The EMBO Journal vol.14 no.17 pp.4156–4167, 1995

Crystal structure of glycyl-tRNA synthetase from *Thermus thermophilus*

D.T.Logan[1,2], M.-H.Mazauric[3], D.Kern[3] and D.Moras[1,4]

[1]Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, BP 163, 67404 Illkirch Cedex and [3]Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, 67084 Strasbourg Cedex, France
[2]Present address: Department of Molecular Biology, University of Stockholm, S-106 91 Stockholm, Sweden

[4]Corresponding author

The sequence and crystal structure at 2.75 Å resolution of the homodimeric glycyl-tRNA synthetase from *Thermus thermophilus*, the first representative of the last unknown class II synthetase subgroup, have been determined. The three class II synthetase sequence motifs are present but the structure was essential for identification of motif 1, which does not possess the proline previously believed to be an essential class II invariant. Nevertheless, crucial contacts with the active site of the other monomer involving motif 1 are conserved and a more comprehensive description of class II now becomes possible. Each monomer consists of an active site strongly resembling that of the aspartyl and seryl enzymes, a C-terminal anticodon recognition domain of 100 residues and a third domain unusually inserted between motifs 1 and 2 almost certainly interacting with the acceptor arm of $tRNA^{Gly}$. The C-terminal domain has a novel five-stranded parallel–antiparallel β-sheet structure with three surrounding helices. The active site residues most probably responsible for substrate recognition, in particular in the Gly binding pocket, can be identified by inference from aspartyl-tRNA synthetase due to the conserved nature of the class II active site.

Keywords: crystal structure/glycyl-tRNA synthetase/prokaryote/sequence motifs/substrate modelling

Introduction

The aminoacyl-tRNA synthetases (aaRS) are among the most primeval enzymes and among the most important for our understanding of the origins of life by virtue of their pivotal role in the translation of genetic information (Schimmel, 1987). In a two-step reaction they catalyse the acylation of tRNA molecules with their cognate amino acids. Sequence alignments (Eriani *et al.*, 1990), in tandem with structural analyses (Brick *et al.*, 1989; Rould *et al.*, 1989; Brunie *et al.*, 1990; Cusack *et al.*, 1990; Ruff *et al.*, 1991; Doublié *et al.*, 1995; Onesti *et al.*, 1995), have shown that the aaRS are divided into two distinct classes. Class I is structurally characterized by an active site based on the oligonucleotide binding 'Rossmann fold', as in the structures of TyrRS (Brick *et al.*, 1989), MetRS (Brunie *et al.*, 1990), GlnRS (Rould *et al.*, 1989) and TrpRS (Doublié *et al.*, 1995). Class II active sites are built around a completely different seven-stranded antiparallel β-sheet motif, e.g. SerRS (Cusack *et al.*, 1990), AspRS (Ruff *et al.*, 1991), PheRS (Mosyak and Safro, 1993), LysRS (Onesti *et al.*, 1995) and HisRS, the structure of which is presented in the accompanying paper (Arnez *et al.*, 1995). This fold was for a time believed to be unique to the class II aaRS, but striking structural similarity has been demonstrated recently between SerRS and the biotin synthetase/repressor protein, which also creates an acyl-adenylate intermediate (Artymiuk *et al.*, 1994).

With the single exception of PheRS, the partition into two classes also correlates entirely with the site of aminoacylation: 2′ OH of the terminal ribose for class I, 3′ OH for class II (Fraser and Rich, 1975; Eriani *et al.*, 1990). Class II aaRS can be further classified in three subgroups based on more local sequence homology (Moras, 1992): subgroup IIa contains HisRS, ProRS, SerRS and ThrRS (similar C-terminal domains except for SerRS, where high local homology is in the active site domain); IIb contains AsnRS, AspRS and LysRS (similar N-terminal domains); IIc contains the aaRS of unusual quaternary structure: AlaRS, PheRS and *Escherichia coli* GlyRS.

The aaRS have a modular structural organization (Delarue and Moras, 1993). AspRS from yeast can be divided into an active site region and an N-terminal anticodon recognition domain built around a five-stranded parallel–antiparallel β-barrel, which has been identified as a general oligonucleotide binding fold (Murzin, 1993). The latter domain has also been found recently in LysRS (Onesti *et al.*, 1995), confirming the hypothesis based on sequence analysis that it defined a class II subgroup (IIb; Gatti and Tzagoloff, 1991). SerRS contains a similar active site unit to which is appended a long N-terminal domain, so far unique in the aaRS, in the form of an α-helical coiled-coil which interacts with three of the arms of $tRNA^{Ser}$, including the unusual long variable arm (Biou *et al.*, 1994). To this modular organization may be added the large 'insertion domain' characteristic of prokaryotic AspRS, which provides supplementary interactions with the acceptor arm of $tRNA^{Asp}$ and which may possess other functions (Delarue *et al.*, 1994).

The class II aaRS bear three characteristic sequence motifs (Eriani *et al.*, 1990a): the first is principally involved in dimerization (all but one known class II synthetase structures are homodimers), but also participates in orientation of motifs 2 and 3, which contain essential components of the catalytic mechanism (Eriani *et al.*, 1993; Cavarelli *et al.*, 1994). Detailed analyses of interactions of the enzyme with amino acid, ATP and aminoacyl-adenylate or its homologues have been published for both AspRS (Cavarelli *et al.*, 1994; Poterszman *et al.*, 1994) and SerRS (Belrhali *et al.*, 1994, 1995). Anticodon recognition is

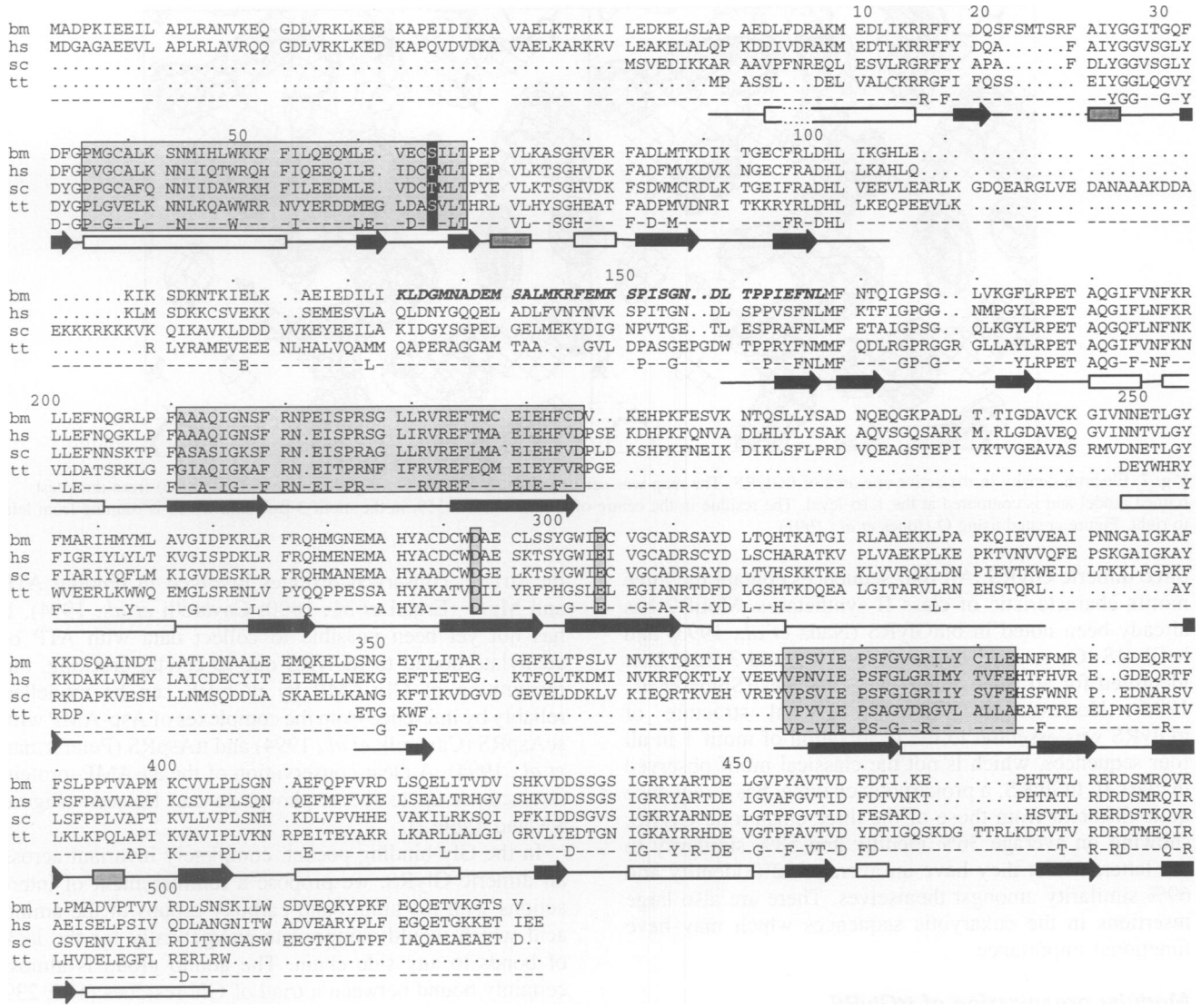


Fig. 1. Multiple sequence alignment of the sequences of the four known dimeric GlyRS, obtained using the program PILEUP of the UWGCG suite (Genetics Computer Group, University of Wisconsin, USA). Abbreviations: bm = *B.mori*, hs = *Homo sapiens*, sc = *Saccharomyces cerevisiae*, tt = *T.thermophilus*. The hsGlyRS and scGlyRS sequences were found by searching a non-redundant compilation database using BLAST (Altschul *et al.*, 1990), which failed to detect any similarity to the ecGlyRS sequence. The numbering presented above the alignment and the secondary structure drawn below it correspond to ttGlyRS. Secondary structure is that defined by the program PROCHECK (Laskowski *et al.*, 1993). Alpha helices are white boxes, 3_{10} helices are grey boxes and β-strands are black arrows. The three class II sequence motifs are boxed. The non-canonical motif 1 Ser/Thr is shown in white on a dark background. The original motif 1 proposed by Cusack (1993) is shown in italics. The fifth line shows residues conserved or conservatively substituted across all four sequences and thus of likely functional or structural importance.

also known at the molecular level for AspRS (Cavarelli *et al.*, 1993). In $tRNA^{Ser}$ the anticodon is not an identity element, does not contact the protein and was disordered in the crystal structure of the complex with SerRS from *Thermus thermophilus* (Biou *et al.*, 1994).

GlyRS is one of the more unusual synthetases as it varies in quaternary structure between organisms: in *E.coli* (ecGlyRS) and *Bacillus brevis* it is an $\alpha_2\beta_2$ tetramer (Ostrem and Berg, 1974; Surgochov and Surgochova, 1975), in *Bacillus stearothermophilus, Bombyx mori* (bm), *T.thermophilus* (tt), yeast (sc) and man (hs) it is an α_2 dimer (Kern *et al.*, 1981; Nada *et al.*, 1993; Ge *et al.*, 1994; Shiba *et al.*, 1994). No unambiguous motif 1 has yet been identified in any GlyRS and, in addition, the *E.coli* enzyme lacks a clear motif 2. This previously led to its classification with AlaRS and PheRS in subgroup IIc (Moras, 1992). The wide divergence between prokaryotic ecGlyRS and the three known eukaryotic GlyRS is correlated with an inability of ecGlyRS to aminoacylate human $tRNA^{Gly}$ and vice versa, leading to the suggestion that the divergence is linked to the discriminator base 73, U in prokaryotes and A in eukaryotes (Shiba *et al.*, 1994). Here we present the structure of the first GlyRS, from the prokaryote *T.thermophilus*, which bears surprising resemblances at the sequence level to the eukaryotic enzymes, despite being apparently unrelated to GlyRS from *E.coli*.

Results and discussion

Sequence similarities between prokaryote and eukaryotes

The ttGlyRS dimer has a molecular weight of 116 kDa and is composed of 2×506 amino acids. Its sequence is presented in Figure 1 in alignment with those of the other

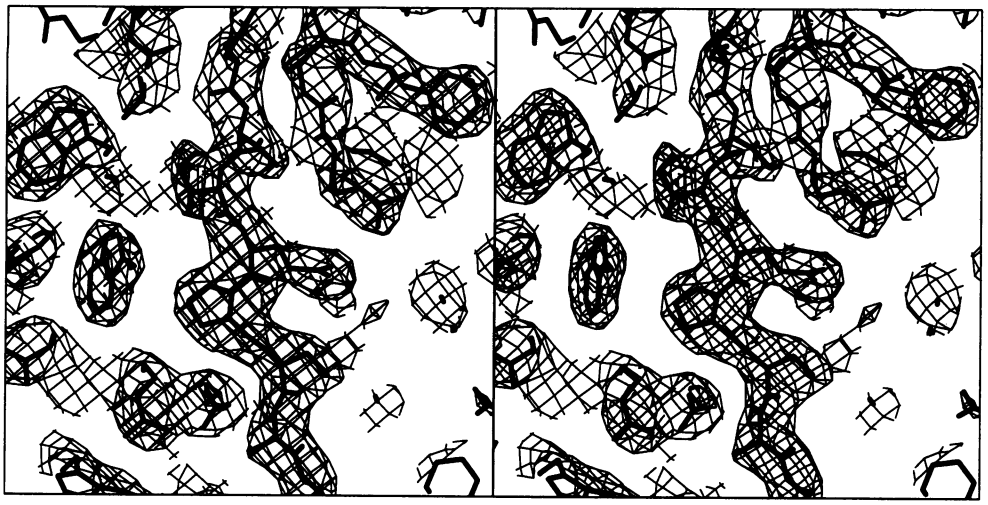


Fig. 2. Electron density in the active site area of ttGlyRS. The map was calculated using SIGMAA-weighted phases (Read, 1986) from the latest refined model and is contoured at the 1.1σ level. The residue in the centre of the view is Glu359, in the motif 3 β-strand, which is running from left to right. Figure created using O (Jones *et al.*, 1991).

three dimeric GlyRS. All four sequences contain the three motifs characteristic of class II synthetases. Motif 2 has already been noted in bmGlyRS (Nada *et al.*, 1993) and hsGlyRS (Ge *et al.*, 1994), and motifs 2 and 3 were later identified for all three eukaryote sequences (Shiba *et al.*, 1994). However, the three-dimensional structure of ttGlyRS was essential to the localization of motif 1 in all four sequences, which is not the classical motif observed in class II. ttGlyRS, a prokaryotic enzyme, has a sequence quite different from those of the three eukaryotic GlyRS, showing on average 36% identity and 58% similarity to the latter, whilst they have an average 52% identity and 69% similarity amongst themselves. There are also large insertions in the eukaryotic sequences which may have functional importance.

Modular organization of ttGlyRS

The structure of ttGlyRS was determined at 2.75 Å by the multiple isomorphous replacement (MIR) method using four derivatives, in conjunction with 2-fold non-crystallographic symmetry (NCS) averaging and solvent flattening. Figure 2 shows a representative portion of the final electron density in the active site region and the fit of the latest model. Each monomer of ttGlyRS is composed of three domains (Figure 3): an active site module, a domain of ~80 residues inserted between motifs 1 and 2 and lying above the active site, and a C-terminal domain of ~100 amino acids consisting of a novel five-stranded mixed β-sheet with three flanking helices. In the dimer, this latter domain, almost certainly involved in anticodon recognition, makes contact with the active site module of both monomers. Like those of AspRS, SerRS and LysRS, the active site consists of a seven-stranded antiparallel β-sheet with flanking helices (Figure 3). Our model for the insertion domain is at present incomplete due to crystalline disorder, but it contains conserved residues likely to interact with the acceptor arm of $tRNA^{Gly}$.

The active site: variations on amino acid recognition

As in all class II synthetases, motifs 2 and 3 in ttGlyRS contribute to the core of the active site and contain all of the strictly conserved residues involved in binding ATP and Mg^{2+} (Eriani *et al.*, 1990; Cavarelli *et al.*, 1994). It has not yet been possible to collect data with ATP or adenylate bound to ttGlyRS due to crystal cracking, but recognition of Gly-AMP by ttGlyRS can be modelled reliably by inference from the complexes of Asp-AMP with scAspRS (Cavarelli *et al.*, 1994) and ttAspRS (Poterszman *et al.*, 1994). A clear conservation of the aa-AMP–protein interactions up to the α phosphate can be seen (Figure 4A and Table I).

In the Gly binding pocket, completely invariant across all dimeric GlyRS, we propose a reinforcement of interactions with the amino and carboxy groups of the amino acid which would compensate energetically for the lack of bonds to any side chain. The amino group is almost certainly bound between a triad of Glu residues (188, 239 and 359), instead of by only Asp342 in scAspRS. Thus modelled, the NH_3^+ moiety of Gly-AMP replaces a strongly bound water molecule in the apoenzyme structure. A single interaction with the carboxy group would then be sufficient to define a unique orientation at the Cα atom of Gly or Gly-AMP. However, binding is strengthened by two interactions, with Gln237 and Ser361. In this orientation, any side chain would clash with a rigid wall of side chains, in particular Glu359 and Arg311, the latter held firmly in place by a bidentate hydrogen bond to Asp315. Rigidity of Arg311 is also important to prevent it from blocking entry of amino acid and ATP by extending into the pocket. This residue, equivalent to Arg485 in scAspRS, was previously thought to be unique to the AspRS family, as in that context it is partly responsible for specific recognition of Asp (Cavarelli *et al.*, 1994; Poterszman *et al.*, 1994). The side chain equivalent to Glu359 is almost always small and non-polar (Gly, Ala, Cys or Ser) in class II synthetases, except for GlyRS and AlaRS, which suggests that an acid at this position may be a major selectivity determinant in the latter enzymes. This strand, known as the 'Gly-rich strand', also contains a Gly or other side chain with no hydrogen bonding capacity at the position equivalent to Ser361 in almost all other class II aaRS, which supports the hypothesis that

this Ser is important for reinforcement of the correct orientation of Gly.

Gly-AMP fits snugly into a relatively buried recognition pocket and a space below, tailor-made for the initial bent conformation of ATP, can be seen (Figure 4B). It seems that here, as in scAspRS (Cavarelli *et al.*, 1994) and ttAspRS (Poterszman *et al.*, 1994), the amino acid binds to a rigid pre-formed pocket. The ATP pocket is also rather well defined, even in the absence of substrate, but the motif 2 loop has higher than average B-factors and probably becomes more ordered on ATP binding, as seen in AspRS and ttSerRS (Belrhali *et al.*, 1994). This would bring into play, in recognition of the adenosine, the class II conserved Glu222. The pocket is highly negatively charged overall due to the presence of numerous Glu residues, of which only some are neutralized by salt bridges. In particular, Glu359 and Glu241 form a hydrogen-bonded pair, which must require alteration of at least one of their pK_a values by the local environment (largely hydrophobic for Glu241). Such clusters of acidic residues have been implicated as switches in folding of tobacco mosaic virus (Namba and Stubbs, 1986) but, as in other virus structures where they occur frequently, they are known or thought to be ion coordination sites, and in ttGlyRS we see no evidence for such ion binding. The strong field generated by these Glu residues may be important in attracting and strongly binding the NH_3^+ group.

A non-canonical motif 1

Motif 1 consists as usual of the long helix H2 followed by a short extra β-strand (B3) parallel to the first strand

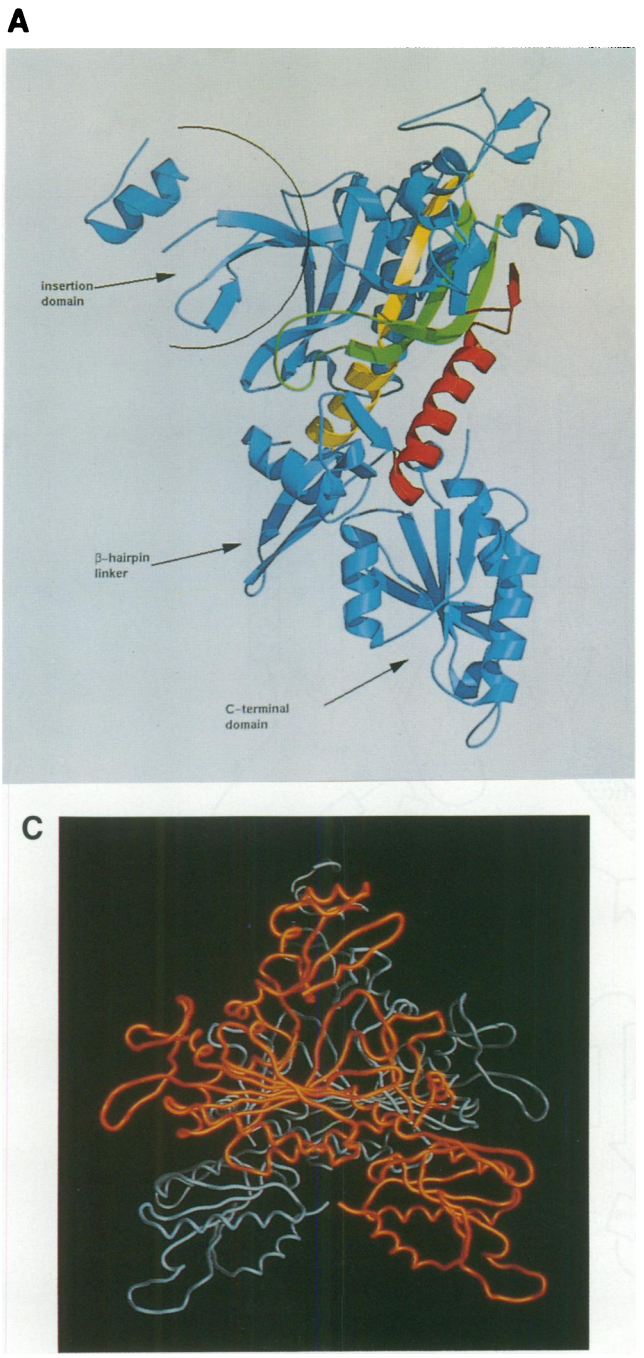
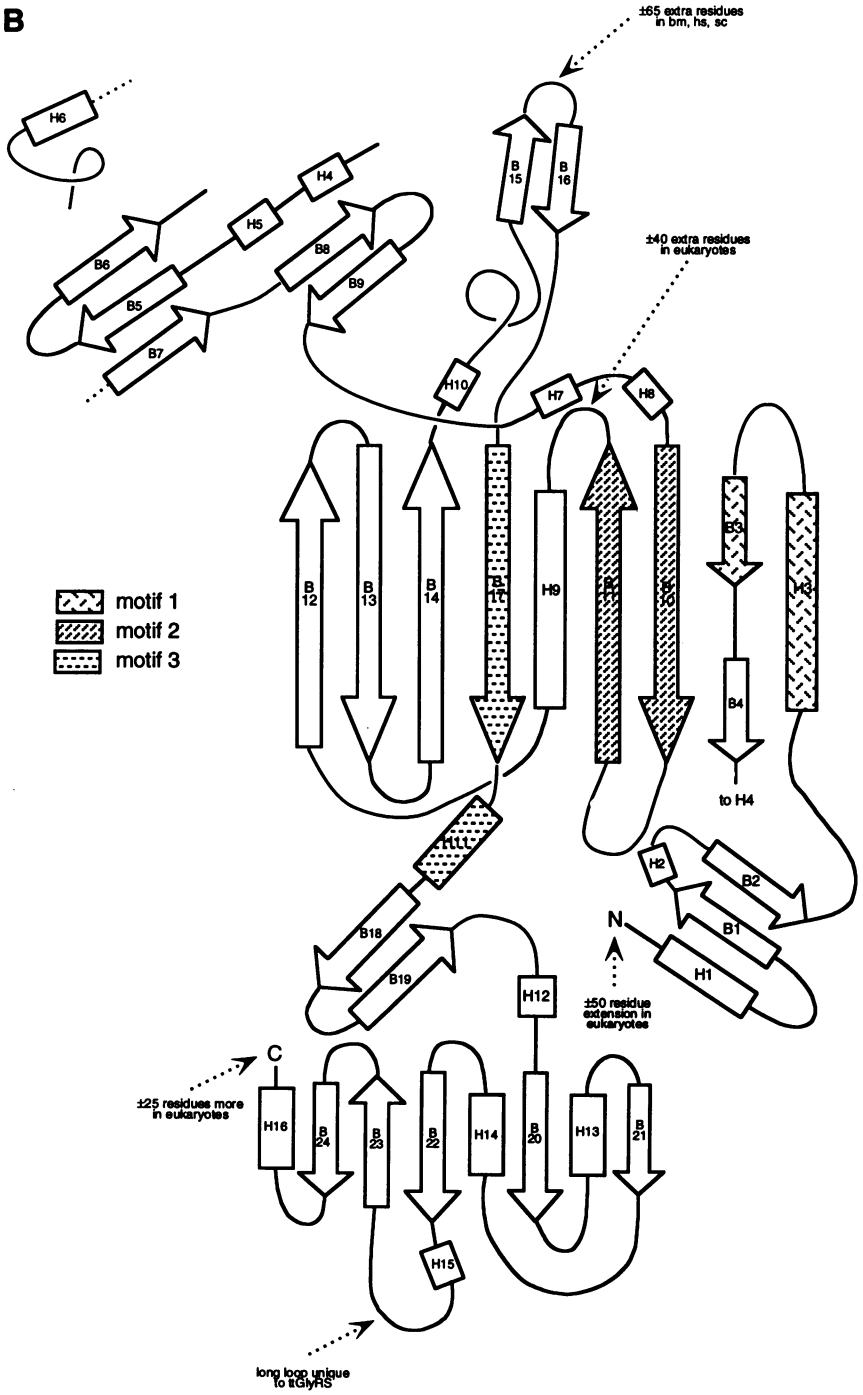


Fig. 3. (**A**) The tertiary structure of ttGlyRS. Only one monomer is shown for clarity. Sequence motifs 1–3 are colour-coded red, green and yellow respectively. (**B**) Topology diagram of the secondary structure elements making up ttGlyRS. The insertion domain and its connections to the active site module are drawn at the top left for clarity. In reality they lie above the active site and strand B7 is in proximity to the motif 2 loop. The positions of the principal insertions and deletions in eukaryotic GlyRS sequences are shown. (**C**) and (**D**) Dimeric association in ttGlyRS. One monomer is shown in orange, the other in blue. The views are perpendicular to the dimer axis, which is running vertically; in (D) the molecule is seen at ~90° to its orientation in (C). The insertion domains of both monomers are at the top and the C-terminal domains at the bottom of this view. Note the differing orientations of the insertion domains with respect to the active site modules of respective monomers.

of the active site β-sheet. However, in contrast to other known class II structures, motif 1 surprisingly does not contain the conserved proline previously thought canonical for this motif (it is normally the only conserved residue). It is replaced by Ser66 in ttGlyRS and by Ser in bmGlyRS and Thr in hsGlyRS and scGlyRS.

The role of motif 1 in communication between monomers of AspRS has been investigated kinetically by mutation of the conserved Pro to Gly in the enzyme from yeast (Eriani *et al.*, 1993) and by analysis of its environment in two AspRS crystal structures (Eriani *et al.*, 1993; Delarue *et al.*, 1994). Two major interactions were seen to be responsible for a dramatic loss of activity observed on the Pro→Gly mutation. In ttGlyRS, both of these interactions are preserved despite a Pro→Ser substitution (Figure 5). The first involves a bidentate H bond between Glu234′ (n′ signifies a residue from monomer 2) and the main chain Ns of Val67 and Leu68. The immediately neighbouring Phe235′ stacks with the adenosine moiety of Gly-AMP. The second involves Leu68, which is packed in a hydrophobic pocket containing, amongst others, Phe219′: the latter's immediate neighbour Arg220′ binds the α phosphate and N7 of the adenylate. In scAspRS, the motif 1 Pro was attributed the role of locking the main chain conformation in the surrounding peptide. In particular, the Ns of Val67 and Leu68 must both point towards Glu234′. In ttGlyRS the motif 1–motif 2 interactions are strictly maintained in the



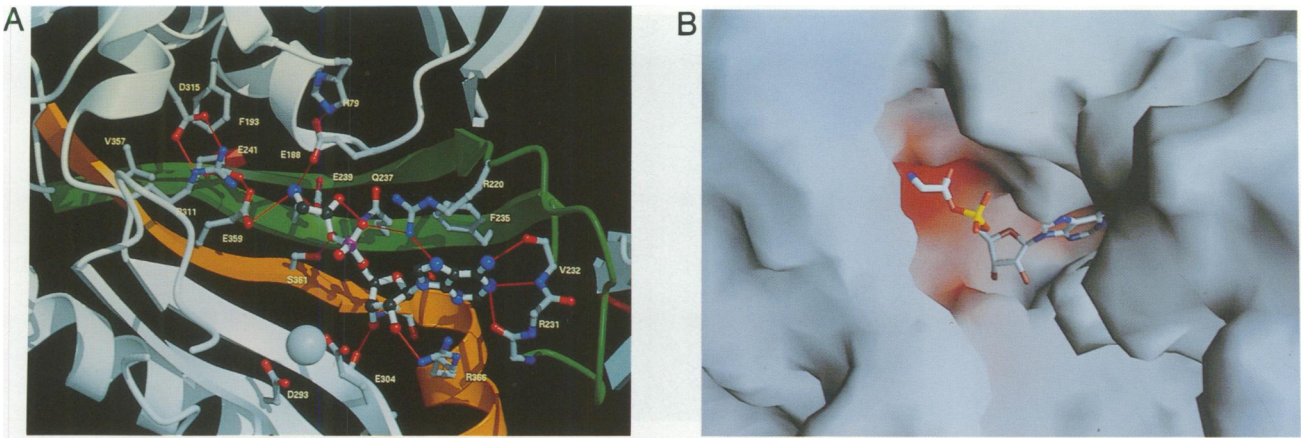


Fig. 4. (**A**) A model for glycyl-adenylate recognition by ttGlyRS based on the refined structure of the scAspRS–aspartyl-adenylate complex. Motif 2 is represented in green, motif 3 in orange. Gly-AMP is drawn in a ball-and-stick representation with white bonds; side chains interacting directly with Gly-AMP are shown with split bonds. Possibly important hydrogen bonds are represented as red lines. (**B**) The molecular surface of the active site region, calculated and displayed using GRASP (Nicholls *et al.*, 1991). The surface is coloured according to surface potential. Negative charge is drawn in red, positive in blue. This pocket represents the most negatively charged area of the whole molecule. Modelled Gly-AMP is drawn with bonds represented as cylinders.

absence of Pro since the main chain conformation of residues 64–69 is very similar to that in scAspRS (Table II) and also ttSerRS (not shown). The H bond involving Oγ of Ser66 is probably essential for achieving this, and suggests why dimeric GlyRS, despite not choosing Pro for their motif 1, use instead residues with small O-containing side chains (Ser and Thr). A number of other interactions also compensate for the loss of rigidity conferred by a Pro (Figure 5 and Table II). The two bonds from the zone 64–68 to the side chain of Arg186 and a single bond to Asp32 explain the strict conservation of the latter residues in all dimeric GlyRS.

Two other motif 1 residues play a crucial role through another hydrogen bonding network: Glu236′ →Lys42′→ Asp64. This underlines the importance of the proximity of motifs 1 belonging to adjacent monomers, as well as that of dimerically related motifs 1 and 2. It also highlights the functional necessity for the two additional N-terminal β-strands B1 and B2, which partially replace C-terminal residues also implicated in dimer stability in scAspRS (Eriani *et al.*, 1993).

C-terminal domain

The C-terminal 100 residues of ttGlyRS constitute a separate domain linked to the active site module by a long β-hairpin linker (B18/B19) immediately following the motif 3 helix. The fold is a five-stranded mixed parallel–antiparallel open twisted β-sheet with three flanking helices. This fold is somewhat reminiscent of nucleotide binding units. A search through a database of known protein structures reduced to oriented secondary structure elements (program DEJAVU by G.J.Kleijwegt and T.A.Jones) revealed limited similarity to part of the FAD binding domain of 3-isopropylmalate dehydrogenase from *T.thermophilus* (Imada *et al.*, 1991), which can be superimposed on the C-terminal domain of ttGlyRS with an r.m.s. deviation of 2.9 Å for 63 Cα atoms compared, including all the β-strands and helices. However, the strand connectivity in ttGlyRS is different and does not represent a Rossmann fold since one of the strands (B23) is antiparallel. Thus, the ttGlyRS C-terminal fold appears to be unique, at least as an independent functional unit.

Table I. Summary of possible interactions between Gly-AMP and ttGlyRS

Adenylate functional group recognized	Residue in scAspRS	Residue in ttGlyRS
Adenine moiety	Phe338	Phe235
(stacking)	Arg531	Arg366
N1	CO Met335	CO Val232
		CO Phe230
N6[a]	N Met335	N Val232
		CO Val232
N7	Nη Arg325	Nη Arg220
O2′	Nη Arg531	Nη Arg366
O3′	CO Ile479	CO Leu305
OPα	Nη Arg325	Nη Arg220
CO		Nε Gln237
OH		Oγ Ser361
NH_3^+	Oδ Asp342	Oε Glu188
		Oε Glu239
		Oε Glu359
Mg^{2+}	Oδ Asp471	Oδ Asp293
	Oε Glu478	Oε Glu304

Interactions between Asp-AMP (excluding the side chain), as observed in the complex with scAspRS (Cavarelli *et al.*, 1994), and Gly-AMP, as modelled in ttGlyRS, are listed. Note the reinforcement of interactions with the carboxy and amino groups of Gly-AMP.
[a]The class II invariant Glu222 in the motif 2 loop is equivalent to Glu225 in ttAspRS which becomes ordered on adenylate formation and makes a supplementary interaction with N6. This is no doubt also its role in ttGlyRS.

This domain spatially replaces the N-terminal anticodon binding domain of AspRS (Ruff *et al.*, 1991; Delarue *et al.*, 1994) and presumably assumes the same function, as already suggested (Cusack, 1993). The anticodon is an identity determinant for $tRNA^{Gly}$ in both *E.coli* (McClain *et al.*, 1991) and *T.thermophilus*. (M.-H.Mazauric *et al.*, in preparation). Assuming a class II conserved binding mode for the acceptor end of tRNA, a hypothesis first proposed by Ruff *et al.* (1991) and confirmed by the crystal structure of the SerRS–$tRNA^{Ser}$ complex (Biou *et al.*, 1994), the anticodon comes into contact with this domain. Figure 6 proposes a model for $tRNA^{Gly}$ recognition based on this assumption: the anticodon lies close to the base of the C-terminal domain. Helices H13

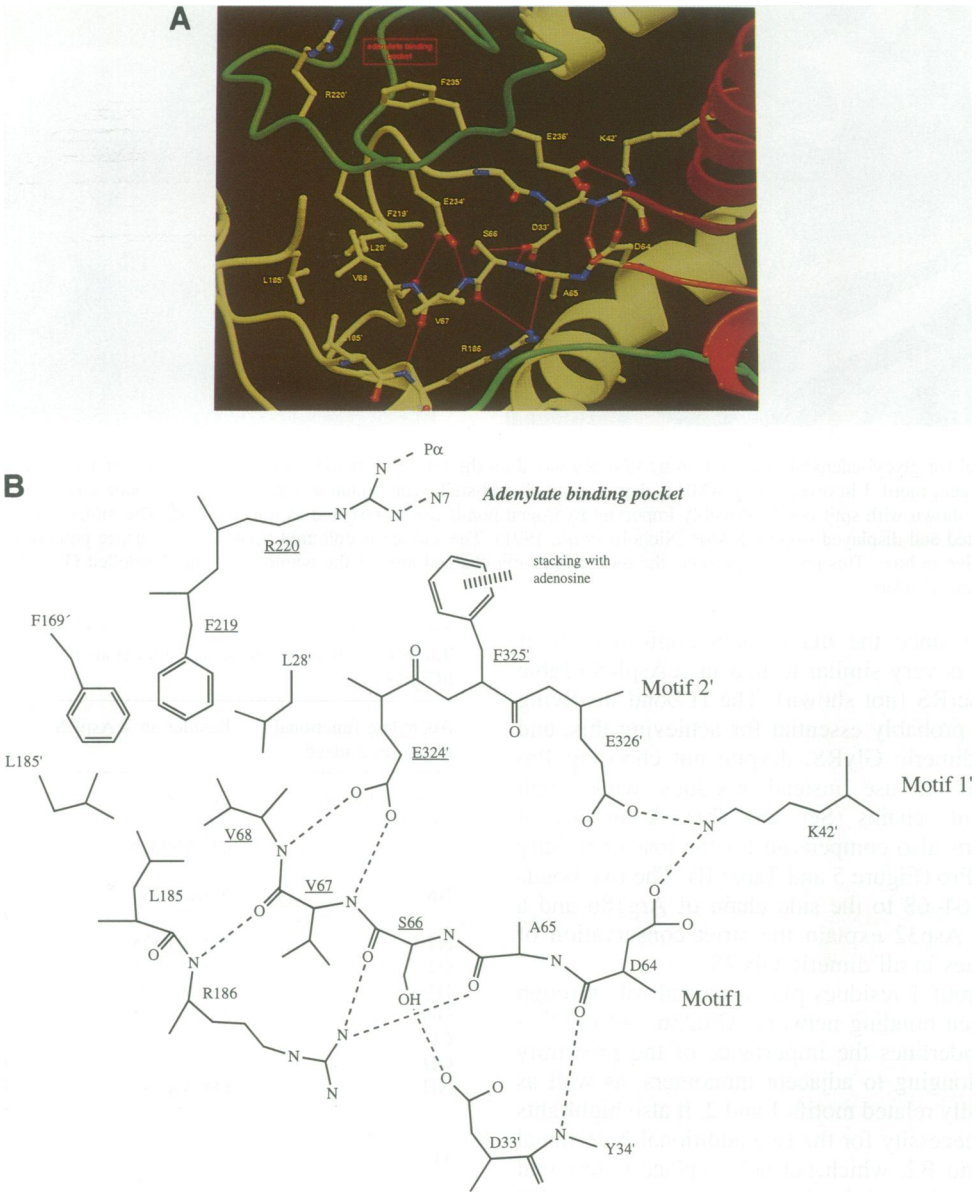


Fig. 5. (**A**) Conservation of interactions essential for communication between motif 1 of one monomer and the active site of the other, despite the absence of the 'canonical' Pro. Motif 1 residues are shown in red, motif 2 in green, motif 1′ in dark red and motif 2′ in dark green (where a prime represents a motif from the second monomer). Critical hydrogen bonds are shown as red lines. The adenosine moiety of Gly-AMP would lie immediately above Phe235′, as seen in Figure 4. (**B**) Schematic diagram showing the essential interactions. Hydrogen bonds are shown as dotted lines, and the labels of the most important residues in the pathway are underlined.

and H14 both run away from the proposed binding site and the dipole of at least H14 may aid in positioning the phosphate backbone, as seen in dinucleotide binding proteins (Wierenga *et al.*, 1985). Residues on the face of the β-sheet in contact with modelled tRNA show much higher conservation than those on the other face and those in helices H13 and H15, which are almost entirely hidden from tRNA. Figure 6C shows the residues in the C-terminal domain which are invariant or conservatively substituted across the four known sequences. A few of these residues are involved in interactions which reinforce tertiary structure and provide links to the active site domain. Nevertheless, a cluster of residues in strands B23 and B24 and in helices H14 and H15 may be involved in base-specific interactions. This region is very close to the anticodon of the modelled tRNA. Interestingly, this is precisely the region which invariably contributes loops to the active site of enzymes having similar α/β folds. However, one must await the structure of the ttGlyRS–tRNA^{Gly} complex in order to understand this fully.

Careful sequence alignments (Cusack *et al.*, 1991) have pinpointed regions of homology in the C-terminal regions of HisRS, ProRS and ThrRS, which can now be seen to extend to the dimeric GlyRS. (SerRS can also be classified in this subgroup on the basis of structural and sequence similarity in the active site, but possesses no anticodon recognition domain.) The five-stranded mixed β-sheet motif has also been found in the structure of HisRS from *E.coli* presented in the accompanying paper (Arnez *et al.*, 1995), which confirms the grouping of these aaRS in the same subclass (IIa), and suggests that the C-terminal domains of ProRS and ThrRS will also have this fold.

Table II. Interactions between motif 1 of one monomer and the active site residues of the other (denoted by ')

ttGlyRS				scAspRS			
Residue	ϕ	ψ	Interactions	Residue	ϕ	ψ	Interactions
Asp64	−139	91	O→N Tyr34′ Oδ→Nζ Lys42′	His271	−132	97	Nδ→ Oγ Thr339′ Nε→ Oε Glu269
Ala65	−90	158	O→ Nζ Arg186	Thr272	−94	160	Oγ→ Oε Glu319
Ser66	−84	163	N→Oδ Asp33′ O→Nζ Arg186 Oγ→Oδ Asp33′	Pro273	−56	165	–
Val67	−74	−40	N→Oε Glu234′ O→N Arg186	Lys274	−120	−19	N→ Oε Glu337′ O→N Ala299 Nζ →Oε Glu319
Leu68	−75	111	N→Oε Glu234′ side chain in hydrophobic pocket with Leu28, Leu185, Leu28′, Leu185′, Phe169′, Phe219′	Leu275	−89	114	N→ Oε Glu337′ side chain in hydrophobic pocket with Phe324′, Phe 287′, Leu298′, Leu289′, Leu289, Leu298

Dimeric GlyRS are thus typical members of subclass IIa rather than IIc, where the tetrameric GlyRS from *E.coli* was grouped along with PheRS.

Insertion domain

For the first time in the class II synthetases, we observe a substantial domain inserted between sequence motifs 1 and 2. However, electron density between residues 85 and 170 is weak and it has been possible to build only 50 of the intervening 85 residues in one monomer and 45 in the other, in most cases with Ala side chains only. The mean main chain atomic B-factor is also much higher than average (90 $Å^2$ compared with 54 $Å^2$ for all the rest of the structure). The connectivity cannot yet be established fully; however, there is clearly a three-stranded antiparallel β-sheet involving the 'entry' and 'exit' strands B5 and B7, plus one other (B6). There is also a helix–loop motif, the helix of which is involved in crystal packing contacts with its partner in the other monomer of a symmetry-related molecule. The insertion domains are not related by the same NCS operator as the body of the protein, and their position may be governed partly by these crystal contacts. The high B-factors are most likely due to static conformational variation rather than dynamic disorder. This would also explain the existence of another space group with almost identical packing (Logan *et al.*, 1994).

The insertion domain is well placed to interact with the minor groove of the acceptor stem of $tRNA^{Gly}$ (Figure 6A and B). Acceptor stem elements are identity determinants for $tRNA^{Gly}$ in *E.coli* (McClain *et al.*, 1991; Yan and Francklyn, 1994), and 7-bp microhelix substrates mimicking only the acceptor stem can be aminoacylated efficiently (Francklyn *et al.*, 1992). $tRNA^{Gly}_{GCC}$, the major isoacceptor, is virtually identical in *E.coli* and *T.thermophilus* (Steinberg *et al.*, 1993), differing by only 2 bp and 4 unpaired bases overall. In particular, the acceptor arm is identical through U73 and the first 6 bp, and the anticodon stem and context are totally invariant, in spite of large differences from eukaryotic $tRNA^{Gly}_{GCC}$s. This suggests that interactions are very similar in the two prokaryotes. It is thus puzzling that no domain homologous to either the insertion or C-terminal domains of ttGlyRS can be found in the ecGlyRS sequence.

Some movement of the insertion domain away from the active site may be necessary to allow tRNA binding, as it currently clashes slightly with the modelled tRNA acceptor arm. This may occur by an allosteric switch upon adenylate formation. Attempts to diffuse buffered ATP/Gly mixtures into crystals at concentrations as low as 0.1 mM ATP caused severe cracking of the crystals within seconds, which is consistent with the above hypothesis since the insertion domain is involved in important crystal contacts. Side chains of residues in the motif 2 loop are in contact with residues in strand B7, and ordering of this loop on ATP binding/adenylate formation may constitute a trigger transmitted to the insertion domain.

A structural distinction between prokaryotic and eukaryotic GlyRS

The sequences of the three eukaryotic GlyRS can be aligned with ttGlyRS (Figure 1), but not that of ecGlyRS. The eukaryotic enzymes contain substantial insertions relative to ttGlyRS. Apart from extensions at the N- and C-termini, there are two major insertions, the first immediately after motif 2 (between B11 and H9) numbering some 40 residues, the second just before motif 3 (between B15 and B16) of ~65 residues. Structural modules inserted at these positions, especially the latter, would be well placed to contact the anticodon stem of a tRNA molecule bound to the neighbouring monomer of a ttGlyRS dimer (Figure 6B), which suggests a specific interaction at this level. However, inspection of the anticodon stem of all four $tRNA^{Gly}_{GCC}$s reveals no common feature shared by the three eukaryotic molecules and not by the one from *T.thermophilus*. This may imply that specificity is determined by indirect action of the additional domains through positioning of the sugar–phosphate backbone.

It has been suggested that the identity of the discriminator base 73 could be correlated with both subunit structure and the prokaryote/eukaryote division: $\alpha_2\beta_2$/U73 in the former, α_2/A73 in the latter (Shiba *et al.*, 1994). However, ttGlyRS is a dimer which recognizes, like ecGlyRS, a $tRNA^{Gly}$ containing U73, and there is sequence similarity in the insertion domains between dimeric GlyRS in prokaryotes and eukaryotes, suggesting that this generalization

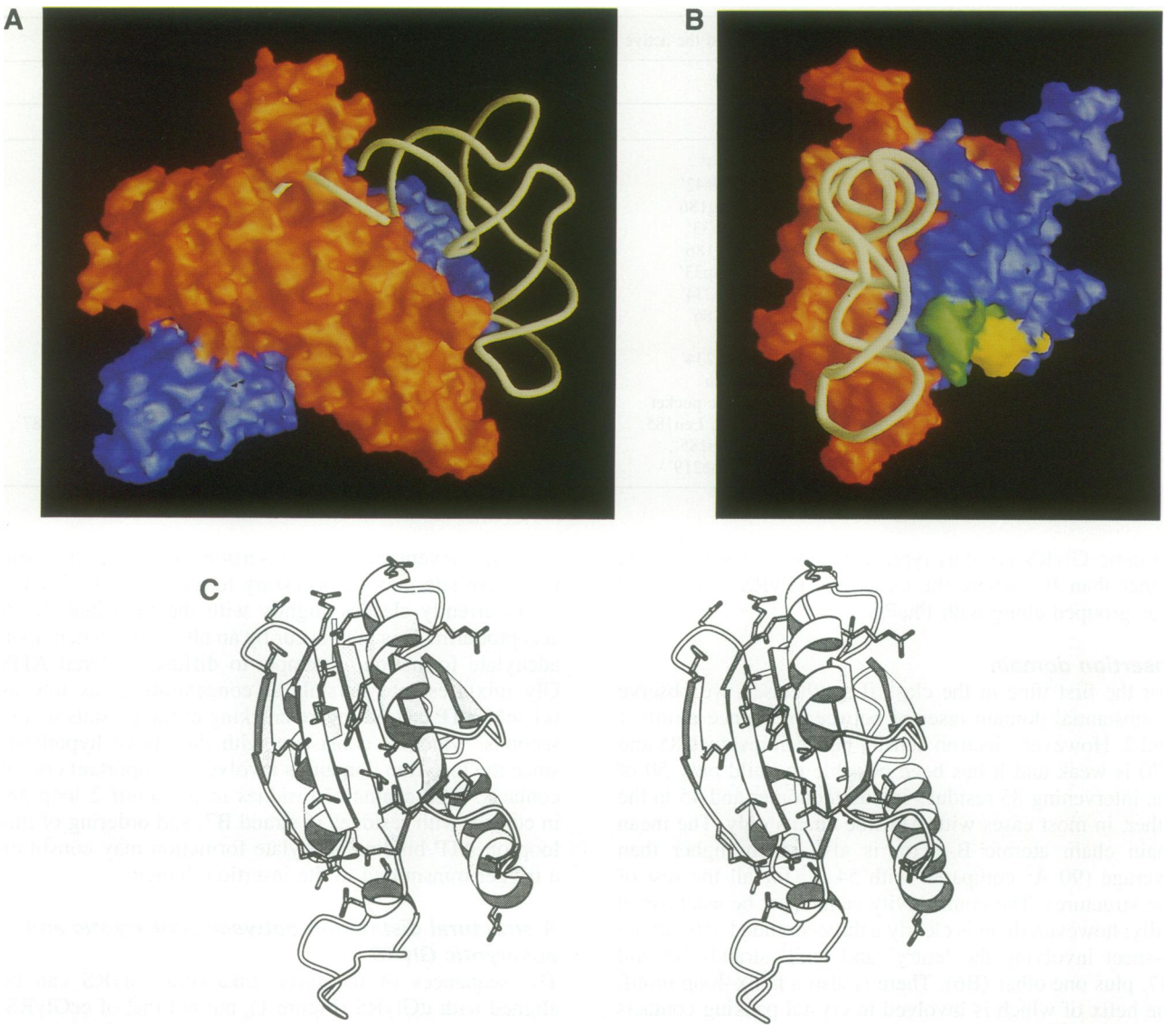


Fig. 6. (**A**) and (**B**) Model for $tRNA^{Gly}$ recognition by GlyRS. The view in (A) is essentially identical to that in Figure 3C. The molecular surface of monomer 1 is depicted in orange and that of monomer 2 in dark blue. One molecule of tRNA is shown as a white tube, interpolated through the phosphate backbone positions, bound to monomer 1 (the tRNA:dimer ratio is, however, presumably 2:1). The view in (B) is rotated by 90° around the dimer axis with respect to (A). The surface location of the eukaryotic insertions between B11 and H9 of monomer 2 is shown in yellow, that of the insertions between B15 and B16 in green. Displayed using GRASP. (**C**) Secondary structure cartoon of the C-terminal domain of ttGlyRS. Residues identical or conserved by functionality in all four sequences are shown in ball-and-stick representation. The anticodon seems most likely to bind on the surface of the β-sheet to the left of helix H12. Figures 3A, 4A and 5 were drawn using Molscript (Kraulis, 1991) and Raster3d (Merritt and Murphy, 1994) with modifications to Molscript by R.M.Esnouf. (C) was drawn with Molscript alone.

is not valid. The major remaining mystery is the relationship of ecGlyRS to the others: otherwise we could assume that our observations on tRNA binding were general to all class II synthetases.

Implications for class II synthetases

In the first partition into two classes, motif 1 and (to a lesser extent) motif 2 could not be found by sequence analysis in some dimeric aaRS. The structure of ttGlyRS confirms that motifs 1, 2 and 3 should be considered as an indissociable functional entity, presenting a global architectural solution to the active site of all dimeric class II synthetases. Motif 1 is thus not only part of the dimer interface but an essential element of the catalytic framework. The fact that the essential features of this architecture can be maintained in spite of a Pro→Ser mutation explains why sometimes motif 1 could not be detected: the absence of Pro, combined with already limited sequence homology, was lethal to a signature-oriented approach.

Why GlyRS appears unique in having chosen Ser and Thr to replace Pro is not obvious, though some evidence suggests that this may also occur elsewhere. The location of motif 1 in AlaRS is currently ambiguous. A recent secondary structure prediction for ecAlaRS (Ribas de Pouplana *et al.*, 1993) suggested a conserved Pro (in the sequence PTLL) in a loop following the motif 1 β-strand as the critical residue; however, it is too far from the end of the extra strand. The secondary structure prediction places a residue which is always Ser or Pro in AlaRS at the end of the appropriate β-strand, a more likely location for the critical residue, which also ensures a hydrophobic

Table III. Native and heavy atom derivative data used in the ttGlyRS structure determination

Data set	Native	Native 2 (+Gly)	PCMBS	TMLA	Sm	Xe
Reagent concentration		35 mM (Gly)	0.75 mM	24 mM	7.5 mM (pH 6.8)	30 atm
Soaking time		3 days	14 days	13 days	4 days	3 h
X-ray source	LURE W32	DESY X-11	LURE W32	lab.	lab.	LURE W32
No. of crystals	5	3	5	1	1	1
Processing	MOSFLM	MARXDS	MARXDS	MARXDS	MOSFLM	MARXDS
Resolution (Å)	2.9	2.75	3.2	3.3	3.4	5.0
No. of unique reflections	33 033	31 176 37 726[a]	23 654	22 999	17 571	6666
Completeness (%)	90.9	77.5, 83.7[a]	87.2	91.8	88.7	92.0
Redundancy	3.8	2.0	4.5	3.2	3.1	3.2
$R_{merge}(I)(\%)$	10.0	7.9	11.1	7.5	9.1	10.0
Mean FID			28.6	30.7	11.9	18.2
No. of sites in a.u.			2	2	1	4
R_{Cullis}			0.87	0.84	0.96	0.97
Phasing power						
centric			0.8	0.9	0.4	0.3
acentric			1.0	1.2	0.5	0.4

Set 'native 2' consists of data collected from crystals into which Gly had been soaked at 35 mM, extending to 2.75 Å. When no Gly was observed in (F_{Gly}–F_{native}) maps around the active site, these were treated as native data and missing reflections were filled in with observations from the suitably scaled 2.9 Å set.

[a]Refers to the composite data set.

The derivatives are: PCMBS, sodium parachloromercuriphenylsulfonate; Sm, $Sm_2(NO_3)_3$; TMLA, $(CH_3)_3Hg(CH_3CO_2)_2$; Xe, gaseous xenon. $R_{merge}(I)(\%) = \Sigma_{hkl}\Sigma_i|<I_{hkl}>-I_{hkl,i}|/\Sigma_{hkl}\Sigma_i I_{hkl,i}|$, where i are the observations of reflection hkl; a.u. = asymmetric unit, FID = fractional isomorphous difference. $R_{Cullis} = \Sigma_{hkl}||F_{PH} \pm F_P|-|F_{H,calc}|/\Sigma_{hkl}|F_{PH} \pm F_P|$; phasing power = $<F_H>/<\varepsilon>$, where $<F_H>$ is the r.m.s. calculated heavy atom structure factor and $<\varepsilon>$ is the r.m.s. lack of closure.

residue in the pocket. This raises the interesting possibility that some AlaRS use Pro and others Ser.

The active site of ttGlyRS suggests that strong interactions with the amino and carboxy moieties of Gly compensate energetically for side chain-specific interactions. The Gly recognition framework is conserved throughout prokaryotes and eukaryotes and differences between the sequences are related rather to aspects of tRNA recognition. The sequence of the anticodon recognition domain confirms that dimeric GlyRS belong to subclass IIa, and its structure provides a model for all such domains in this subclass.

Materials and methods

Sequencing

Genes of ttGlyRS were isolated from genomic DNA obtained from *T.thermophilus* HB8 cells. Protein was first isolated from this strain at sufficient purity to sequence the first 50 residues (M.-H.Mazauric *et al.*, in preparation). A specific DNA probe was then obtained by PCR amplification of the 5′ coding end of the gene using two degenerate oligonucleotides, sense and antisense, constructed from protein sequence information. This probe allowed the localization of the gene in a 4.0 kb fragment by Southern hybridization on a *Bam*HI digest of DNA. A minilibrary was then constructed in pUK18 containing a collection of fragments of this size; the fragment containing the ttGlyRS gene was purified by colony hybridization using the specific DNA probe. Both strands of the gene were sequenced using the dideoxynucleotide method (Sanger *et al.*, 1977) and primers distributed along the gene.

Data collection and phase determination

ttGlyRS was purified from HB8 cells and crystallized as described previously (Logan *et al.*, 1994). Crystals used for structure determination belong to space group $C222_1$, with unit cell dimensions a = 124 Å, b = 254 Å, c = 104 Å, $\alpha = \beta = \gamma = 90°$. There is a dimer in the asymmetric unit and a solvent content of 61%. The activity of the enzyme from redissolved crystals has been found to be similar to that of fresh protein (data not shown). Native and heavy atom data are summarized in Table III. For soaking, crystals were transferred to stabilizing solutions containing 1–2% more PEG 6000 than used for crystallization, in 0.1 M bis-Tris propane (BTP) buffer at pH 7.5, except for the Sm derivative which was insoluble at pH >7.0. Heavy atom concentrations were increased stepwise for the 'liquid phase' derivatives. Data were collected at 0°C on 18 cm MarResearch imaging plates using 1° or 0.6° rotations. The Xe derivative was prepared by diffusion of gaseous Xe into crystals mounted in quartz capillaries (Schiltz *et al.*, 1994) and a 'quick sweep' was carried out from one crystal using 1.5° images, at 9°C to avoid formation of Xe hydrate. Synchrotron data were collected at wiggler station W32 of the Laboratoire pour l'Utilisation du Rayonnement Synchrotron (LURE), Orsay, France (λ = 0.91 Å) and station X-11 of the EMBL outstation at the Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany (λ = 0.927 Å). Laboratory data were collected on Rigaku rotating anode sources with graphite monochromators. Data were reduced using either MOSFLM (A.G.W. Leslie, LMB, Cambridge, UK) or MARXDS (MarResearch, Hamburg, Germany) and further processed using programs from the CCP4 package (Collaborative Computational Project, 1994).

The Sm derivative, though phasing poorly, gave a very isomorphous difference Patterson which allowed the other derivatives to be solved by difference and residual Fourier techniques. Heavy atom parameters were refined using the CCP4 program MLPHARE. For the mercury derivative, anomalous data were collected from five crystals. The ratio of mean anomalous difference to mean intensity varied from 1.6 to 7.6% for data in the shell between 3.7 and 3.2 Å, for data above 1σ. Treating each crystal individually with respect to both isomorphous and anomalous differences resulted in strong bias towards essentially single isomorphous replacement phases from this derivative and overestimation of the figure of merit (FOM). To overcome this, but to avoid drowning the weak anomalous signal, one real occupancy was refined against merged isomorphous differences and five anomalous occupancies were refined against unmerged anomalous differences. The Xe derivative is rather poor, partially since data collection conditions were not optimized, but we are convinced that the sites are valid since they are found in buried pockets as observed elsewhere (Schiltz *et al.*, 1995).

Phase improvement, extension and model building

Phases obtained from MLPHARE at 4 Å were used to calculate an MIR map which was modified by solvent flattening. Phases from back transformation of this map were used to refine further the heavy atom parameters (Cura *et al.*, 1992), which resulted in changes averaging ~15% in occupancies and B-factors. This process was carried out twice, at which point refinement had converged and the mean FOM for acentric reflections between 12 and 3.2 Å had improved from 0.38 to 0.41. All solvent flattening and averaging was carried out using GAP (J.G.Grimes and D.I.Stuart) and CCP4 programs. Solvent and molecular envelopes

were generated using GAP, and optimized using MAMA (T.A.Jones and G.J.Kleijwegt).

The solvent-flattened (SF) MIR map calculated at 3.5 Å using optimized heavy atom parameters allowed a model for the two active site regions of a dimer of scAspRS to be placed in the density. The NCS operator was determined by rigid body superposition of the monomers, and refined against the SF MIR map using a spherical envelope at the dimer centre. With this operator, one cycle of envelope-free averaging was carried out to enhance the molecular envelope, which was then defined using the scAspRS model with spheres to cover the remaining density. Seventy five percent of the model could be traced as polyalanine using O (Jones *et al.*, 1991), with the aid of skeletonized electron density, in the map obtained after seven cycles of averaging and solvent flattening using this mask and recombination with MIR phases at 3.5 Å using SIGMAA (Read, 1986). An improved envelope based on this trace was used for a phase extension to 3.3 Å, which allowed 50% of the side chains to be constructed. Model phases were then combined with MIR phases at 3.3 Å, and the resolution extended to 3.1 Å. At this point, 90% of main and side chain were built. MIR phases were then abandoned and calculated phases only used to 2.9 Å.

Model refinement

The model was refined using X-PLOR (Brünger *et al.*, 1987), by energy minimization and two rounds of slow cooling, interspersed with manual model rebuilding. NCS restraints were applied in the final stages of refinement using the weight which produced the optimal reduction in free R-factor (Brünger, 1992). Initial refinement was carried out using the 2.9 Å data set described previously (Logan *et al.*, 1994) but, at a later stage, set 'native 2' extending to 2.75 Å was substituted. The current model consists of 472/506 residues in monomer 1 and 467/506 in monomer 2. There are also 23 well-ordered water molecules, but no model for bulk solvent. This gives a total of 7547 non-hydrogen atoms for which the R-factor is 21.8%, R_{free} = 31.0% for all 36 365 reflections between 8 and 2.75 Å (21.3 and 30.7% respectively to 2.9 Å). R_{free} is calculated using 8% of the reflections, randomly chosen, which are not used in refinement. The model is of good quality, scoring more than 1σ better than average for every stereochemical parameter tested by the program PROCHECK (Laskowski *et al.*, 1993). The r.m.s. deviations from the ideal geometry values of Engh and Huber (1991) are 0.014 Å for bonds, 1.9° for angles and 25.0° for dihedrals. The mean thermal B-factors are rather high: 47 $Å^2$ for main chain atoms and 51 $Å^2$ for side chain atoms, even when five mobile loops in each monomer and the entire 'insertion' domain are not included. This is reflected in the rapid decay of reflection intensities beyond 3 Å. However, electron density is generally of very good quality (Figure 2).

Modelling of GlyRS–Gly-AMP interactions

A best fit superposition was found between the main chain, O and Cβ atoms of eight residues in interaction with aspartyl-adenylate and their strictly conserved equivalents in ttGlyRS. The r.m.s. deviation was 0.85 Å for 41 atoms superimposed. This rotation was then applied to the coordinates of Asp-AMP to place it in the active site of ttGlyRS, and the Asp side chain moiety was removed. A minimal amount of manual readjustment (almost entirely rigid body movement) was then carried out to optimize hydrogen bonds with the protein.

Acknowledgements

We wish to thank V.Cura for early crystallization and data processing, A.Mitschler for invaluable assistance with data collection, J.-C.Thierry for much help and encouragement, J.Reinbolt for help with N-terminal sequencing, R.Giegé for continuous support of the biochemistry and critical reading of the manuscript, R.Kreutzer (Laboratorium für Biochimie, Bayreuth) for biochemical help, staff at EMBL Hamburg and LURE (M.Schiltz), M.M.G.M. Thunnissen (Groningen) for a gift of TMLA and for help with Figure 5B, J.G.Grimes and D.I.Stuart (Oxford) for supplying and helping with GAP, R.M.Esnouf (Oxford) for extensions to Molscript and to colleagues for discussion at many times during this project. D.T.L. was supported by CNRS and EEC grants, M.H.M. by an MRE grant. The project benefited from CNRS and Université Louis Pasteur (Strasbourg) support.

References

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 413–410.

Arnez,J.G., Harris,D.C., Mitschler,A., Rees,B., Francklyn,C.S. and Moras,D. (1995) *EMBO. J.*, **14**, 4143–4155.

Artymiuk,P.J., Rice,D.W., Poirrette,A.R. and Willet,P. (1994) A tale of two synthetases. *Nature Struct. Biol.*, **1**, 758–760.

Belrhali,H. *et al.* (1994) Crystal structures at 2.5 Ångstrom resolution of seryl-tRNA synthetase complexed with two analogs of seryl adenylate. *Science*, **263**, 1432–1436.

Belrhali,H., Yaremchuk,A., Tukalo,M., Berthet-Colominas,C., Rasmussen,B., Bösecke,P., Diat,O. and Cusack,S. (1995) The structural basis for seryl-adenylate and AP_4A synthesis by seryl-tRNA synthetase. *Structure*, **3**, 341–352.

Biou,V., Yaremchuk,A., Tukalo,M. and Cusack,S. (1994) The 2.9Å crystal structure of *T.thermophilus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$. *Science*, **263**, 1404–1410.

Brick,P., Bhat,T.N. and Blow,D.M. (1989) Structure of tyrosyl-tRNA synthetase refined at 2.3Å resolution. Interaction of the enzyme with tyrosyl adenylate intermediate. *J. Mol. Biol.*, **208**, 83–98.

Brunie,S., Zelwer,C. and Risler,J.-L. (1990) Crystallographic study at 2.5Å resolution of the interaction of methionyl-tRNA synthetase from *Escherichia coli* with ATP. *J. Mol. Biol.*, **216**, 411–424.

Brünger,A. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.

Brünger,A.T., Kuryian,J. and Karplus,M. (1987) Crystallographic R-factor refinement by molecular dynamics. *Science*, **235**, 458–460.

Cavarelli,J., Rees,B., Ruff,M., Thierry,J.-C. and Moras,D. (1993) Yeast $tRNA^{Asp}$ recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature*, **362**, 181–184.

Cavarelli,J. *et al.* (1994) The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. *EMBO J.*, **13**, 327–337.

Collaborative Computational Project No. 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Cura,V., Krishnaswamy,S. and Podjarny,A.D. (1992) Heavy-atom refinement against solvent-flattened phases. *Acta Crystallogr.*, **A48**, 756–764.

Cusack,S. (1993) Sequence, structure and evolutionary relationships between class 2 aminoacyl-tRNA synthetases: an update. *Biochimie*, **75**, 1077–1081.

Cusack,S., Berthet-Colominas,C., Härtlein,M., Nassar,N. and Leberman,R. (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase. *Nature*, **347**, 249–255.

Cusack,S., Härtlein,M. and Leberman,R. (1991) Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. *Nucleic Acids Res.*, **19**, 3489–3498.

Delarue,M. and Moras,D. (1993) The aminoacyl-tRNA synthetase family: modules at work. *Bioessays*, **15**, 675–687.

Delarue,M., Poterszman,A., Nikonov,S., Garber,M., Moras,D. and Thierry,J.-C. (1994) Crystal structure of a prokaryotic aspartyl-tRNA synthetase. *EMBO J.*, **13**, 3219–3229.

Doublié,S., Bricogne,G., Gilmore,C. and Carter,C.W.J. (1995) Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure*, **3**, 17–31.

Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.*, **A47**, 392–400.

Eriani,G., Delarue,M., Poch,O., Gangloff,J. and Moras,D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, **347**, 203–206.

Eriani,G., Cavarelli,J., Martin,F., Dirheimer,G., Moras,D. and Gangloff,J. (1993) Role of dimerization in yeast aspartyl-tRNA synthetase and importance of the class II invariant proline. *Proc. Natl Acad. Sci. USA*, **90**, 10816–10820.

Francklyn,C., Shi,J.-P. and Schimmel,P. (1992) Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices. *Science*, **255**, 1121–1125.

Fraser,T.H. and Rich,A. (1975) Amino acids are not all initially attached to the same position on transfer RNA molecules. *Proc. Natl Acad. Sci. USA*, **72**, 3044–3048.

Gatti,D.L and Tzagoloff,A. (1991) Structure of a group of related aminoacyl-tRNA synthetases. *J. Mol. Biol.*, **218**, 557–568.

Ge,G., Trieu,E.P. and Targoff,I.N. (1994) Primary structure and functional expression of human glycyl-tRNA synthetase, an autoantigen in myositis. *J. Biol. Chem.*, **269**, 28790–28797.

Imada,K., Sato,M.,Tanaka,N., Katsube,Y., Matsuura,Y. and Oshima,T. (1991) Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of *Thermus thermophilus* at 2.2Å resolution. *J. Mol. Biol.*, **222**, 725–738.

Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.

Kern,D., Giegé,R. and Ebel,J.-P. (1981) Glycyl-tRNA synthetase from baker's yeast. Interconversion between active and inactive forms of the enzyme. *Biochemistry*, **20**, 122–131.

Kraulis,P.J. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.

Logan,D.T., Cura,V., Touzel,J.-P., Kern,D. and Moras,D. (1994) Crystallisation of the glycyl-tRNA synthetase from *Thermus thermophilus* and initial crystallographic data. *J. Mol. Biol.*, **241**, 732–735.

McClain,W., Foss,K., Jenkins,R.A. and Schneider,J. (1991) Rapid determination of nucleotides that define $tRNA^{Gly}$ acceptor identity. *Proc. Natl Acad. Sci. USA*, **88**, 6147–6151.

Merritt,E.A. and Murphy,M.E.P. (1994) Raster3D version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr.*, **D50**, 869–873.

Moras,D. (1992) Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem. Sci.*, **17**, 159–164.

Mosyak,L. and Safro,M. (1993) Phenylalanyl-tRNA synthetase from *Thermus thermophilus* has four antiparallel folds of which only two are catalytically functional. *Biochimie*, **75**, 1091–1098.

Murzin,A.G. (1993) OB (oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.*, **12**, 861–867.

Nada,S., Chang,P.K. and Dignam,J.D. (1993) Primary structure of the gene for glycyl-tRNA synthetase from *Bombyx mori*. *J. Biol. Chem.*, **268**, 7660–7667.

Namba,K. and Stubbs,G. (1986) Structure of tobacco mosaic virus at 3.6Å resolution: implications for assembly. *Science*, **231**, 1401–1406.

Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–286.

Onesti,S., Miller,A.D. and Brick,P. (1995) The crystal structure of the lysyl-tRNA synthetase (LysU) from *Escherichia coli*. *Structure*, **3**, 163–176.

Ostrem,D.L. and Berg,P. (1974) Glycyl transfer ribonucleic acid synthetase from *Escherichia coli;* purification, properties and substrate binding. *Biochemistry*, **13**, 1338–1348.

Poterszman,A., Delarue,M., Thierry,J.-C. and Moras,D. (1994) Synthesis and recognition of aspartyl-adenylate by *Thermus thermophilus* aspartyl-tRNA synthetase. *J. Mol. Biol.*, **244**, 158–167.

Read,R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr.*, **A42**, 140–149.

Ribas de Pouplana,L., Buechter,D.D., Davis,M.W. and Schimmel,P. (1993) Idiographic representation of conserved domain of a class II tRNA synthetase of unknown structure. *Protein Sci.*, **2**, 2259–2262.

Rould,M.A., Perona,J.J., Söll,D. and Steitz,T.A. (1989) Structure of *E.coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8Å resolution. *Science*, **246**, 1135–1142.

Ruff,M., Krishnaswamy,S., Boeglin,M., Poterszman,A., Mitschler,A., Podjarny,A., Rees,B., Thierry,J.-C. and Moras,D. (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. *Science*, **252**, 1682–1689.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Schiltz,M., Prangé,T. and Fourme,R. (1994) On the preparation and X-ray data collection of isomorphous xenon derivatives. *J. Appl. Crystallogr.*, **27**, 950–960.

Shiltz,M., Fourme,R., Broutin,I. and Prangé,T. (1995) The catalytic site of serine proteinases as a specific binding cavity for xenon. *Structure*, **3**, 309–316.

Schimmel,P. (1987) Aminoacyl-tRNA synthetases: general scheme of structure–function relationships in the polypeptides and recognition of transfer RNAs. *Annu. Rev. Biochem.*, **56**, 125–158.

Shiba,K., Schimmel,P., Motegi,H. and Noda,T. (1994) Human glycyl-tRNA synthetase. Wide divergence of primary structure from bacterial counterpart and species-specific aminoacylation. *J. Biol. Chem.*, **269**, 30049–30055.

Steinberg,S., Misch,A. and Sprinzl,M. (1993) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **21**, 3011–3015.

Surgochov,A.P. and Surgochova,I.G. (1975) Two enzymically active forms of glycyl-tRNA synthetase from *Bacillus brevis*. *Eur. J. Biochem.*, **54**, 175–184.

Wierenga,R.K., De Maeyer,M.C.H. and Hol,W.G.H. (1985) Interaction of pyrophosphate moieties with α-helices in dinucleotide binding proteins. *Biochemistry*, **23**, 1346–1357.

Yan,W. and Francklyn,C. (1994) Cytosine 73 is a discriminator nucleotide *in vivo* for histidyl-tRNA in *Escherichia coli*. *J. Biol. Chem.*, **269**, 10022–10027.

Received on April 13, 1995; revised on May 22, 1995