Figures 1b, 2 and 3). The loop connecting α5 and α6 forms two hydrogen bonds with β13 at the edge of the parallel β sheet of the Rossmann fold. Moreover, helices α5 and α6 are fixed by the hydrophobic interaction with the surface of the Rossmann fold. In the corresponding site between the second and third β strands of the CP-domain core, IleRS has an antiparallel pair of α helices, which is fixed hydrophobically on the Rossmann-fold domain, with the β–β loop anchored to the fifth β strand of the Rossmann fold by three hydrogen bonds (Figure 3). In contrast, GluRS has an antiparallel pair of β strands. Intriguingly, this β hairpin of GluRS is fixed hydrophobically on the Rossmann-fold domain, and the β–β loop is anchored to the fifth β strand of the Rossmann fold, in a similar manner to those in the MetRS and IleRS structures (Figure 3). On the other hand, GlnRS has a 'β loop' (Figure 3), which is involved in disrupting the end of the tRNA acceptor stem [6]

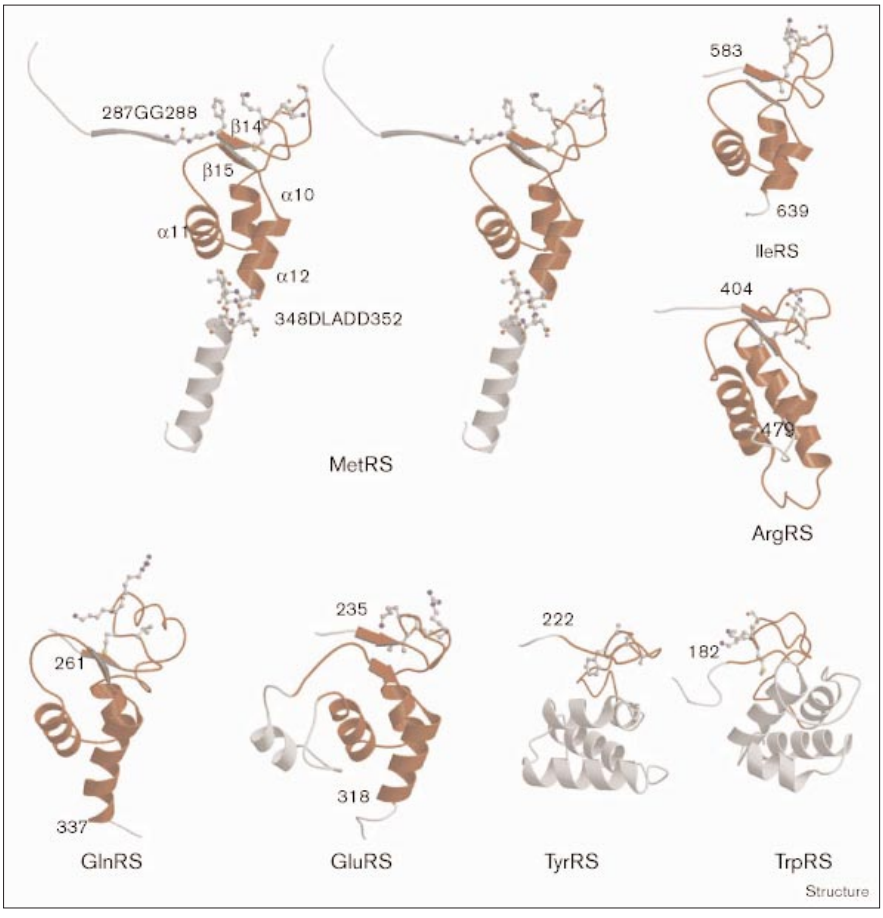
between the second and third strands of the CP-domain antiparallel β sheet.

Between the third and fourth β strands (β9 and β11) of the CP-domain core, a β strand (β10; shown in green in Figures 1b, 2 and 3) is inserted and forms hydrogen bonds with β11 in an antiparallel manner. This β-strand insertion is considered to be a MetRS-specific element. Between the corresponding β strands, IleRS has a much longer insertion with a β–β–β–α–β–β topology: a couple of antiparallel pairs of β strands face each other, and at their β-turn ends two pairs of cysteine ligands are coordinated to the second Zn^{2+} ion (Figure 3).

The 'stem contact fold' domain

As shown in Figures 1b and 2, the region that connects the N-terminal Rossmann-fold domain (shown in orange) and the C-terminal α-helix-bundle domain (shown in magenta) has a β–α–α–β–α topology (shown in red), and is tentatively designated as the β–α–α–β–α domain. As described above, the N-terminal end of the β–α–α–β–α domain (Phe289) is linked by Gly288 to the C-terminal end of the Rossmann-fold domain (Figure 4). The last α helix (α12) of the β–α–α–β–α domain is nearly coaxial to the first α helix (α13) of the α-helix-bundle domain (Figure 4). These two α helices, however, are clearly separated by the Asp348–Leu349–Ala350–Asp351– Asp352 segment (Figure 4) in an unusual conformation: the mainchain C=O group of Ala347 forms a hydrogen bond with the mainchain N–H group of Asp352(= 347 + 5), and the mainchain C=O group of Asp352 forms a hydrogen bond with the sidechain NH_2 group of Asn415 in α15. The C-terminal end of the β–α–α–β–α domain is Ala347. The secondary structure elements of the β–α–α–β–α domain tightly interact within the globule. The two β strands (β14 and β15) of the β–α–α–β–α domain form a short parallel β sheet. The three α helices (α10, α11 and α12) are perpendicularly arranged relative to each other. These secondary structure elements are tightly packed through a hydrophobic core made up of eleven residues including six leucine and three tyrosine residues. On the other hand, it is primarily α11 that contacts the N-terminal Rossmann-fold domain and the C-terminal α-helix-bundle domain (Figure 1). The Rossmann-fold domain interacts with four sidechains of the β–α–α–β–α domain and, to a similar extent, with five sidechains of the α-helix-bundle domain.

Figure 4



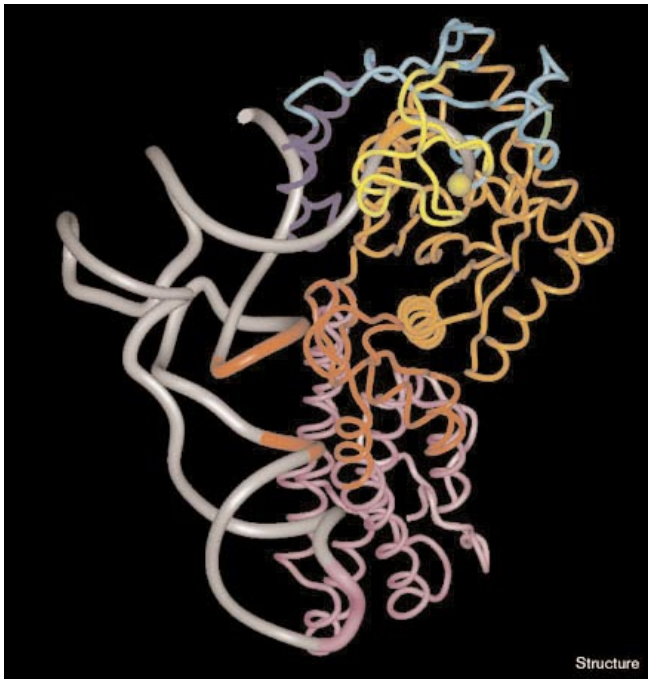
Schematic representations of the structures of the β–α–α–β–α domain of *T. thermophilus* MetRS and the corresponding regions of IleRS and ArgRS (class Ia), GlnRS and GluRS (class Ib), and TyrRS and TrpRS (class Ic), indicated in red (viewed from the same direction as in Figure 1b). For the MetRS structure (stereoview), in addition to the β–α–α–β–α domain (ribbon diagram), the linker regions connecting this domain to the N-terminal Rossmann-fold domain and the C-terminal α-helix-bundle domain are shown in ball-and-stick representation. The fold of this β–α–α–β–α domain is named the stem contact (SC) fold. The SC-fold domain is conserved very well in other class Ia/Ib synthetases, IleRS, ArgRS, GlnRS and GluRS (the first and the last residues in the domain are numbered). In the long loop between the first β strand and the first α helix of the SC fold, there is the 'KMSKS' motif, shown in ball-and-stick representation. In contrast, the class Ic synthetases (TyrRS and TrpRS) do not have the SC fold; the long loop with the KMSKS motif does not form a particular fold, and is followed by the C-terminal domain with a totally different topology (α–α–α–α), which might be involved in the anticodon binding. In the *B. stearothermophilus* TyrRS structure the sidechains of Lys230, Lys233 and Thr234, of the KMSKS motif, and of Lys230–Phe231–Gly232–Lys233–Thr234 are missing.

On the basis of the characteristic boundary and hydrophobic-core structures of the β–α–α–β–α domain, we found that a homologous topology is well conserved in the class Ia and class Ib synthetases (IleRS, ArgRS, GlnRS and GluRS; Figure 4) [3,6,8,9], although these polypeptide segments had never been described as structural domains. The glycine/proline linker structure at the boundary between the Rossmann-fold and β–α–α–β–α domains is well conserved, as described above. The last α helix of the β–α–α–β–α domain is followed by a short linker segment that clearly separates this domain from the anticodon-binding domain, whereas the anticodon-binding domain structures are totally different among MetRS/IleRS/ ArgRS, GlnRS and GluRS, as described below. The hydrophobic interactions among the secondary structure elements within the β–α–α–β–α domain are more extensive than the contacts of this domain with the N- and C-terminal domains. Our sequence alignment suggests that the hydrophobic-core residues of the class Ia MetRS/IleRS/ β–α–α–β–α domain are well conserved in the three remaining class Ia synthetases, ValRS, LeuRS and CysRS (not shown). On the other hand, in the class Ic synthetase structures, the Rossmann fold is followed by a long loop and a domain with an α–α–α–α topology, which is different from those of the β–α–α–β–α domains of the class Ia/Ib synthetases [4,5] (Figure 4). The β–α–α–β–α domain is characteristic of the monomeric class Ia/Ib synthetases. On the other hand, the α–α–α–α domain might be suggested to be the anticodon-binding domain of the class Ic, dimeric TyrRS/TrpRS.

The crystal structure of the *E. coli* GlnRS–$tRNA^{Gln}$ complex reveals that the β–α–α–β–α domain (a part of it is named as the connective subdomain in [34,35]) interacts with the acceptor, D (dihydrouridine), and anticodon stems of the tRNA [6]. The Glu323 residue in the loop, corresponding to that between β15 and α12 of MetRS, contacts the highly conserved G10–C25 base pair of the tRNA [6]. The β–α–α–β–α topology of this domain is, therefore, denoted hereafter as the 'stem contact' (SC) fold. On the basis of the GlnRS–$tRNA^{Gln}$ structure, a synthetase–tRNA docking model can be made, in which the SC-fold domain broadly contacts the D stem and the terminal region of the anticodon stem (the inner side of the L-shaped tRNA structure) and the C-terminal domain recognizes the anticodon of the tRNA (Figure 5). The SC-fold domain is connected with the Rossmann-fold and anticodon-binding α-helix-bundle domains, through two short hinges (Figures 1 and 4). Thus, the SC-fold domain might have a common function of docking the inner side of the L-shaped tRNA so as to present precisely the anticodon trinucleotides to the proper recognition site of the enzyme. Because most of the identity determinants of the tRNA have been identified outside of the inner side of the L shape, the putative contact of the SC-fold domain with the tRNA might depend on the shape of the RNA stem, but not so much on the nucleotide sequence in the contact region. A number of water molecules might be released when the SC-fold domain fits into the inner-side surface of the L-shaped tRNA. On the other hand, we now point out that the class IIb enzyme structures (AspRS, AsnRS and LysRS) have a subdomain with a well-conserved fold, the topology of which is different from that of the class I SC fold, for the contact with the inner side of the L-shaped tRNA. It is intriguing that the two different folds serve as a domain connecting the catalytic and anticodon-binding domains and contacting the inner side of the tRNA.

The KMSKS motif, Lys297–Met298–Ser299–Lys300–Thr301, is situated around a β turn in the loop between β14 and α10 in the SC-fold domain of *T. thermophilus* MetRS (Figure 4). The hydrophobic part of the Lys297 sidechain interacts with the phenyl ring of Phe289 (Figure 4). As the KMSKS motif is considered to contribute to the catalytic site, the SC-fold domain carrying this motif might link the

Figure 5



A docking model of *T. thermophilus* MetRS and $tRNA^{Met}$. The color definitions for MetRS are the same as those in Figure1b, and the docked $tRNA^{Met}$ molecule is colored in gray. On $tRNA^{Met}$, the discriminator A73, the enzyme-contact nucleotides on the inner side of the L-shaped tRNA, and the anticodon trinucleotides are colored in orange, red, and magenta, respectively. Correspondingly, on MetRS, the Rossmann-fold, SC-fold, and anticodon-binding domains are in orange, red and magenta, respectively. The bound Zn^{2+} ion is represented as a yellow ball. This model was constructed by superposition of the Rossmann-fold domain of the *T. thermophilus* MetRS structure onto that of the *E. coli* GlnRS–$tRNA^{Gln}$ complex structure [6], followed by removal of the GlnRS structure. This figure was prepared using the program Insight II [53].

tRNA binding to the catalytic-site activity. In this context, the β15–α12 of the MetRS SC-fold domain corresponds to the β–α segment that was proposed (as a result of genetic studies) to transmit the signal of the anticodon recognition to the catalytic site of *E. coli* GlnRS [35], whereas the polypeptide chain corresponding to the SC-fold domain was not regarded as a structural domain [6]. The conformation of the KMSKS loop region is well conserved in the structures of IleRS (class Ia) and GluRS (class Ib) in the substrate-free form [8,9] (Figure 4). On the other hand, the substrate-bound forms of the class Ia/Ib enzymes, such as $tRNA^{Gln}$–GlnRS–ATP [6] and ArgRS–arginine [3], exhibited KMSKS loop conformations different from those of MetRS, IleRS and GluRS (Figure 4). Large conformational changes of the KMSKS loop region should, therefore, occur upon binding with substrate(s). Moreover, in the present *T. thermophilus* MetRS structure, Tyr12 would cause a steric hindrance against ATP or Met–AMP in the putative catalytic site corresponding to that of $tRNA^{Gln}$–GlnRS–ATP. The conformational change of the KMSKS loop region is, therefore, likely to be induced upon substrate binding, which might also affect the properties of the region around the other catalytic motif, HIGH, of *T. thermophilus* MetRS. Such dynamic movement of the KMSKS loop might be related to the dynamic nature of the aminoacylation reaction, where the same catalytic site is used successively during the putative two-step mechanism, including aminoacyl–AMP formation in the absence of tRNA and the following transfer of the aminoacyl moiety of aminoacyl–AMP to the 3′-end of tRNA. Soaking native *T. thermophilus* MetRS crystals in a solution containing ATP (or methionine) and co-crystallization from the solution of the protein and ATP (or methionine) were tried. The electron-density maps, however, showed no evidence of an ATP (or methionine) molecule. On the other hand, the dimeric class Ic enzymes have the KMSKS motif in a long loop just following the Rossmann-fold domain (Figure 4). The structures of TyrRS bound with Tyr or Tyr–AMP and TrpRS bound with Trp–AMP exhibited different KMSKS loop conformations, not only from the class Ia/Ib KMSKS loop conformations, but also from each other (Figure 4).

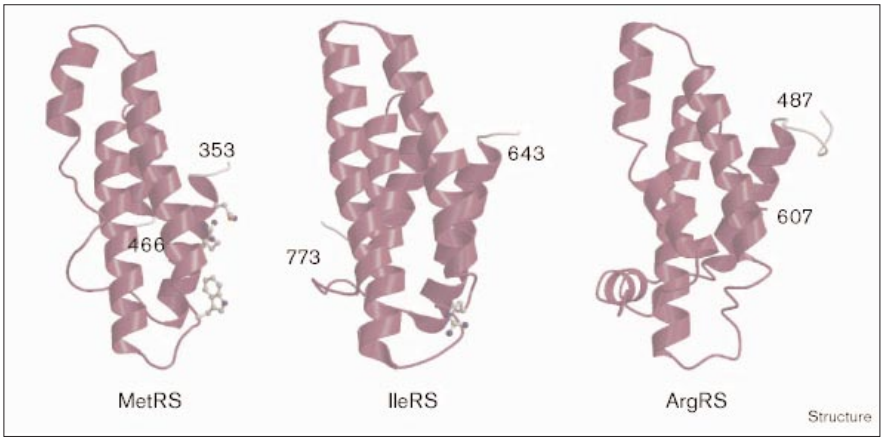
For proteins other than the class Ia/Ib synthetases, we found no structure homologous to the SC-fold domain by the program Dali version 2.0 [36]. It is not, as yet, clear if any domain of the SC fold exists independently of the Rossmann fold, or if the SC fold is a C-terminal extension of the Rossmann fold. Nevertheless, we would propose that the SC-fold domain found in the class Ia/Ib synthetase structures is to be regarded as a structural domain because of the following reasons: first, the SC-fold domain has a clearly defined hydrophobic core; second, as described previously for *T. thermophilus* MetRS, the interaction of the SC-fold domain with the Rossmann-fold domain is as moderate as that between the Rossmann-fold and anticodon-binding domains; third, the position of the SC-fold domain relative to the Rossmann-fold domain is different among the class Ia/Ib structures (data not shown), whereas the topology of the SC fold is well conserved. On the other hand, it should be noted here that the two class Ic synthetases lack the SC-fold domain and have the KMSKS motif in the loop following the Rossmann-fold domain. The SC-fold domain, therefore, might have originated from such a simple KMSKS loop and have evolved to the present complex form of the SC fold downstream of the Rossmann-fold catalytic domain. Otherwise, it is still possible that the ancient synthetase already had the SC-fold domain connecting the Rossmann-fold and anticodon-binding domains, and that the class Ic synthetases have lost the SC fold when they evolved to be dimeric. Anyway, the presence of the SC fold between the Rossmann-fold and anticodon-binding domains characterizes the class Ia/Ib synthetases.

The anticodon-binding α-helix-bundle domain

The α-helix-rich cylindrical domain of *T. thermophilus* MetRS (residues 353–466, shown in magenta in Figures 1b and 2) is a bundle of five antiparallel α helices (α13–α17). The N-terminal end of this α-helix-bundle domain is connected to the SC-fold domain through the characteristic linker structure (Figure 4). Thus, the three class Ia synthetases, MetRS, IleRS and ArgRS, have in common the α-helix-bundle domain of the same topology (Figure 6). In the cases of MetRS and IleRS, the first α helix of the α-helix-bundle domain is nearly co-axial to the last α helix of the SC-fold domain. Previous studies overlooked the characteristic turn structure between these two α helices, and therefore regarded them as one continuous α helix. In contrast, the two α helices of ArgRS are separated by a linker structure different from those of MetRS and IleRS, and are less co-axial [3]. The bundle-forming helices of *T. thermophilus* MetRS make rigid contacts through hydrophobic interactions between the sidechains of the leucine and valine residues that are clustered inside the α-helix-bundle (data not shown). The α13 and α15 helices might constitute a recognition surface for the tRNA anticodon; the invariant Asn355 and Arg359 residues protruding from α13, and Trp424 in the loop connecting α15 and α16 come in close proximity (Figure 6). For *E. coli* MetRS, the corresponding Asn391, Arg395 and Trp461 residues have been shown using mutational analyses to recognize the three nucleotides of the $tRNA^{Met}$ anticodon [19–22]. In contrast, the previously reported structure of *E. coli* MetRS [7] did not demonstrate the clustering of these functional residues. Now, we clearly show that the amino acid residues involved in the anticodon recognition are clustered on one face of the α-helix-bundle domain in the *T. thermophilus* MetRS structure (Figure 7). The lower half of this anticodon-recognition surface is particularly hydrophobic, possibly for interacting with the bases, whereas the upper half is hydrophilic (Figure 7). In the *T. thermophilus* IleRS structure, the Arg726 residue, which corresponds to the

Figure 6

Schematic representations of the 'anticodon-binding' α-helix-bundle domains of the class Ia synthetases (MetRS, IleRS and ArgRS) viewed from the same direction as in Figure 1b. The first and the last residues in the domain are numbered. The amino acid residues that are important for the anticodon recognition by MetRS and IleRS are shown in ball-and-stick representation. Figures 3, 4 and 6 were prepared using the program Molscript [54] and Raster3D [55].



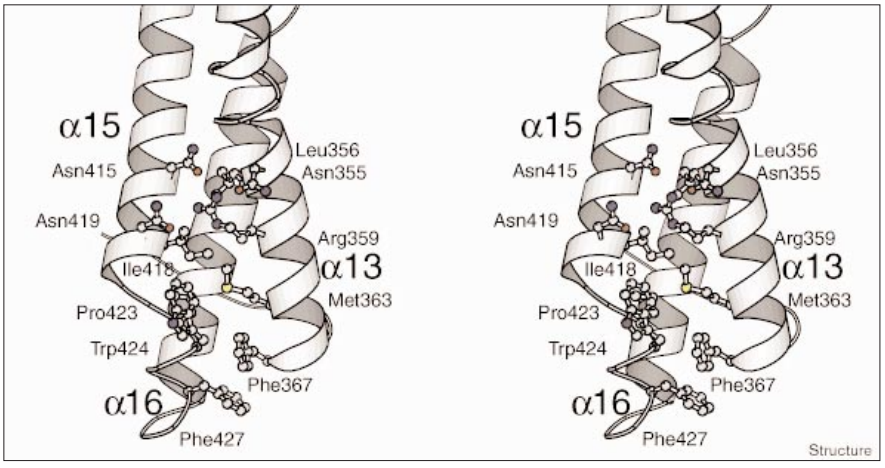
E. coli IleRS Arg734 residue involved in the recognition of the $tRNA^{Ile}$ anticodon [37,38], is located in the same position as that of the anticodon recognizing the Trp424 residue of the *T. thermophilus* MetRS (Figure 6). The previous sequence alignment suggested that all of the class Ia synthetases have an α-helical domain for therefore anticodon recognition [37], and this domain seems to be specific to class Ia synthetases.

On the other hand, the two class Ib synthetases have different architectures in their anticodon-binding domains, which might be ascribed to their particular evolutionary scenario [8]: GlnRS has two β-barrel domains [6,39], whereas GluRS exhibits two α-helical domains [8]. A five-stranded β barrel structure, termed the oligonucleotide binding (OB) fold, is responsible for the anticodon recognition by class IIb enzymes, such as AspRS [14], AsnRS [40] and LysRS [16]. In the cases of HisRS and GlyRS, an α/β fold, comprising a five-stranded mixed β sheet surrounded by three α helices, serves as the anticodon-recognition domain [41,42]. The RNA-binding domain of the U1 small nuclear ribonucleoprotein (snRNP) A folds into a four-stranded β sheet flanked on one side by two α helices, and recognizes the single-stranded loop region of a stem-loop RNA structure [43,44]. As compared with these RNA-binding modules, the anticodon-recognition α-helix-bundle domain of the class Ia aminoacyl-tRNA synthetase is unique.

In summary, the class I aminoacyl-tRNA synthetases have in common the Rossmann-fold domain (indicated in orange in Figures 1b and 2), and the class Ia and Ib synthetases have in common the CP-domain core (cyan) and the SC-fold domain (red). The CP-domain core consists of a four-stranded antiparallel β sheet, and is inserted between the N- and C-terminal halves of the Rossmann-fold domain. Distinct subdomains (shown in yellow, blue and green in the figures) are inserted into the three β–β loops of the

Figure 7

Stereoview of the 'anticodon-binding' α-helix-bundle domain of *T. thermophilus* MetRS. The putative anticodon-binding surface is illustrated with the ball-and-stick sidechains of the exposed amino acid residues (viewed from a direction rotated by 90° around the vertical axis). Figures 1 and 7 were prepared using the program Molscript [54].



CP-domain core. The SC-fold domain is located just after the Rossmann-fold domain and is followed by the anticodon-binding domain. The α-helix-bundle anticodon-binding domain (magenta) characterizes the class Ia synthetases, whereas the class Ib synthetases, GlnRS and GluRS have distinct anticodon-binding domains.

Biological implications

Each of the 20 aminoacyl-tRNA synthetases strictly recognizes its specific amino acid and tRNA(s). The 10 'class I' aminoacyl-tRNA synthetases are thought to have evolved from an ancestral enzyme with the catalytic-domain structure based on the Rossmann fold, which is totally different from that conserved in the 10 'class II' aminoacyl-tRNA synthetases. Nevertheless, it had not been well established what is conserved and what is diverged in the protein structures of the class I synthetases, with respect to the recognition of their characteristic nucleotide residues located, for example, in the anticodon and the amino-acid-accepting terminus. In order to understand the evolutionary relationships among the class I synthetases, the sequence-based classification into three subclasses, Ia, Ib, and Ic, must be examined at the tertiary-structure level.

In the present study, we determined the crystal structure of the 'class Ia' methionyl-tRNA synthetase (MetRS) from *Thermus thermophilus*. The *T. thermophilus* MetRS structure does not agree with the previously reported, probably incorrect, crystal structure of the *Escherichia coli* MetRS, and is in full agreement with the biochemical and genetic data from *E. coli* MetRS. The conserved 'anticodon-binding' residues are spatially clustered on an α-helix-bundle domain. We show that the topology of the MetRS α-helix-bundle domain is conserved in the structures of the class Ia isoleucyl- and arginyl-tRNA synthetases. The new domain boundaries of the Rossmann fold domain and the connective peptide (CP) domain characteristically inserted into the Rossmann-fold domain are proposed in this study for the class I synthetases. The region between the Rossmann-fold and the anticodon-binding domains was defined as a domain with a β–α–α–β–α topology (named the SC fold), which was found to be conserved in the class Ia and class Ib synthetases, together with the features of the two short linkers connecting the three domains. The SC-fold domain contains the class I specific KMSKS motif for the enzyme reaction, and furthermore, is likely to dock the inner side of the L-shaped tRNA. These structural features that define class I and/or subclass Ia/Ib of the aminoacyl-tRNA synthetases will serve as a solid basis for understanding how the primordial Rossmann-fold enzyme evolved by C-terminal acquisition of the SC-fold domain for holding the L-shaped tRNA molecule, and how specific types of RNA-binding domains have evolved for distinguishing distinct anticodons.

Materials and methods

Preparation of the monomeric fragment of T. thermophilus *MetRS*

A 1.5 kb DNA fragment encoding amino acid residues 1–502 of *T. thermophilus* MetRS was amplified from the cloned *metS* gene [27] using polymerase chain reaction (PCR). This PCR fragment, which has the TAG termination codon in place of the AAG codon for Lys503 and the *Nde*I and *Sal*I sites created at the 5′ and 3′ ends, respectively, was cloned into the *Nde*I/*Sal*I sites of the T7 polymerase expression vector, pK7 [45]. *E. coli* JM109(DE3) cells were transformed with the recombinant plasmid. The cells were cultured at 37°C in 20 l of medium (pH 7.0) containing 10 g/l Bactotryptone, 5 g/l yeast extract, 5 g/l NaCl, and 50 μg/ml kanamycin. After cultivation for 8 h, isopropylthiogalactoside (IPTE) was added to a concentration of 0.5 mM, and the cultivation was continued overnight. About 200 g of wet cells were harvested using centrifugation. The extract of the harvested cells was incubated at 70°C for 20 min to denature the *E. coli* proteins. The recombinant *T. thermophilus* MetRS was purified to near homogeneity by chromatography on a DEAE-Sephacel column, as described previously [27]. The MetRS fraction from the DEAE-Sephacel chromatography was further applied to a Phenyl Superose column (FPLC, Pharmacia), and the elution was performed, at a flow rate of 1 ml/min, with a linear gradient (from 800 to 0 mM) of $(NH_4)_2SO_4$ in 50 mM potassium phosphate buffer (pH 7.3) containing 25 mM 2-mercaptoethanol. The MetRS fractions were collected and used for the crystallization. The Lys3→Cys (K3C) mutant was prepared using site-directed mutagenesis of the MetRS gene by PCR, followed by the same purification procedures as described previously.

The selenomethionylated MetRS was prepared as follows. Cells of an *E. coli* methionine-auxotroph strain, B834(DE3), were transformed with the monomeric *T. thermophilus* MetRS construct. The cells were cultured in 5 ml of luria broth medium overnight at 37°C and then were grown in 100 ml of LeMaster medium [46] to an OD_{600} of 0.7. This culture was used to inoculate 8 l of LeMaster medium. The culture was induced with IPTE at an OD_{600} of 0.37 and was harvested at an OD_{600} of 1.0. This growth medium is that described by LeMaster and Richards, supplemented with 30 μg/l selenomethionine, 100 μM $CaCl_2$, 1 μM $ZnSO_4$, vitamin mixture (100 mg/l each of biotin, choline chloride, folic acid, nicotinamide, *d*-pantothenate, and pyridoxal, and 100 μg/l of riboflavin), and 5 μg/l thiamin. The purification was carried out by the same procedures as those described above, except that all of the buffers were degassed and contained 10 mM dithiothreitol (DTT) and 1 mM ethylenediaminetetraacetic acid.

Crystallization

Crystals of the monomeric *T. thermophilus* MetRS were grown at 20°C by vapor diffusion in hanging drops and were enlarged by microseeding. A 10 μl drop containing 15 mg/ml protein buffered with 5% PEG 6000, 1 mM DTT, and 100 mM HEPES pH 7.5 was equilibrated against a reservoir solution containing 25% PEG 6000, 1 mM DTT, and 100 mM HEPES pH 7.5. One crystal harvested after four days was inserted as a seed into a 10 μl hanging drop containing 10 mg/ml protein buffered with 3–5% PEG 6000, 1 mM DTT, and 100 mM HEPES, pH 7.5. The rod-shaped crystals grew to average dimensions of $0.2 \times 0.2 \times 0.4$ mm in about one week. The crystals belong to the orthorhombic space group $P2_12_12_1$, with unit-cell dimensions of $a = 57.15$ (3) Å, $b = 82.75$ (3) Å, and $c = 117.42$ (4) Å, and contain one molecule in the crystallographic asymmetric unit. The measured density, $D_{obs} = 1.22$ g/cm^3, gives a calculated solvent content of 48.1% and a V_m of 2.39 Å^3/Da.

Three heavy-atom derivatives were prepared by soaking the native crystals in solutions of 15% PEG6000, 100 mM HEPES buffer (pH 7.5) containing 1.45 mM $EtHgSC_6H_4CO_2Na$ (EMTS), for nine days, 2 mM $Sm(OAc)_3$, for two days, and 10 mM K_2PtCl_4, for eight days, respectively. Crystals of the site-directed Lys3→Cys (K3C) mutant were grown under the same conditions as those for the native crystal. The ethyl mercury chloride (EMC) derivative was prepared by soaking the

Table 2

Data collection and phasing statistics.

Data set	Native	Selenomethionine	EMTS	K_2PtCl_4	K3CEMC	$Sm(OAc)_3$
Source	PF (BL6A)	RAXIS-IIc	PF (BL6A)	PF (BL6A)	PF (BL6A)	PF (BL6A)
λ (Å)	1.0	1.54	1.0	1.0	1.0	1.0
Resolution limit (Å)	2.0	3.0	2.5	2.5	2.5	2.5
Total reflections	146,275	30,658	36,603	60,136	52,140	69,690
Redundancy	4.3	2.9	2.3	3.5	3.3	3.0
I/σ	9.6 (1.5)	14.2 (4.5)	5.9 (1.9)	7.4 (1.4)	15.9 (4.3)	17.6 (4.3)
Unique reflections	34,121	10,767	16,055	17,266	15,869	17,444
Completeness (%)	89.6 (74.7)	92.9 (87.2)	81.3 (65.2)	87.4 (67.2)	80.3 (59.3)	88.4 (70.6)
R_{merge} (%)	8.0 (28.8)	5.7 (12.1)	11.2 (27.2)	12.4 (47.9)	5.6 (17.2)	7.4 (21.5)
Phasing (15.0 ~ 3.0 Å)						
Number of sites	–	9	1	5	6	1
R_{cullis}	–	0.58	0.78	0.88	0.86	0.90
Phasing power (centric)	–	2.21	1.10	1.36	1.04	0.76
Phasing power (acentric)	–	2.32	1.44	1.46	1.07	0.97
Mean FOM after SHARP phasing		0.618				

EMTS: $EtHgSC_6H_4CO_2Na$. K3CEMC: K3C mutant and ethyl mercury chloride. The values in parentheses refer to the highest resolution shell. $R_{merge} = \Sigma | I_i - \langle I \rangle | / \Sigma I$, where I_i is the intensity of an individual measurement,and <I> is the mean intensity of multiple observations of symmetry-related reflections. $R_{cullis} = \Sigma | | F_{PH} - F_P | - F_{hcal} | / \Sigma | F_{PH} - F_P |$ defined for centric reflections, where F_P, F_{PH} and F_{hcal} are native, derivative and calculated heavy-atom structure factor amplitudes, respectively. Phasing power = rms of the heavy-atom structure-factor amplitides divided by the lack of closure.

K3C crystals in the solution of 15% PEG 6000, buffered with 100 mM HEPES, and containing 0.5 mM EMC, for three days. Crystals of the selenomethionine-substituted MetRS were also grown by microseeding. A small native crystal was inserted as a seed into a 10 µl drop containing the selenomethionine-substituted MetRS buffered with 5% PEG 6000, 10 mM DTT, 1 mM EDTA, and 100 mM HEPES pH 7.5. A crystal harvested from this drop was again inserted into another drop under the same conditions. The crystals of the selenomethionine-substituted MetRS were enlarged to dimensions of $0.1 \times 0.1 \times 0.3$ mm.

Native protein crystals were soaked in the solution of 15% PEG 6000, 100 mM HEPES buffer (pH 7.0) containing 0.6 mM ATP. The native crystals soaked in the solution containing 0.5 mM methionine were cracked. Co-crystallizations with ATP or methionine were also tried in the protein solution containing 0.5–5 mM ATP or 0.2–0.3 mM methionine under the same conditions as those of the native crystallization.

Data collection and MIR phasing

All data sets, except those for the selenomethionylated crystals, were collected at 10°C using a Weissenberg-type imaging-plate diffractometer installed on beam line BL6A at the Photon Factory (Tsukuba, Japan). The data set of the selenomethionylated crystal was collected at 20°C on an RAXIS-IIc imaging-plate using a Rigaku rotating anode generating Cu Kα radiation. The intensities were processed with the DENZO and SCALEPACK programs [47]. The crystal structure was determined using multiple isomorphous replacement (MIR). The Patterson map calculation, heavy-atom refinement, solvent flattening (DM), histogram-matching algorithm of DM, and electron-density map calculations were performed using the CCP4 program suite [48]. The Patterson map of the EMTS derivative was interpreted, and the sites of the other three heavy-atom derivatives and the Se atom sites of the selenomethionylated protein were determined using difference Fourier maps phased with the EMTS derivative. Given that the selenomethionylated crystals had excellent isomorphism, the native and selenomethionylated proteins were used mainly to determine the phases. On the other hand, the four heavy-atom derivatives were used for reinforcement, because their isomorphism was poor, caused by soaking with the heavy metal. The MIR phases calculated using the program SHARP [49] gave more meaningful maps in the regions of the loop on the boundary with the solvent, especially the loop containing the KMSKS motif. The phasing statistics are listed in Table 2.

Model building and refinement

Chain tracing in the solvent-flattened MIR map at 3.0 Å was carried using the graphic program O [50], and the identified positions of nine out of the ten selenomethionine residues became certain starting points in the model building. The model was refined initially using slow torsion in the program X-PLOR [51], followed by a simulated annealing refinement. Two residues at the C-terminal end, from a total of 502 residues, were missing. The final model contains 4098 protein atoms and 136 water molecules. The refined model gave an R factor of 21.0% at 6.0–2.0 Å resolution and an R_{free} of 29.1% for 5% of the randomly chosen reflections [52]. All of the non-glycine residues, except Val226, have ψ and φ angles that lie in the energetically allowed region of the Ramachandran diagram.

Accession numbers

The coordinates have been deposited in the Protein Data Bank, with accession number 1A8H.

Acknowledgements

We thank N Watanabe and N Sakabe for support in the synchrotron data collection at the Photon Factory (Tsukuba, Japan), and B Rees for his kind help with the calculations by the program SHARP. This project was supported in part by Grants-in-Aid for Scientific Research on Priority Areas to MK and SY from the Ministry of Education, Science, Culture and Sports of Japan.

References

1. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347,** 203-206.
2. Cusack, S. (1995). Eleven down and nine to go. *Nat. Struct. Biol.* **2,** 824-831.
3. Cavarelli, J., Delagoutte, B., Eriani, G., Gangloff, J. & Moras, D. (1998). L-arginine recognition by yeast arginyl-tRNA synthetase. *EMBO J.* **17**, 5438-5448.
4. Brick, P., Bhat, T.N. & Blow, D.M. (1989). Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. *J. Mol. Biol.* **208**, 83-98.
5. Doublié, S., Bricogne, G., Gilmore, C. & Carter Jr., C.W. (1995). Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure* **3**, 17-31.

6. Rould, M.A., Perona, J.J., Söll, D. & Steitz, T.A. (1989). Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. *Science* **246**, 1135-1142.
7. Brunie, S., Zelwer, C. & Risler, J.L. (1990). Crystallographic study at 2.5 Å resolution of the interaction of methionyl-tRNA synthetase from *Escherichia coli* with ATP. *J. Mol. Biol.* **216**, 411-424.
8. Nureki, O., *et al*., & Morikawa, K. (1995). Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. *Science* **267**, 1958-1965.
9. Nureki, O., *et al*., & Yokoyama, S. (1998). Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* **280**, 578-582.
10. Silvian, L.F., Wang, J., & Steitz T.A. (1999). Insights into editing from an Ile-tRNA synthetase structure with $tRNA^{Ile}$ and mupirocin. *Science* **285**, 1074-1077.
11. Kohda, D., Yokoyama, S. & Miyazawa, T. (1987). Functions of isolated domains of methionyl-tRNA synthetase from an extreme thermophile, *Thermus thermophilus* HB8. *J. Biol. Chem.* **262**, 558-563.
12. Cassio, D. & Waller, J.-P. (1971). Modification of methionyl-tRNA synthetase by proteolytic cleavage and properties of the trypsin-modified enzyme. *Eur. J. Biochem*. **20,** 283-300.
13. Blanquet, S., Dessen, P. & Iwatsubo, M, (1976). Antico-operative binding of bacterial and mammalian initiator $tRNA^{Met}$ to methionyl-tRNA synthetase from *Escherichia coli*. *J. Mol. Biol*. **103**, 765-784.
14. Cavarelli, J., Rees, B., Ruff, M., Thierry, J.-C. & Moras, D. (1993). Yeast $tRNA^{Asp}$ recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature* **362**, 181-184.
15. Biou, V., Yaremchuk, A., Tukalo, M. & Cusack, S. (1994). The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$. *Science* **263**, 1404-1410.
16. Cusack, S., Yaremchuk, A. & Tukalo, M. (1996). The crystal structure of *T. thermophilus* lysyl-tRNA synthetase complexed with *E. coli* $tRNA^{Lys}$ and a *T. thermophilus* $tRNA^{Lys}$ transcript: anticodon recognition and conformational changes upon binding of a lysyl-adenylate analogue. *EMBO J.* **15**, 6321-6334.
17. Cusack, S., Yaremchuk, A., Krikliviy, I. & Tukalo, M. (1998). $tRNA^{Pro}$ anticodon recognition by *Thermus thermophilus* prolyl-tRNA synthetase. *Structure* **6**, 101-108.
18. Goldgur, Y., Mosyak, L., Reshetnikova, L., Ankilova, V., Lavrik, O., Khodyreva, S. & Safro, M. (1997). The crystal structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus* complexed with cognate $tRNA^{Phe}$. *Structure*, **15**, 59-68.
19. Ghosh, G., Pelka, H. & Schulman, L.H. (1990). Identification of the tRNA anticodon recognition site of *Escherichia coli* methionyl-tRNA synthetase. *Biochemistry* **29**, 2220-2225.
20. Meinnel, T., Mechulam, Y., Corre, D.L., Panvert, M., Blanquet, S. & Fayat, G. (1991). Selection of suppressor methionyl-tRNA synthetases: mapping the tRNA anticodon binding site. *Proc. Natl Acad. Sci. USA* **88**, 291-295.
21. Ghosh, G., Kim, H.Y., Demaret, J.-P., Brunie, S. & Schulman, L.H. (1991). Arginine-395 is required for efficient *in vivo* and *in vitro* aminoacylation of tRNAs by *Escherichia coli* methionyl-tRNA synthetase. *Biochemistry* **30**, 11767-11774.
22. Kim, H.Y., Pelka, H., Brunie, S. & Schulman. L.H. (1993). Two separate peptides in *Escherichia coli* methionyl-tRNA synthetase form the anticodon binding site for methionine tRNA. *Biochemistry* **32**, 10506-10511.
23. Fourmy, D., Meinnel, T., Mechulam, Y. & Blanquet, S. (1993). Mapping of the zinc binding domain of *Escherichia coli* methionyl-tRNA synthetase. *J. Mol. Biol.* **231,** 1068-1077.
24. Fourmy, D., Mechulam, Y. & Blanquet, S. (1995). Crucial role of an idiosyncratic insertion in the Rossmann fold of class I aminoacyl-tRNA synthetases: The case of methionyl-tRNA synthetase. *Biochemistry* **34,** 15681-15688.
25. Fourmy, D., Dardel, F. & Blanquet, S. (1993). Methionyl-tRNA synthetase zinc binding domain. Three-dimensional structure and homology with rubredoxin and *gag* retroviral proteins. (1993). *J. Mol. Biol.* **231**, 1078-1089.
26. Nureki, O., Kohno, T., Sakamoto, K., Miyazawa, T. & Yokoyama, S. (1993). Chemical modification and mutagenesis studies on zinc binding of aminoacyl-tRNA synthetases. *J. Biol. Chem.* **268**, 15368-15373.
27. Nureki, O., *et al.,* & Yokoyama, S. (1991). Methionyl-tRNA synthetase gene from an extreme thermophile, *Thermus thermophilus* HB8. *J. Biol. Chem.* **266**, 3268-3277.
28. Mechulam, Y., *et al.,* & Blanquet, S. (1999). Crystal structure of *Escherichia coli* methionyl-tRNA synthetase highlights species-specific features. *J. Mol. Biol.* **294**, 1287-1297.
29. Landro, J.A. & Schimmel, P. (1994). Zinc-dependent cell growth conferred by mutant tRNA synthetase. *J. Biol. Chem.* **269**, 20217-20220.
30. Landro, J.A., Schmidt, E., Schimmel, P., Tierney, D.L. & Penner-Hahn, J.E. (1994). Thiol ligation of two zinc atoms to a class I tRNA synthetase: evidence for unshared thiols and role in amino acid binding and utilization. *Biochemistry* **33**, 14213-14220.
31. Landro, J.A. & Schimmel, P. (1993). Metal-binding site in a class I tRNA synthetase localized to a cysteine cluster inserted into nucleotide-binding fold. *Proc. Natl Acad. Sci. USA* **90**, 2261-2265.
32. Lin, L., Hale, S.P. & Schimmel, P. (1996). Aminoacylation error correction. *Nature* **384,** 33-34.
33. Liu, J., *et al*., & Lapointe, J. (1995). The zinc-binding site of *Escherichia coli* glutamyl-tRNA synthetase is located in the acceptor-binding domain. *J. Biol. Chem.* **270**, 15162-15169.
34. Rould, M.A. & Steitz, T.A. (1992). Structure of the glutaminyl–tRNA synthetase–$tRNA^{Gln}$–ATP complex. *Nucl. Acids Mol. Biol.* **6**, 225-245.
35. Rogers, M.J., Adachi, T., Inokuchi, H. & Söll, D. (1994). Functional communication in the recognition of tRNA by *Escherichia coli* glutaminyl-tRNA synthetase. *Proc. Natl Acad. Sci. USA* **91**, 291-295.
36. Holm, L., Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138
37. Shiba, K. & Schimmel, P. (1992). Functional assembly of a randomly cleaved protein. *Proc. Natl Acad. Sci. USA* **89,** 1880-1884.
38. Auld, D.S. & Schimmel, P. (1995). Switching recognition of two tRNA synthetases with an amino acid swap in a designed peptide. *Science* **267,** 1994-1996.
39. Rould, M.A., Perona, J.J. & Steitz, T.A. (1991). Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature* **352**, 213-218.
40. Berthet-Colominas, C., Seignovert, L., Hartlein, M., Grotli, M., Cusack, S. & Leberman, R. (1998). The crystal structure of asparaginyl-tRNA synthetase from *Thermus thermophilus* and its complexes with ATP and asparaginyl-adenylate: the mechanism of discrimination between asparagine and aspartic acid. *EMBO J.* **17**, 2947-2960.
41. Arnez, J.G., Harris, D.C., Mitschler, A., Rees, B., Francklyn, C.S. & Moras, D. (1995). Crystal structure of histidyl-tRNA synthetase from *Escherichia coli* complexed with histidyl-adenylate. *EMBO J.* **14**, 4143-4155.
42. Logan, D.T., Mazauric, M.-H., Kern, D. & Moras, D. (1995). Crystal structure of glycyl-tRNA synthetase from *Thermus thermophilus*. *EMBO J.* **14**, 4156-4167.
43. Oubridge, C., Ito, N., Evans, P.R., Teo, D.-H. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432-438.
44. Allain, F.H.-T., Gubser, C.C., Howe, P.W.A., Nagai, K., Neuhaus, D. & Varani, G. (1996). Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* **380,** 646-650.
45. Kigawa, T., Muto, Y. & Yokoyama, S. (1995). Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J. Bio. NMR* **6**, 129-134.
46. LeMaster, D.M. & Richards, F.M. (1985). ^{1}H-^{15}N heteronulcear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**, 7263-7268.
47. Otwinowski, Z. & Minor, W. (1996). Processing X-ray diffraction data collected in oscillation mode. *Methods Enz.* **276**, 307-326.
48. Collaborative computing project number 4. (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.
49. De la Fortelle, E. & Bricogne, G. (1996). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enz.* **276**, 472-494.
50. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgarrd, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110-119.
51. Brünger, A.T. (1993). *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*.: Yale University Press, New Haven, CT.
52. Brünger, A.T. (1992). Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.
53. Biosym Technotogies (1993). *Insight II User Guide, version 2.3.0.* San Diego: Biosym Technologies.
54. Kraulis, P. J. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.
55. Mettitt, E.A. & Murphy, E.P. (1994). *Raster3D* Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869-873.