Nucleic Acids Research, Vol. 19, No. 13 3489–3498

Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases

Stephen Cusack, Michael Härtlein and Reuben Leberman
European Molecular Biology Laboratory, Grenoble Outstation, c/o ILL, 156X 38042 Grenoble, France

Received March 12, 1991; Revised and Accepted May 24, 1991

ABSTRACT

Class 2 aminoacyl-tRNA synthetases, which include the enzymes for alanine, aspartic acid, asparagine, glycine, histidine, lysine, phenylalanine, proline, serine and threonine, are characterised by three distinct sequence motifs 1,2 and 3 (reference 1). The structural and evolutionary relatedness of these ten enzymes are examined using alignments of primary sequences from prokaryotic and eukaryotic sources and the known three dimensional structure of seryl-tRNA synthetase from *E. coli*. It is shown that motif 1 forms part of the dimer interface of seryl-tRNA synthetase and motifs 2 and 3 part of the putative active site. It is further shown that the seven α_2 dimeric synthetases can be subdivided into class 2a (proline, threonine, histidine and serine) and class 2b (aspartic acid, asparagine and lysine), each subclass sharing several important characteristic sequence motifs in addition to those characteristic of class 2 enzymes in general. The $\alpha_2\beta_2$ tetrameric enzymes (for glycine and phenylalanine) show certain special features in common as well as some of the class 2b motifs. In the alanyl-tRNA synthetase only motif 3 and possibly motif 2 can be identified. The sequence alignments suggest that the catalytic domain of other class 2 synthetases should resemble the antiparallel domain found in seryl-tRNA synthetase. Predictions are made about the sequence location of certain important helices and β-strands in this domain as well as suggestions concerning which residues are important in ATP and amino acid binding. Strong homologies are found in the N-terminal extensions of class 2b synthetases and in the C-terminal extensions of class 2a synthetases suggesting that these putative tRNA binding domains have been added at a later stage in evolution to the catalytic domain.

INTRODUCTION

The aminoacyl-tRNA synthetases play a crucial role in protein biosynthesis by specifically charging tRNAs with their cognate amino acids. Primary structures of enzymes for all twenty amino acids are now available and recent publications (1–3) have demonstrated conclusively that these enzymes may be partitioned into two quite distinct groups, each containing ten members. One group, designated class 1, contains the enzymes for cysteine, methionine, valine, isoleucine, leucine, arginine, glutamine, glutamic acid, tyrosine and tryptophan and is characterised by two primary structural motifs, H-I-G-H and K-M-S-K-S (3–5). High resolution X-ray structures of three class 1 synthetases are known, those of the tyrosyl-, methionyl- and glutaminyl-enzymes (6–8), the last in a complex with its cognate tRNA. These all contain a Rossmann nucleotide-binding fold (9) in the catalytic domain comprising parallel β-strands and connecting helices, with both the H-I-G-H and K-M-S-K-S motifs close to the ATP binding site.

The second group is characterised by three quite different sequence motifs 1, 2 and 3 which have been identified in the aminoacyl-tRNA synthetases for aspartic acid, asparagine, histidine, lysine, phenylalanine, proline, serine and threonine (see Table 2 and reference 1). The alanyl- and glycyl-enzymes apparently contain only motif 3 (1). The only X-ray crystal structure of a class 2 synthetase that has been determined is that of seryl-tRNA synthetase from *Escherichia coli* (SRSEC*) which is an α_2 dimeric enzyme, each subunit having 430 residues (2). This structure has a C-terminal catalytic domain of 330 residues based around a seven-stranded antiparallel β-sheet (Figure 1) which is quite different from the Rossmann nucleotide binding fold. The N-terminal 100 residues form a remarkable solvent-exposed arm comprising a 60Å long antiparallel coiled-coil. The cave-like active site is lined by a section of the β-sheet and rimmed by loops (Figure 2); in particularly there is a disordered loop, part of motif 2, which appears to form a flexible flap over the active site. The two universally conserved arginines in motifs 2 and 3 and second conserved basic residue (R/H) in motif 2 all protrude into the active site cave suggesting their extreme functional importance. In this enzyme motif 1 forms part of the dimer interface and this is consistent with the fact that all class 2 synthetases that possess motif 1 are α_2 dimers. It is suggested below that phenylalanyl-tRNA synthetase which is in fact an $\alpha_2\beta_2$ tetramer does not actually possess motif 1 as originally proposed in reference 1.

These observations lead to the proposition that the major structural difference between the two classes of aminoacyl-tRNA synthetases is that the catalytic domain of class 1 enzymes is based on a Rossmann fold whereas that of class 2 enzymes is based on an anti-parallel fold. Furthermore it has been proposed (1) that this difference might reflect a difference in enzymatic

*The aminoacyl-tRNA synthetase for amino acid X from organism Y.z. is referred to as XRSYZ.

reaction; class 1 synthetases preferentially aminoacylate the 2′ OH of the ribose of the tRNA 3′ terminal adenosine and class 2 synthetases the 3′ OH, although there are exceptions to this rule (10,11). On the other hand, judging by the four known crystal structures, the tRNA binding domains of aminoacyl-tRNA synthetases seem to be more idiosyncratic for each enzyme.

The purpose of this paper is to examine in more detail the sequence alignments of class 2 aminoacyl-tRNA synthetases in the light of the known three-dimensional structure of seryl-tRNA synthetase with the aim of (a) defining the extent of structural homology betweeen class 2 synthetases and hence the nature of the new anti-parallel ATP-binding fold (b) locating functional domains and active site residues (c) permitting transfer of structure/function relationships from one synthetase to another and (d) defining the evolutionary relationship between the class 2 synthetases.

The results indicate the importance of careful visual, multiple sequence alignment of other synthetases with the SRSEC sequence/secondary-structure after first comparing the same synthetase from different organisms. This first step enables important conserved motifs to be located and also indicates regions where gaps and deletions are permissible. The synthetases are clearly a favourable case for this strategy because of the overall low homology, even inter-species for the same amino acid. However the fact that the synthetases are enzymes with similar activities and are presumably evolutionary related, allows one to infer with more confidence structural and functional homology from what may appear to be statistically marginal sequence homologies, provided that there is also topological equivalence.

An alignment of all known aspartyl-tRNA synthetases and of the aspartyl-, asparaginyl- and lysyl-tRNA synthetases from *E. coli* has been given previously (12), and of the seryl-, threonyl- and prolyl-enzymes from *E. coli* (1), but in the absence of any structural interpretation. These alignments differ in several details from those presented here.

METHODS

As a starting point, the alignments given in reference 1 were re-examined in the light of the known three-dimensional structure of SRSEC shown diagrammatically in Figure 1.

The crucial observation was that despite the low overall sequence homology between seryl-tRNA synthetase from *E. coli* (SRSEC) and from *Saccharomyces cerevisiae* (SRSSC) (29% identity i.e. 126/430 residues in SRSEC), when the SRSSC sequence is placed in the SRSEC structure, of the 23 mostly hydrophilic residues which line the active site (Figure 2 and marked * in Table 2), nineteen are identical and three have conservative changes. Apart from those in motif 2 (β-strand A6, loop L2 and β-strand A7) and motif 3 (β-strand A5 and the helix H12), this highlighted the importance of fully conserved motifs in β-strands A4 (S-C-S-N) and A3 (Y-D-L-E) as well as conserved residues in the loops between strands βA2 and βA3, between helix H8 and strand βB3 and between strand βB2 and helix H9 (see Figure 1 and Table 2).

This led to the hypothesis that analagous active-site elements might be identified in other class 2 synthetases as corresponding to conserved motifs between the prokaryotic and eukaryotic enzymes. Amino acid sequences of prokaryotic and eukaryotic class 2 synthetases (except for proline and glycine for which only the *E. coli* enzyme sequences are known) were therefore aligned with the aid of the program BESTFIT and visual inspection. Multiple alignments were then made by visual inspection, starting from the well-defined motifs 1, 2 and 3 of reference 1 and taking into account the known structure of SRSEC. Location of

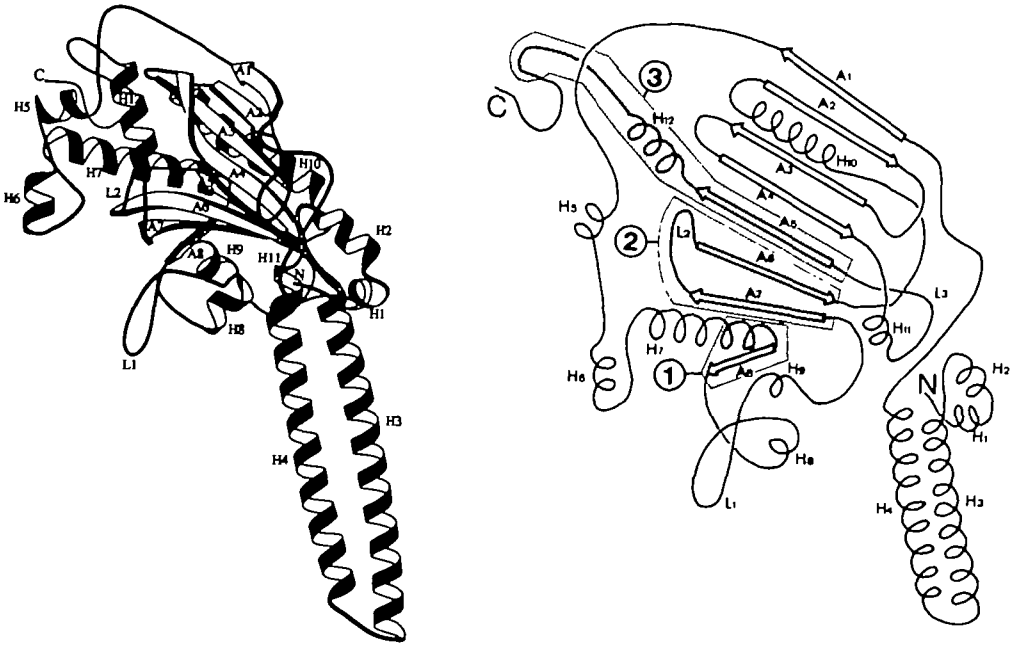


Figure 1. Left: Ribbon diagram of the tertiary fold of one subunit of seryl-tRNA synthetase from *E. coli*, showing the major secondary structural elements, β-strands bA1 to βA8 and a-helices H1 to H12. (see also reference 2). Figure prepared with program RIBBON (J. Priestle). Right: Simplified diagram of the structure of seryl-tRNA synthetase from *E. coli* (based on left hand figure). The location of the three sequence motifs, characteristic of class 2 aminoacyl-tRNA synthetases and defined in reference 1, are indicated. In this paper it is proposed that at least strands βA3 and βA4 and helices H7 and H10 are also common to the seven dimeric class 2 synthetases. Figure drawn by E. DiCapua.

interspecies conserved motifs preceeding motif 3 enabled assignments of putative β-strands A4 and A3 in all class 2 synthetases except alanyl-tRNA synthetase. Further examination showed that other secondary structural features occurring in SRSEC (especially helices H7 and H10) are also probably conserved in all dimeric class 2 synthetases.

For simplicity, not all known aminoacyl-tRNA synthetase sequences are presented in Table 2. Many excluded are highly homologous to those in the table e.g. the recently determined sequences of two threonyl-tRNA synthetases from *Bacillus subtilis* (13) are homologous to the *E. coli* enzyme; the histidyl-enzyme from hamster is very homologous to the human enzyme (14) and the two distinct lysyl-tRNA synthetases in *E. coli* (encoded by the genes *lysS* and *lysU*) are 88% identical (15). On the otherhand, in the case of the histidyl- and aspartyl-tRNA synthetases, three sequences are included, as in each case, the *E. coli* enzyme shows lower than usual homology with the yeast enzyme (see discussion). Yeast mitochondrial synthetase sequences are known for the threonyl-(16), histidyl-(17), aspartyl-(18), phenylalanyl-(19) and lysyl-tRNA synthetases (20). These sequences often exhibit additional insertions and deletions as well as overall lower identity compared to cytoplasmic or bacterial sequences for the same amino acid, sometimes making alignment less certain. They are not included here for simplicity (except in Tables 3a and 3b), although it has been verified that motifs claimed to be important for a particular synthetase are also identifiable in the mitochondrial synthetase sequence. The eukaryotic alanyl-tRNA synthetase sequence is from *Bombyx mori* (ARSBM, reference 21), which is a monomer whereas the *E. coli* enzyme is a homo-tetramer. In the case of asparaginyl-tRNA synthetase, the eukaryotic sequence used is that of an antigenic protein (22) from the nematode *Brugia malayi* (designated ANTBM). This protein has not been shown to be an asparaginyl-tRNA synthetase, but has high homology with the *E.coli* asparaginyl-enzyme (reference 1) except that the published sequence for ANTBM shows no apparent motif 2. However an alternative reading frame over residues 315−330 (marked ANTBM*) in Table 2, would give a perfect motif 2. We suggest that there maybe a frameshift sequencing error but this has not been confirmed.

RESULTS

The results of the sequence alignments are shown in Table 1 which summarises the secondary structural elements predicted to be found in the different class 2 synthetases with reference to that of seryl-tRNA synthetase (SRSEC). The results suggest that class 2 synthetases can be subdivided into class 2a comprising the enzymes for serine, threonine, proline and histidine which must have structures very similar to the known 3-D fold of SRSEC, and class 2b comprising those for aspartic acid, asparagine and lysine, which have fewer structural features in common with SRSEC. Phenylalanyl-tRNA synthetases (β-subunits) and glycyl-tRNA synthetases (α-subunits) are related to class 2b. The inclusion of the histidyl-tRNA synthetases in class 2a follows from the details of the alignment, the similar sequence topology particularly compared to PRSEC and the considerable homology in the C-terminal extension with the prolyl- and threonyl-enzymes (see below and Table 3a).

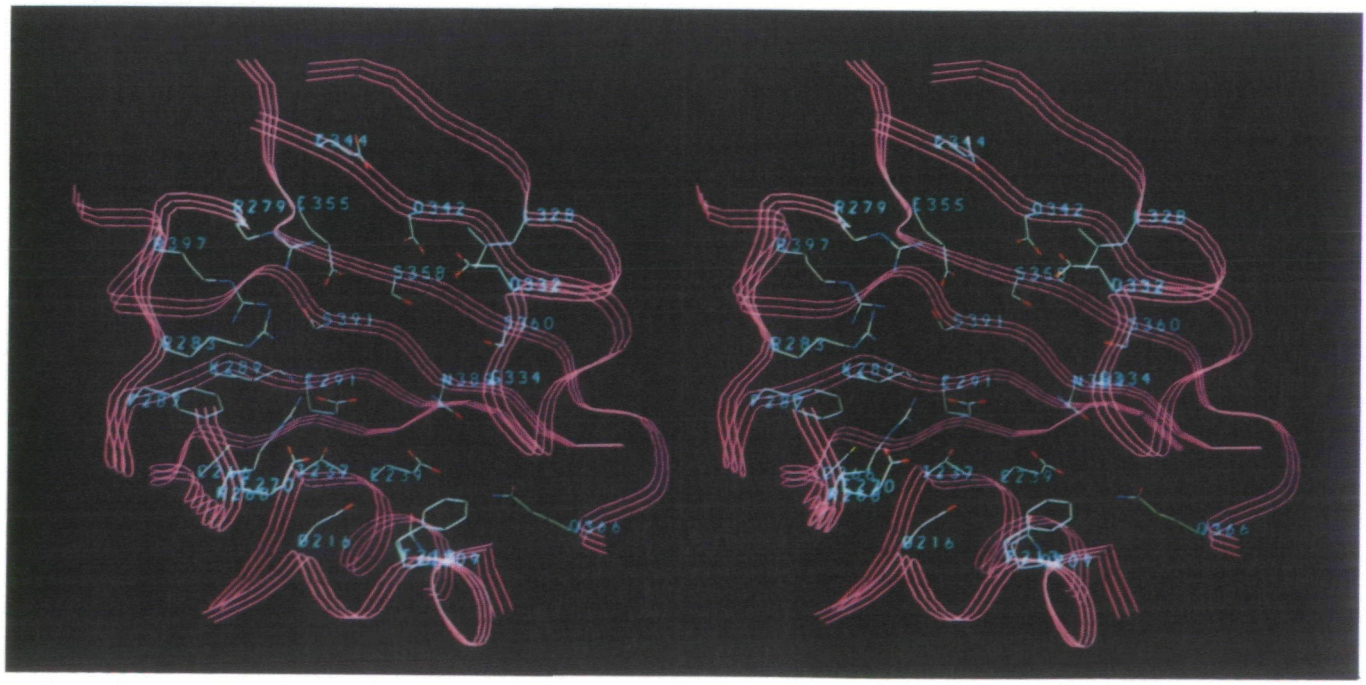


Figure 2. Stereo view of part of the three-dimesional structure of seryl-tRNA synthetase from *E. coli*., showing all residues exposed in the putative active site (marked by a * in Table 2). These residues are strongly conserved between the *E. coli* and *S. cerevisiae* enzymes. View direction is similar to that of Figure 1. Ribbons indicate the polypeptide backbone. Six of the seven antiparallel strands are shown starting with strand βA2 at the top right. Sequentially down then comes strand βA3 (with asp-342 and glu-344); strand βA4 (glu-355, ser-358, ser-360); strand βA5 (asn-389, ser-391) and helix H12 (arg-397) which form motif 3; motif 2 comprising strand βA6 (phe-287, lys-289, glu-281) and strand βA7 (cys-266) with the intervening loop L2 (arg-268, glu-270, arg-279), disordered between residues 271−279; at the bottom are regions near helix H9 (thr-237, glu-339) and after helix H8 (gln-209, phe-213, asp-216) Figure prepared with program FRODO (P. Evans).

Table 1. Predicted secondary structural elements of class 2 aminoacyl-tRNA synthetases with reference to those of seryl-tRNA synthetase from *E. coli*. Each element is scored with a + (certain or almost certain), a ? (possible) or a − (probably not or cannot be identified). The table is based on the sequence alignments in Table 2.

	N-ext	βA1	Motif 1 (H7,βA8)	Sheet B (LoopL1)	H9	Motif 2 (βA7,loop-L2,βA6)	H10	βA2	Insert	βA3	βA4	Motif 3 (βA5,H12)	C-ext
Class 2a													
Ser	+	+	+	+	+	+	+	+	−	+	+	+	short
Thr	+	?	+	+	+	+	+	+	short	+	+	+	long
Pro	−	−	+	+	+	+	+	+	long	+	+	+	long
His	−	−	+	?	+	+	+	+	long	+	+	+	long
Class 2b													
Asp*Ec*	+	?	+	−	?	+	+	?	long	+	+	+	medium
Asp*Sc*	+	?	+	−	?	+	+	?	short	+	+	+	short
Asn	+	?	+	−	?	+	+	?	short	+	+	+	short
Lys	+	?	+	−	?	+	+	?	long	+	+	+	short
Pheβ	−	−	−	−	−	+	−	−	short	+	+	+	short
Glyα	−	−	−	−	−	−	−	−	−	+	+	+	long
Ala	−	−	−	−	−	?	−	−	−	−	−	+	long

The detailed sequence comparisons with reference to the secondary structure of SRSEC, upon which Table 1 is based, are presented in Table 2. This table should not be read to imply sequence or structural homology with SRSEC throughout the alignments; certain places where there is clearly no homology are indicated (see caption to Table 2). Important features of the alignment will now be discussed sequentially with particular emphasis on those regions where the structural homology with SRSEC appears sure and those where it is uncertain. The following symbols are used for certain amino acid types: small residues λ = {C, S, T, P, A, G}, hydrophobic residues φ = {M, I, L, V, F, W}.

Motif 1-the dimer interface

The alignments show that much of the dimer interface including the interface helix H7 and strand βA8 and possibly strand βA9*, can be found in all members of class 2a and 2b. The amphiphilic interface helix is characterised by conserved patterns of hydrophobic and basic residues (the latter most strongly in class 2b) and terminates with a strongly conserved glycine (followered by a tyrosine in class 2a and a phenylalanine in class 2b). Motif 1 should thus be extended from its original definition to include all of the interface helix. This makes it unlikely that the phenylalanyl-tRNA synthetases β-subunits actually possess motif 1 as originally proposed (reference 1); the interface helix terminated by a glycine is certainly not present. This may be related to the $\alpha_2\beta_2$ quartenary structure of the phenylalanyl-tRNA synthetase whereas the other members of class 2a and 2b are all α_2 dimers.

The region of strand βB2 and helix H9 contains two residues, Thr-237 and Glu-239 of the putative active site of SRSEC conserved in the yeast seryl-enzyme and in PRSEC and HRSEC. In general, both class 2a and class 2b synthetases have specific conserved sequence motifs in this region (for example Q-S-P-Q, S-F-L-T and R-I-A-P-E-L in the aspartyl-, asparaginyl- and lysyl-enzymes respectively). The lack of strong homology means that the nature of the secondary structure cannot be predicted for class 2b. However the alignment of the active site residues T-X-E found in several class 2a synthetases with the residues S-X-Q in two class 2b synthetases is suggestive. It is doubtful whether the phenylalanyl-tRNA synthetases fit with class 2b in this region.

Motif 2

Motif 2, which includes strand βA7, the variable active site loop and strand βA6 and is clearly present in all members of class 2a and 2b as well as the β-subunit of phenylalanyl-tRNA synthetases. We have also identified for the first time a putative motif 2 in the alanyl-tRNA synthetases, but not in the glycyl-tRNA synthetase. In SRSEC the variable loop, given its position and the fact that it is disordered in the absence of substrates, maybe involved in interaction with and recognition of the tRNA acceptor stem, known to contain important seryl-identity elements. There are however clear differences between class 2a and 2b. The former tend to have longer variable loops and characteristically different motifs after the loop. These differences in detail reinforce the assignments of histidyl-tRNA synthetases to class 2a and phenylalanyl-tRNA synthetases to class 2b, the latter being most closely homologous to the lysyl-tRNA synthetases.

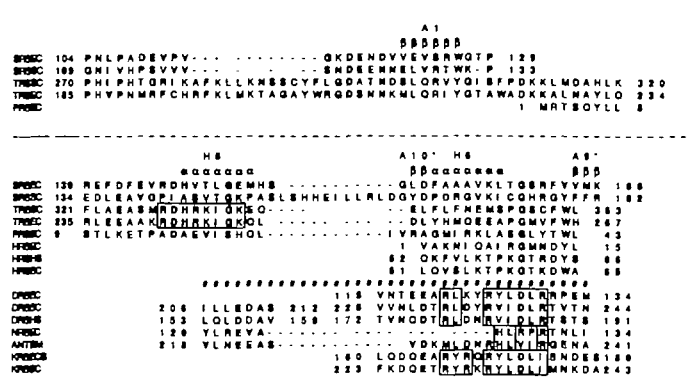
Crossover helix H10

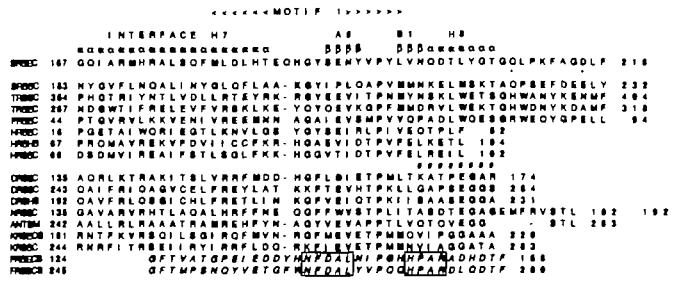
The amphiphilic crossover helix H10, which connects strands βA6 and βA2 in SRSEC, is apparently conserved in class 2a synthetases with a characteristic heptad repeat of hydrophobic residues and terminating with a strongly conserved φ-G-φ. The occurrence of an analagous helix immediately following motif 2 is also probable in class 2b synthetases (see table 2) although with the hydrophobic residues differently phased to class 2a synthetases.

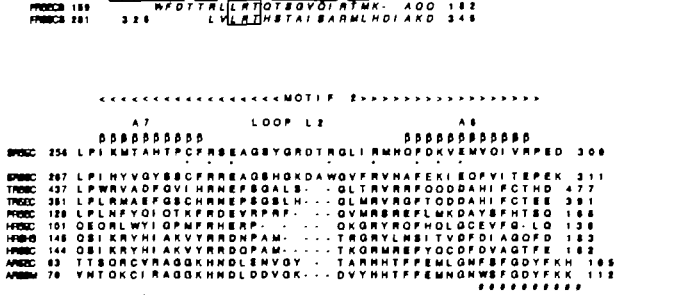
Insertion between strands βA2 and βA3

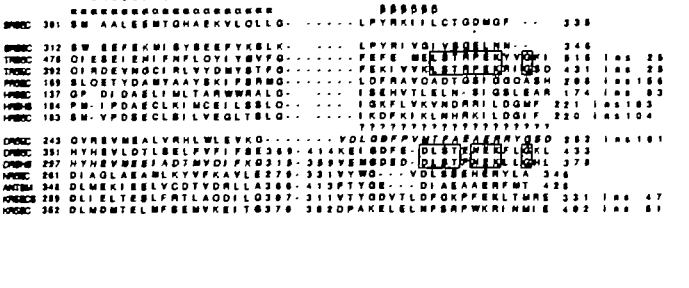
In SRSEC the crossover helix H10 and the strands βA2, βA3 and βA4 succeed each other separated by only short loops. This is consistent with the fact that the seryl-tRNA synthetases have the minimal distance of about 80 residues (apart from the phenylalanyl-tRNA synthetases) between motifs 2 and 3 (reference 1). Other class 2 synthetases have sequences of very variable length (e.g. up to 175 residues in the case of PRSEC) and low homology between the putative crossover helix and strands βA3. It is thus difficult to identify strand βA2 in class 2b synthetases. However there is a motif in most class 2b synthetases D-φ-λ-X-X-X-E-K/R-X-φ-λ which is strikingly similar to a conserved motif in the threonyl-tRNA synthetases.

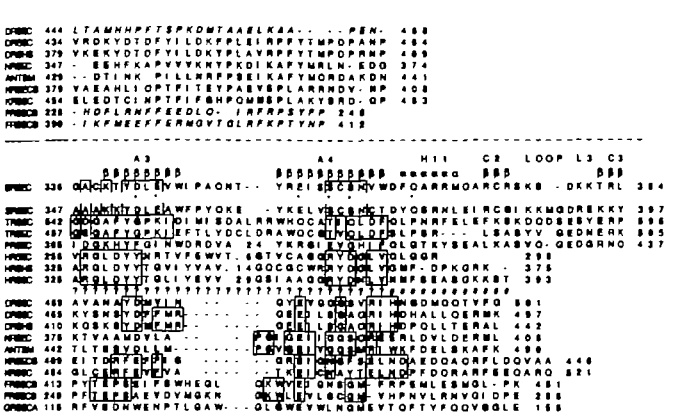
Table 2. Alignment of sequences from class 2 synthetases with reference to the secondary structure of seryl-tRNA synthetase from *E. coli*.

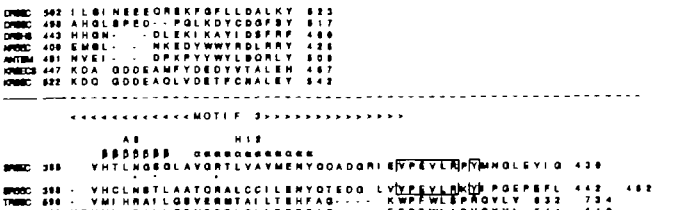


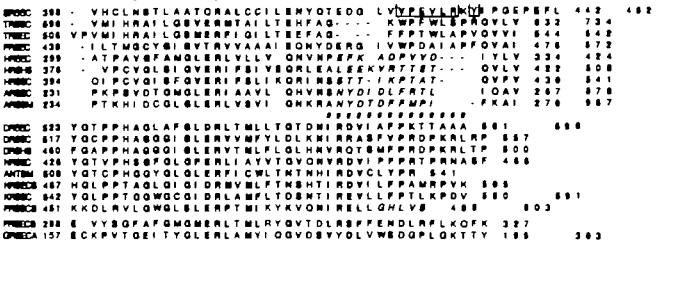












The aminoacyl-tRNA synthetase for amino acid X from organism *Y.z.* is referred to as XRSYZ. Abbreviations are as follows: EC: *Escherichia coli*, ARSBM: alanyl-tRNA synthetase from *Bombyx mori*, SC: *Saccharomyces Cerevisiae*, HS: *Homo sapiens*, ANTBM: antigen from *Brugia malayi* homologous to asparaginyl-tRNA synthetase, ANTBM*: alternative reading frame for ANTBM (see text). Note that the sequence for the human histidyl-tRNA synthetase (HRSHS) has been corrected from that published where there is a frameshift sequencing error in the motif 3 region (Dr. Tsui, personal communication).

The following symbols are used:

H5:	designation of secondary structural element in SRSEC (Figure 1 and Reference 2).
α:	α-helix
β:	β-strand
*:	residue exposed in the putative active site of SRSEC (in row between SRSEC and SRSSC).
Bold:	residues are given bold-type if they are conserved (according to modified Dayhoff class, see below) in at least three synthetases in class 2a or two synthetases in class 2b.
Boxes:	used to draw attention to motifs conserved between the same synthetase from different organisms and referred to in the text.
# # # # #:	indicates absence of apparent homology between class 2a and class 2b.
????:	indicates possible homology between class 2a and class 2b.
Italics:	indicates absence of apparent homology for a particular synthetase.

Modified Dayhoff classes means the following classification:

1,2= C, S, T, P, A, G = λ
3= N, D, E, Q
4= H, R, K
5= M, I, L, V, F, W = ϕ
6= F, Y, W

References for aminoacyl-tRNA synthetase sequences used are given below.

Abbreviation	Reference
ANTBM (antigen from *Brugia malayi*)	22
ARSEC (EC: *Escherichia coli*):	36
ARSBM (BM: *Bombyx mori*):	21
DRSEC:	12
DRSSC (SC: *Saccharomyces Cerevisiae*):	37
DRSSCM:(M: mitochondrial)	18
DRSHS (HS: *Homo sapiens*):	38
FRSEC:	39
FRSSC:	40
FRSSCM:	19
GRSEC:	41
HRSEC:	42
HRSSC:	17
HRSSCM:	17
HRSHS:	14
KRSECS: constitutive *lysS* gene product	15
KRSSC:	28
KRSSCM:	20
NRSEC:	43
PRSEC.	1
SRSEC:	32
SRSSC:	44
TRSBS:	13
TRSEC:	45
TRSSC:	46
TRSSCM:	16

This may be used to give a possible alignment of the class 2a and class 2b in this region. Just preceeding putative strand βA3, there is a well conserved motif found in the asparaginyl- and eukaryotic aspartyl-tRNA synthetases (F/Y-P-X-D/E-I/V-R/K-P/A-F-Y) part of which is possibly also present in the lysyl- and phenylalynyl-tRNA synthetases.

Strands βA3 and βA4

In SRSEC, strands βA3 and βA4 contain several residues (conserved in SRSSC) that are exposed in the putative active site. Similarly, groups of synthetase specific conserved motifs in fairly close proximity are found in all other class 2a and 2b synthetases (including the phenylalanyl- and glycyl-tRNA synthetases) which are also tentatively assigned to strands βA3 and βA4. These regions are characterised by the frequent occurrence of small residues (G, A, C, S, T). Within class 2a, the assignment of strand βA3 is largely determined by the presence of a strongly conserved λ-X-λ-X-X-Y motif and that of strand βA4 by the conserved features at the end of strand βA4 and beginning of helix H11 (which contains a conserved basic residue). In SRSEC the conserved Tyr-340 in strand βA3 burrows through the protein *away* from the active site as does Lys-338. These two residues hydrogen bond with Glu-306 (also conserved between SRSEC and SRSSC) in the crossover helix H10. PRSEC may share this feature.

In class 2b, strand βA3 is tentatively aligned on a conserved acidic residue and strand βA4 has the consensus E-φ-φ-λ-λ. The *E.coli* glycyl- and phenylalanyl-tRNA synthetases are clearly homologous in this region. The relative alignment between class 2a and class 2b is uncertain. The loop between putative strands βA3 and βA4 is rather variable even between the same synthetase from different organisms.

The structural motif comprising strands bC2 and bC3 and the intervening loop which is solvent exposed and poorly ordered in SRSEC is apparently conserved in the threonyl-tRNA synthetases and PRSEC, but not in histidyl-tRNA synthetases. In class 2b there is a corresponding but non-homologous insertion of about 20 residues in the aspartyl- and asparaginyl- and lysyl-tRNA synthetases, but not in the phenylalanyl-tRNA synthetases.

Motif 3 and Helix H12

Motif 3, comprising strand βA5 and helix H12, is found in all ten class 2 synthetases and here again there are characteristic differences between the subclasses. The first part of motif 3 is much more glycine and proline-rich in class 2b (consensus sequence Y-G-X-P-P-H-λ-G-φ-G-φ-G) than 2a. Class 2b has a highly conserved methionine three residues after the universally conserved arginine in helix H12 and a consensus motif N-φ-R-D/E-φ-X-X-F/Y-P-R/K after helix H12. Class 2a has different motifs at the end of helix H12 followed by φ-P-X-φ-φ-X-P in the seryl-, threonyl- and prolyl-enzymes. These distinctions suggest that GRSEC (α-subunit) has a motif 3 derived from class 2b and that the alanyl-tRNA synthetases are most closely related to the histidyl-tRNA synthetases in class 2a.

In SRSEC, helix H12 is almost completely buried, hence its largely hydrophobic nature (conserved in other synthetases) apart from two long side chains with charged groups that can reach the protein surface (in particular the active site residue arg-397). The motifs after helix H12 (E-N-Y-Q and V-P-E-V-L-R-X-Y) are well away from the active site and the presumed tRNA binding domain. Their conservation in the seryl-tRNA synthetases therefore may imply functional interactions with another molecule. In this connection we note an intriguing homology in this region between SRSEC and the gene product of *Selb*, two proteins that both involved in the co-translational incorporation of selenocysteine into certain enzymes. The *Selb* gene product is the particular elongation factor that specifically recognises selenocysteinyl-$tRNA^{SelCys}$ which is produced by charging $tRNA^{SelCys}$ with serine by SRSEC and subsequent conversion of the bound serine to selenocysteine (23,24). The *Selb* gene product contains in the presumed nucleotide binding domain (by analogy with the elongation factor, EF-Tu) the sequence 129-V-K-E-V-L-R-P-Y which has 6/8 residues in common with the second C-terminal motif found in SRSEC and SRSSC.

The alignments described above reveals certain motifs conserved in at least two of the three members of class 2b but absent in class 2a. This includes the motifs after helix H12 mentioned above as well as the consensus motif E/D-λ-R-L/Y-X-X-R-X-φ-D-L-R just before the interface helix H7; E-λ-λ-λ just after motif 1; D-φ-λ-X-X-X-E-K/R-X-φ after the crossover helix H10 and F/Y-P-X-D/E-I/V-R/K-P/A-F-Y before motif 3. These are likely to imply conserved structural and/or functional features.

DISCUSSION

The sequence comparisons of the putative catalytic domain of class 2 aminoacyl-tRNA synthetases presented above demonstrates that the α_2 dimeric enzymes can be subdivided into a closely related sub-class 2a comprising the seryl-, threonyl-, prolyl- and histidyl-tRNA synthetases and a second sub-class 2b comprising the aspartyl-, asparaginyl- and lysyl-tRNA synthetases. This subdivision is further confirmed on consideration of the N- and C-terminal extensions to the catalytic domain (see below). The tetrameric class 2 synthetases can also be assigned on the basis of more limited homology, alanyl-tRNA synthetase is related to class 2a and the phenylalanyl-tRNA synthetases (β-subunits) and GRSEC (α-subunit) to class 2b.

Table 3. (3a) Alignment of C-terminal extensions to the putative catalytic domain in class 2a synthetases.

```
TRSSCM 348  YQAVII#CTALQKKLRNEL#VPLN------DWHFNVDLDIRNEPVGYRIK    408
TRSSC  628  RQVLVVPVGVKYQGYAEDVRNKLH------DAGFYADVDLTGNTLQKKVR-   671
TRSEC  540  VQVVIMNITDSQSEYVNELTQKLS------NAGIRVKADLRNEKIGFKIR-   583
PRSEC  474  FQVAILPMNMHKSFRVQELAEKLYSEL--RAQGIEVLLDDRKERPGV-MFA   521
HRSEC  330  DIYLVASGADTQSAAMA-LAERLRDEL----PGVKLMTN-HGGGNFKKQFA   374
HRSHS  418  TQVLVASASAQKKLAGGETKACLR----LWDAGIKAELLYKKNPKLLNQLQ   453
HRSSC  426  TQVFVMA-FGGGKDWTGYLPERMKVTKQLWDAGIEAEYVYKAKANPRKQFD   476

TRSSCM 409  SAILKNYSYLIIVGDEEVQLQKYNIRERDNRKSFEKLTMSQIWEKFIELEKNYK  462
TRSSC  672  NGQMLKYNFIFIVGEQEMNEKSVNIRNRDVMEQQGK-NATVSVEEVLKQLRN  722.734
TRSEC  584  EHTLRRVPYMLVCGDKEVESGKVA[VRTR]RGKDLGSM-DVNEVIEKLQQEIRS  634.644
PRSEC  522  DMELIGIPHTIVLGDRNLDNDDIEYKYRRNGEKQLI-KTGDIVEYLVKQIKG  572
HRSEC  375  RADKWGARVAVVLGESEVANGTAVVKDLRSGEQTAV-AQDSVAAHLRTLLG   424
HRSHS  454  YCEETGIPLVAIIGEQELKDGVIKLRSVASREEVDV-RREDLVEEIRRRTNQ  503.508
HRSSC  476  TTKKAGCHIAVILGKEEYLEGKLRVKRLGQEFADDD-GELVSAADIVPIVQE  526.541
```

Shaded arginines 583 and 612 in TRSEC indicate sites where mutagenesis causes loss of autoregulatory activity and reduction of affinity for tRNA (30). Boxed motif 608-V-R-T-R-R in TRSEC is found also in N-terminal extension of NRSEC (residues 26–30, see Table 3b and text).

(3b) Alignment of N-terminal extensions to the putative catalytic domain in class 2b synthetases.

```
DRSSCM 44   LSS-GQKIVLNGWIEQKPKRVGKNLIFGLLRDSNGD--IIQLVDNKSLLKG   91
DRSEC  11   LSHVGQQVTLCGWV-NRRRDLGS-LIFIDMRDREGI--VQVFFDPDRADAL   57
DRSSC  102  AKDSDKEVLFRARV-HNTRQQGATLAFLTLRQQASL--IQGLVKANKEGTI  149
DRSHS  52   IQKADEVVWVRARV-[HTSR]AKGKQ-CFLVLRQQQFN--VQALVAVGDHASK   98
NRSEC  13   RVAVDSEVTVRGW[V-RTRR]DSKAGISFLAVYDGSCFDPVQAVINNSLPNYN   62
ANTBM  120  VKHRNERVCIKGWI-HRMRRQGKSLMFFILRDGTGF--LQVLLMDKLCQTY  167
KRSECS 61   LEALNIEVAVAGRM-MTRRIMGKA-SFVTLQDVGGR--IQLYVARDDLPEG  107
KRSSC  117  ETLPEEKVSIAGRI-HAKRESGSKLKFYVLHGDGVE--VQLMSQLQDYCDP  164
KRSSCM 94   EDNPNLLLSINGRI-KSIRFSGQKIVFIDLYNGSS 7 LQLIVNYNKIGGS  147
```

See caption to Table 3a and text.

Table 1 summarises the secondary structural elements predicted to be found in the different synthetases with reference to that of SRSEC. The three other members of class 2a must have very high structural homology to SRSEC. For instance the threonyl-tRNA synthetases are predicted to possess 7/7 of the antiparallel β-strands found in SRSEC and PRSEC and the histidyl-tRNA synthetases 6/7. All class 2a members are predicted to possess the same major secondary structural elements of the dimer interface and the crossover helix H10. Class 2b members would appear to have more limited structural homology with SRSEC, it being only possible to identify 5/7 antiparallel β-strands, the crossover helix and the dimer interface. Of course as is often found to be the case, the structural homology may well exceed the apparent sequence homology. The β-subunit of the phenylalanyl-tRNA synthetases are also predicted to possess 5/7 β-strands but not the dimer interface. There is a convincing homology between the small (a) subunit of GRSEC and the small (b) subunit of FRSEC in the region of putative strands βA3 and βA4 in addition to motif 3. In the case of the alanyl-tRNA synthetases, despite the presence of a number of conserved motifs between the enzymes from *E. coli* and *Bombyx mori* (21), only motif 3 and possibly motif 2 can be identified (although the otherwise universally conserved arginine in motif 2 becomes a histidine, see Table 2). The clear absence of a canonical motif 1 preceeding this putative motif 2 in alanyl-tRNA synthetases is consistent with the fact that the Bombyx mori enzyme is a monomer, and the oligomeristion domain of the E. coli enzyme is known to be in the C-terminal region. Further analysis of the glycyl-tRNA synthetases will require sequences from other organisms than just *E. coli*.

(a) Active site. The arguments given above suggest strongly that all class 2 synthetases have a closely similar active site topology to SRSEC. Apart from the universally conserved residues in motifs 2 and 3 (which are found on strands βA5, βA6 and βA7) it can be predicted that there are also important active site residues on the two strands βA3 and βA4. As these are not conserved between synthetases for different amino acids but are conserved between the same synthetase from different organisms, these may be involved in amino acid binding specificity. This suggestion is supported by the geometry of the active site (Figure 2). This is such that all universally conserved residues from motif 2 and motif 3 (e.g. Arg-268, Glu-270, Arg-283, Phe-287, Glu-291, Arg-397) are concentrated on the left side of the putative site (viewed as in Figure 2). These residues might then form a conserved binding site for the ATP and 3′ end of the tRNA with the significant concentration of positively charged residues in the vicinity (in addition to those mentioned above, SRSEC also has Arg-279 and Lys-289) being involved in phosphate interactions and in the catalytic mechanism (25) . More to the right of the putative active site are the synthetase specific conserved residues of strands βA5, βA4 and βA3 which may form the amino acid binding site. Interestingly there are no basic residues on this side of the active site in SRSEC, but many with hydrogen bonding capability. These observations lead one to expect the aminoacyl-adenylate to lie across the active site from left to right, with the

adenosine to the extreme left and the amino acid to the right. One piece of evidence lend supports to this proposition. A tightly bound detergent molecule from the crystallization medium is found at the left of the active site (2,26). The rather well defined shape of the electron density leads to the suggestion that the molecule might be a derivative of hexyl-ribofuranoside (the detergent used in the crystallization is known to be an highly impure preparation of octyl-glucoside, reference 2). The putative hexyl-chain of this molecule is buried in the protein to the left of the active site and the sugar moiety is between helix H12 and the strand 281−284. This corresponds closely to the proposed position for the adenine and ribose moietes of the ATP.

We note that site-directed mutagenesis of Tyr-426 in NRSEC which is located at the beginning of motif 3 has shown that this residue has an important role in determining the kinetic parameters for ATP binding (27).

(b) Dimer interface. The dimer interface can be identified in all seven class 2 synthetases that are α_2 dimers. As described for SRSEC (2), this involves short β-strands A8 and A9* which extend the major antiparallel β-sheet across the dimer interface as well as the 'interface' helix H7. In SRSEC, helix H7 is perpendicular to the dimer axis and interacts with its antiparallel symmetry related partner with a relatively large interhelical separation of 12.5Å. The interactions are hydrophilic, via complementary charged residues and water molecules. The frequent occurrence of long hydrophilic and/or charged residues (glutamine, glutamate, arginine and lysine) in the putative interface helices would suggest that very similar interactions occur in the other α_2 synthetases.

The strong conservation of the dimer interface between all dimeric class 2 synthetases suggests that there maybe important functional interactions between the two subunits or that there may be cross-subunit tRNA binding. In this respect we note that the partially disordered loop L1 in the SRSEC structure crosses over from one monomer towards the active site of the second monomer and forms and inter-monomer β-sheet around the two-fold axis (2). The sequence alignments suggest that this structure is conserved in class 2a synthetases.

(c) tRNA binding domains. Putative tRNA binding and/or recognition domains can be identified as insertions within or N- or C-terminal extensions to the main framework of the catalytic domain or, in the case of the phenylalanyl- and glycyl-tRNA synthetases, additional subunits. In addition, eukaryotic synthetases often have extra often highly basic extensions compared with the corresponding prokaryotic sequences; it has been speculated that these promote compartmentalization of synthetases (28).

In SRSEC the 60Å long N-terminal antiparallel coiled-coil domain (residues 1−100) is presumed to be involved in tRNA binding (2). It is interesting to speculate whether any other synthetase could have such a domain as SRSEC. The two very long a-helices H3 and H4 in SRSEC are characterised by 4−3 repeats of hydrophobic residues (including arginine, alanine and threonine) and a very high proportion (39%) of charged residues. An alignment of SRSEC and TRSEC shows that significant homology exists (1) in the region of the first long helix H3 and possibly for the return helix H4 although in TRSEC the latter would be slightly shorter and kinked due to the occurrence of a proline (P-170) in the middle. The 4−3 pattern of hydrophobic residues (including arginine, lysine, alanine and threonine) is however largely conserved as well as the high content of charged residues (45% in TRSEC).

In class 2a, PRSEC and the histidyl-tRNA synthetases do not possess an N-terminal extension, but instead have substantial inserts between strands βA2 and βA3. There is no obvious homology between the insertions of these two synthetases. They also have a C-terminal extension of about 100 residues as do the threonyl-tRNA synthetases. Table 3a shows that there are several strongly conserved regions amongst the three class 2a synthetases with C-terminal extensions (threonyl-, prolyl- and histidyl-enzymes). This is further support for the inclusion of the histidyl-tRNA synthetases in class 2a. All these extra domains may be involved with tRNA binding and in the case of threonyl-tRNA synthetases, with translational auto-regulation (29−31). Two arginines (Arg-583 and Arg-612), whose mutation causes a significant decrease in the affinity for tRNA and a loss of autoregulation activity of *E.coli* threonyl-tRNA synthetase are shown shaded in Table 3a.

There is a remarkable homology in the N-terminal putative tRNA binding region of all three class 2b synthetases (Table 3b). Over a region of about 30 residues, there are five fully conserved residues and a significant concentration of basic residues. Class 2b synthetases show very variable length insertions between the crossover helix H10 and putative strand βA3 and only short or non-existant C-terminal extensions to the catalytic domain.

There are two occurrences of cross sub-class homologies in the putative tRNA binding regions, which however should be treated with caution as they occur in topologically inequivalent positions. The first is a five residue motif V-R-T-R-R found at the C-terminal of TRSEC and the N-terminal of NRSEC, both in conserved regions (boxed in Tables 3a and 3b). Secondly, there is nine residue identical peptide found in the middle of the long helix H4 in SRSEC as well as in an N-terminal alanine-rich region of FRSECβ (32). We recall also that in the catalytic domain the class 2b motif after the crossover helix (D-ϕ-λ-X-X-X-E-K/R-X-ϕ-l) is also found in the threonyl-tRNA synthetases (boxed in Table 2).

(d) Evolutionary relatedness. It is interesting to consider the sequence relatedness of the class 2 synthetases, since this might throw light on their evolution and more generally on that of the genetic code. However it is difficult to evaluate this quantitatively as it depends on the extent of sequence alignment included and on the species origin of the enzymes. If we restrict attention to the *E. coli* synthetases and quantify the number of amino acid identities or conservative changes as a percentage of the number of residues aligned for each pair of synthetases in Table 2 (i.e. over the putative catalytic domain), the following results are obtained. Each entry gives on the first line the percentage identity and the number of residues compared (%/Number) and on the second line, the score obtained using the Dayhoff substitution matrix (33) normalised to the number of residues compared. The latter assesses conservative changes according to classes 1=C, 2=S,T,P,A,G 3=N,D,E,Q 4=H,R,K 5=M,I,L,V 6=F,Y,W. The diagonal elements in italics give the comparisons between corresponding synthetases from *E.coli* and *S. cerevisiae* (ANTBM in the case of asparaginyl-tRNA synthetase).

Class 2a

	SRSEC	TRSEC	PRSEC	HRSEC	HRSSC
SRSEC	*34.2/319*	15.6/307	18.2/286	11.6/225	14.7/232
	1.69	0.38	0.67	−0.23	−0.07
TRSEC	–	*48.8/338*	20.0/290	14.7/224	–
		2.80	0.80	0.07	
PRSEC	–	–	–	16.4/226	–
				0.31	
HRSEC	–	–	–	–	*28.5/235*
					1.25

Class 2b

	DRSEC	KRSECS	NRSEC	ANTBM
DRSEC	*29.1/278*	28.9/280	26.8/261	–
	1.31	1.47	1.15	
KRSECS	–	*53.7/298*	21.5/265	–
		2.97	1 045	
NRSEC	–	–	–	*34.3/297*
				1.86
DRSHS	–	–	28.5/298	28.2/298
			1.29	1.48

These results show that for *E. coli*, in class 2a, PRSEC is the closest relative to all the three other subclass members, the prolyl- and threonyl-enzymes being the most closely related. HRSEC and SRSEC are the most distantly related pair; indeed on purely statistical arguments they would probably be classed as unrelated were it not for the fact they are both synthetases with certain motifs in common. It is interesting to note that there is an apparent closer homology between SRSEC and HRSSC. Within class 2b, the homologies are generally higher, with DRSEC being the closest relative of both NRSEC and KRSEC.

Quantitation of the alignments between synthetases from class 2a and class 2b have not been made. This is because the cross-subclass alignments are only significant over a more limited range including mainly motifs 1, 2 and 3 which would then give an artificially high apparent homology. In any case the values need to be regarded with caution as the extent of the region of overlap differs considerably (e.g. 286–307 residues when comparing TRSEC, SRSEC and PRSEC, about 225 when comparing HRSEC with other members of class 2a and 260–300 when comparing members of class 2b).

It is also interesting to note the significant differences between extent of homology between eukaryotic and prokaryotic synthetases. The lysyl- and threonyl-tRNA synthetases are most conserved (respectively 54 and 49% identity) and the histidyl- and aspartyl-tRNA synthetases are the least conserved (both about 29% identity). Furthermore, results for cross-organism comparison can give higher homology than within the same organism. For example SRSEC is closer to HRSSC, than to HRSEC. Similar there is extensive homology and similarity in topology over more than 300 residues between the asparaginyl-tRNA synthetases and the eukaryotic aspartyl-tRNA synthetases; DRSEC has a much larger insertion between the crossover helix and strand βA3. These observations suggest that the *E. coli* enzymes for histidine and aspartic acid have evolved relatively quickly.

There are three observations which suggest a relationship between *E. coli* glycyl- and phenylalanyl-tRNA synthetases. Firstly, both are $\alpha_2\beta_2$ tetramers with a close correspondence between subunit size (glycine: a 303 residues, b 689; phenylalanine: a 795, b 307). Secondly there is a significant homology between GRSECα and FRSECβ over about 100 residues including apparently important residues in putative strands βA3 and βA4 and motif 3. Thirdly, there is a small region of homology between GRSECα and FRSECβ, which would probably not be considered significant was it not equidistant from the N-terminus in both sequences:

```
FRSECβ   39 TLQMTTLRELPPEERPAAGA
                | | || | ||   ||
GRSECα   40 SHPMTCLRALGPEPMAAAYV
```

It is however puzzling that a motif 2 cannot be identified in GRSECα.

CONCLUSION

The alignments presented in this paper show that class 2 synthetases can be divided into class 2a, with in *E. coli*, the prolyl-enzyme as the representative member (being most closely related to each of the three other members, seryl-, threonyl- and histidyl-) and class 2b, with the aspartyl-enzyme as the representative member (being most closely related to both the asparaginyl- and lysyl-enzymes). Alanyl-tRNA synthetase is a putative member of class 2a (based on motif 3 alone), and phenylalanyl- and glycyl-tRNA synthetases form a related pair associated with class 2b. The strongest evidence for classification of histidyl-tRNA synthetase as a member of class 2a is homologies in the C-terminal extension to the catalytic domain (Table 3a). Although details in the catalytic domain also suggest this classification (as mentioned in the results), it could also be argued that the histidyl-enzyme is intermediate between class 2a and 2b.

It is not yet clear whether this sub-classification or indeed the partition of all synthetases into two major classes has any relationship with the evolution of the genetic code. However it is noteworthy that, with the exception of phenylalanine, the members of each sub-class all occupy contiguous boxes in the standard representation of the genetic code. In particular all amino acids with codons of the form XCX are members of class 2a (including alanine). On the other hand the classes of synthetases only partially correlate with the amino acid families postulated by the co-evolution theory of the genetic code (34), which is based on amino acid biosynthetic pathways. According to this hypothesis there are two major amino acid families, both containing class 1 and class 2 members, one comprising aspartic acid, lysine, asparagine, threonine, isoleucine and methionine and the second comprising glutamic acid, glutamine, arginine, proline and histidine. However a recent publication (20) has revealed an interesting sequence homology (particularly in the region of motif 3) between the ammonia-dependent asparagine synthetase and aspartyl-tRNA synthetase which lends some support to the co-evolution theory. Both enzymes activate aspartic acid with ATP to form an aminoacyl-adenylate intermediate.

The alignments in Table 2 suggest that the putative tRNA binding domains have been added to the primordial class 2 catalytic domain either as C or N terminal extensions or insertions between the crossover helix H10 and strand βA3. Class 2b members have evolved from a common N-terminal extension (Table 3b) and class 2a members (except seryl-tRNA synthetase) from a common C-terminal extension (Table 3a). In addition seryl- and threonyl-tRNA synthetases have an N-terminal extension with features in common. These observations are in accord with the hypothesis that early class 2 aminoacyl-tRNA synthetases comprised a common catalytic domain with the ability to bind ATP, an amino acid and the 3′ end of early tRNAs. With the expansion of the genetic code and the evolution of tRNAs, synthetases acquired more specificity first by improved discrimination at the level of the tRNA acceptor stem and subsequently by addition of extra tRNA-binding domains capable of, for example, specific anticodon recognition.

A scenario for the evolution of modern protein aminoacyl-tRNA synthetases from precursors in the RNA world has been recently proposed in which it was hypothesised that the first synthetases would be those for basic amino acids (35). Such speculation is clearly imprudent before the relationships between the modern aminoacyl-tRNA synthetases are clearly established. For one thing, the sequence analysis in reference 1 and its extension presented here show that the aminoacyl-tRNA

synthetases do not exhibit such a 'puzzling diversity' as previously thought. However, without further structural information to confirm or refute the hypotheses of this paper and in particular to clarify the relationship of the three tetrameric synthetases to the seven dimeric ones, it is still too early to propose any kind of evolutionary tree for the class 2 synthetases.

Finally it is interesting to remark that the ATP binding domain, based on an antiparallel sheet, found in seryl-tRNA synthetase and by presumption in all other class 2 synthetases has not yet been found in any other nucleotide binding protein. On the otherhand the Rossmann fold characteristic of class 1 synthetases is widely found in other enzymes (9). An important clue in the understanding of the partition of the synthetases into two classes (i.e. whether they co-evolved or which one came first?) may come when other ATP-binding proteins with an antiparallel fold are discovered.

ACKNOWLEDGEMENTS

We thank Dino Moras, Marc Delarue and their colleagues at the IBMC in Strasbourg for drawing our attention to the sequence homology between the asparaginyl-tRNA synthetase from *E. coli* and the antigen from the nematode *Brugia malayi* (21) and for a preliminary glimpse of the partial three-dimensional crystal structure of the complex between aspartyl-tRNA synthetase/$tRNA^{Asp}$ from *S. cerevisiae*. We also thank Nicolas Nassar for pointing out some of the sequence homologies and Elisabeth DiCapua for useful suggestions regarding the manuscript and figures.

REFERENCES

1. Eriani, G., Delarue, M., Poch, O., Gangloff, J. and Moras, D., (1990a), *Nature* **347**, 203–206.
2. Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N. and Leberman,R., (1990),*Nature* **347**, 249–255.
3. Eriani,G., Dirheimer, G. and Gangloff, J. (1991) *Nucleic Acids Res.* **19** 265–269.
4. Burbaum, J.J., Starzyk, R.M. and Schimmel, P., (1990), *Proteins* **7**, 99–111.
5. Hountondji, C., Dessen, P. and Blanquet, S., (1986), *Biochimie*, **68**, 1071–1078.
6. Brick,P., Bhat, T.N. and Blow, D.M., (1989), *J.Mol.Biol.*, **208**, 83–98.
7. Brunie, S., Zelwer, C. and Risler, J-L., (1990), *J. Mol. Biol.* **216**, 411–424.
8. Rould, M.A., Perona, J.J., Söll, D. and Steitz, T.A., (1989),*Science*, **246**, 1135–1142.
9. Rossmann, M.G., Liljas, A., Branden, C-V. and Banaszak, L.J. (1975) In Boyer, P. D. (ed.) The Enzymes. Academic Press, New York. Vol.11a, p 61.
10. Fraser, T. H. and Rich, A., (1975), *Proc. natn. Acad. Sci. U.S.A*, **72**, 3044–3048.
11. Sprinzl, M. and Cramer, F. (1975), *Proc. natn. Acad. Sci. U.S.A*, **72**, 3049–3053.
12. Eriani,G., Dirheimer, G. and Gangloff, J. (1990b) *Nucleic Acids Res.* **18** 7109–7117.
13. Putzer, H., Brakhage, A.A. and Grunberg-Manago, M., (1990), *J.Bacteriol.*, **172**, 4593–4602.
14. Tsui, F.W.L. and Siminovitch, L., (1987), *Nucleic Acids Res.*, **15**, 3349–3367.
15. Lévêque, F., Plateau, P., Dessen, P. and Blanquet, S. (1990), *Nucleic Acids Res.* **18** 305–312.
16. Pape, L.K., Koerner, T.J. and Tzagoloff, (1985) *J.Biol.Chem.* **260**, 15362–15370.
17. Natsoulis, G., Hilger, F. and Fink, G.R., (1986), *Cell*, **46**, 235–243.
18. Gampel, A. and Tzagoloff, A., (1989), *Proc.Natl.Acad.Sci. USA*, **86**, 6023–6027.
19. Koerner, T.J., Myers, A. M., Lee, S., Tzagoloff, A. (1987) *J.Biol.Chem.* **262** 3690–3696.
20. Gatti, D.L. and Tzagoloff, A. (1991) *J. Mol.Biol.* **218** 557–568.
21. Chang, P. K. and Dignam, J.D. (1990) *J. Biol. Chem.* **265** 20898–20906.
22. Nilsen, T.W. et al. (1988), *Proc. natn.Acad.Sci. U.S.A* **85** 3604–3607.
23. Leinfelder, W., Zehelein,E., Mandrand-Berthelot, M-A. and Böck, A., (1988), *Nature*, **331**, 723–725.
24. Forchhammer, K., Leinfelder, W. and Böck, A., (1989), *Nature*, **342**, 453–456.20. Fersht, A.R., (1987), *Biochemistry*, **26**, 8031–8037.
25. Fersht, A.R., (1987), *Biochemistry*, **26**, 8031–8037.
26. Nassar, N and Cusack, S. Unpublished results.
27. Anselme, J. and Härtlein, M. *(1991)* FEBS Letters **280** 163–166.
28. Mirande,M. and Waller, J-P., (1988), *J.Biol.Chem.*, **263**, 18443–18451.
29. Moine, H. *et al.* (1988),*Proc. natn. Acad.Sci. U.S.A* **85** 7892–7896.
30. Springer, M. et al. (1989) *EMBO J.* **8** 2417–2424.
31. Moine, H. et al. (1990) *J. Mol. Biol.* **216**, 299–310.
32. Härtlein, M., Madern, D. and Leberman,R., (1987), *Nucleic Acids Res.*, **15**, 1005–1017.
33. Dayhoff, M. O., Barker, W.C. and Hunt, L. T., (1983), *Methods Enzymol.* **91**, 524–545.
34. Wong, J. T. (1975) *Proc. natn. Acad.Sci. U.S.A.* **72** 1909–1912.
35. Weiner, A. M. and Maizels, N. (1987) *Proc. natn. Acad.Sci. U.S.A.* **84** 7383–7387
36. Putney,S.D., Sauer,R.T. and Schimmel, P., (1981), *J.Biol.Chem* , **256**, 198–204.
37. Sellami, M., Fasiola, F., Dirheimer, G., Ebel, J-P. and Gangloff,J., (1986), *Nucleic Acids Res* , **14**, 1657–1666.
38. Jacobo-Molina, A., Peterson, R. and Wang, D.C.H., (1989), *J.Biol.Chem.*, **264**, 16608–16612.
39. Mechalum, Y., Fayat, G. and Blanquet, S., (1985), *J.Bacteriol.*, **163**, 787–791.
40. Sanni, A., Mirande, M., Ebel, J-P., Boulanger, Y., Waller, J-P., and Fasiolo, F., (1988), *J.Biol. Chem.*, **263**, 15407–15415.
41. Webster, T.A., Gibson,B.W., Keng, T., Biemann, K and Schimmel,P., (1983), *J.Biol.Chem* , **258**, 10637–10641.
42. Freedman, R., Gibson, B., Donovan, D., Biemann, K., Eisenbeis, S., Parker, J. and Schimmel, P., (1985), *J.Biol.Chem.*, **260**, 10063–10068.
43. Anselme, J. and Härtlein, M (1989), *Gene*, **84**, 481–485.
44. Weygand-Durasevic, I., Johnson-Burke, D. and Soll, D., (1987), *Nucleic Acids Res.*, **15**, 1887–1904.
45. Mayaux, J-F., Fayat, G., Fromant, M., Springer, M., Grunberg-Manago, M. and Blanquet, S., (1983), *Proc. natn. Acad.Sci. U.S.A*, **80**, 6152–6156.
46. Pape, L.K. and Tzagoloff, A., (1985), *Nucleic Acids Res.*, **13**, 6171–6183.