Aminoacyl-tRNA Synthetases: Partition into two Classes

M. DELARUE and D. MORAS[1]

1 Introduction

Aminoacyl-tRNA synthetases (aaRS) constitute a family of enzymes that catalyze the specific attachment of one amino acid (aa) to its cognate tRNA in what is a key step in the translation of the genetic information during protein biosynthesis. This enzymatic reaction requires ATP and can be decomposed in two steps:

$$aa + ATP \rightarrow aa\text{-}AMP + PPi$$
$$aa\text{-}AMP + tRNA \rightarrow aa\text{-}tRNA + AMP$$

Over the years, considerable information has been gathered on the structural and kinetic properties of this family of enzymes (for a recent review, see Schimmel 1987). The net result of these studies, however, was disturbing since synthetases were essentially characterized by their structural diversity: little, if any, sequence homology, very different molecular weights and quaternary structures. On the functional level, the situation was similar since some synthetases were shown to interact with the anticodon of their cognate tRNA, some were not, some had a proofreading mechanism and some had none.

It was therefore tempting to try to correlate the different behaviour of aaRS to differences in the amino acids themselves; or to differences in the tRNAs: for instance, a lot of work based on sequence analysis and site-directed mutagenesis is currently being done on various tRNAs to look for tRNA identity (for recent references, see McClain et al. 1991). Unfortunately, while extremely valuable, this genetic and biochemical type of approach tends to reveal more differences than unifying principles.

Since the evolution of aaRS may be connected to the existence and the emergence of the genetic code, it can easily be understood that the search for common features in aaRS has been very intense. Just looking at the reaction performed by all aaRS (see above), one can see that all these enzymes require ATP; it was therefore reasonable to postulate that all of them should have one common structural domain, responsible for the

[1] Laboratoire de Cristallographie Biologique, I.B.M.C. du C.N.R.S., 15 Rue R. Descartes, 67000 Strasbourg, France

Nucleic Acids and Molecular Biology, Vol. 6
ed. by F. Eckstein and D.M.J. Lilley

binding of ATP; indeed, the so-called Rossmann fold found in the structure of three different aaRS, MetRS, TyrRS and GlnRS (Blow et al. 1983; Perona et al. 1991), seemed to fulfill this requirement (Rould et al. 1989).

Recently, however, this view has changed and the presence of this domain is postulated in only 10 out of the 20 aaRS, meaning that the 10 remaining aaRS belong to another class. This prediction was made possible because more and more sequences of different aaRS became known in different organisms, due to the widespread use of molecular biology techniques. A few years ago, two short peptides ("HIGH" and "KMSKS", also called the "consensus sequences") had been identified in at least nine different aaRS (Webster et al. 1984; Hountoudji et al. 1986; Burbaum et al. 1990). Using careful sequence comparisons (Eriani et al. 1990b), it was possible to delineate three new sequence motifs common to 10 aaRS that do not display the so-called two "consensus sequences", thereby defining at least two classes of aaRS (hereafter referred to as class I and class II). The only unknown sequence at that time, CysRS, was later shown to be a member of class I (Eriani et al. 1991; Hou et al. 1991).

At about the same time, structural results confirmed the validity of such a partition of all aaRS into two separate classes. Indeed, the structure of SerRS (Cusack et al. 1990), and, more recently, the one of AspRS (Ruff et al. 1991), two members of class II, were solved and displayed a new and common ATP-binding and catalytic domain entirely different from the one of class I (see Figs. 3 and 4), confirming the sequence alignments mentioned above (Eriani et al. 1990b). It is therefore now apparent that aaRS have evolved into two separate families. The aim of this review is to describe in more detail all the experimental facts relevant to this partition of all aaRS, at the structural, functional and evolutionary level.

2 Partition of the 20 aaRS into Two Separate Classes on the Basis of Sequences Alone (Some Like It HIGH)

2.1 Class I aaRS

This class contains ten members, namely MetRS, CysRS, LeuRS, IleRS, ValRS, TyrRS, TrpRS, ArgRS, GluRS and GlnRS. The two consensus sequences (HIGH and KMSKS) that are present in these aaRS were discovered a few years ago (see Burbaum et al. 1990 for a recent review) and only recently has CysRS been added to this group (Eriani et al. 1991; Hou et al. 1991). These two consensus sequences, which are displayed in Fig. 1, are not especially well conserved and, in our opinion, certainly not sufficient to be called motifs in the sense of Gribskov et al. (1987). Judging from the size of these consensus sequences compared to the total length of these proteins (which can average 900 residues for LeuRS, IleRS, ValRS for instance), it can be seen that the homology between them is very weak

ALIGNMENT OF THE TWO CONSENSUS SEQUENCES
OF CLASS I tRNA-SYNTHETASES

METRS	(Ec)	14	PYANGSIHLGHMLE...	320	GAKMS	K	SRGTFI
METRS	(Sc)	205	PYVNNVPHGLNIIG...	523	NGKFS	K	SRGVGV
ILERS	(Ec)	58	PYANGSIHIGHSVN...	600	GRKMS	K	SLKNYP
ILERS	(Sc)	47	PFATGTPHYGHSVN...	601	GRKMS	K	SIGNTV
LEURS	(Ec)	42	PYPSGRLHMGHVRN...	617	MSKMS	K	SKNNGI
LEURS	(Scm)	56	PYPSGATHMGHLRV...	644	YEKMS	K	SKYNGA
VALRS	(Ec)	42	PNVTGSLHMGHAFQ...	701	GRKMS	K	SLGNVI
VALRS	(Sc)	190	PNVTGALHIGHALT...	552	GQKMS	K	SKGNVI
VALRS	(Bst)	49	PNVTGKLHLGHAWD...	523	GRKMS	K	SLGNGV
CYSRS	(Ec)	32	ITVYDLCHIGHGRT...	264	REKMS	K	SLGNFF
TYRRS	(Ec)	32	DPTADSLHIGHLVP...	220	GTKFG	K	TEGGAV
TYRRS	(Nc)	95	DPTAPSLHVGHLLP...	310	GAKFG	K	SAGNAI
TYRRS	(Bsu)	49	DPSAPDVHLGHTVV...	224	VEKMS	K	SK
TRPRS	(Scm)	34	QPT.GCFHLGNYLG...	222	EKKMS	K	SDPNKQ
TRPRS	(Ec)	3	QPS.GELTIGNYMG...	172	TKKMS	K	SDDNRN
TRPRS	(Bst)	1	QPS.GVITIGNYIG...	169	TKKMS	K	SDPNHD
ARGRS	(Ec)	31	PNVAKEMHVGHLRS...	430	YADLS	K	NRTTDY
GLNRS	(Ec)	27	PEPNGYLHIGHAKS...	264	YTVMS	K	RKLNLL
GLNRS	(Sc)	252	PEPNGYLHIGHSKA...	494	GTVLS	K	RKIAQL
GLNRS	(Hs)	31	PEASGYLHIGHAKA...	262	NTVLS	K	RKLTWF
GLURS	(Ec)	3	PSPTGYLHVGGART...	234	GKKLS	K	RHGAVS
GLURS	(Rm)	6	PSPTGEPHVGTAYI...	250	KSKLS	K	RKNPTS
GLURS	(Bst)	5	PSPTGHLHIGGART...	250	RKKLS	K	RDESIA
GLURS	(Bsu)	5	PSPTGHLHIGNART...	249	RKKLS	K	RDESIA

Fig. 1. Alignments of the two consensus sequences of class I aaRS, as derived from Landes et al. (1991). The position of the beginning of each stretch of sequence is indicated. All the known sequences, as found in the last version of the GenBank distributed with the UWGCG package, are included. Strongly conserved residues are *underlined*. Strictly conserved residues are *boxed*. Abbreviations: *Escherichia coli* (*Ec*), *Bacillus stearothermophilus* (*Bst*), *Bacillus subtilis* (*Bsu*), *Neurospora crassa* (*Nc*), *Saccharomyces cerevisiae* (*Sc*), *Saccharomyces cerevisiae* mitochondria (*Scm*), rat mitochondria (*Rm*), *Homo sapiens* (*Hs*)

indeed. The distance between the two consensus sequences can be quite variable and their location relative to the N- or C-terminus is not well conserved either. This is one of the reasons why it has taken so long to define groups of synthetases on the basis of sequences alone. Nevertheless, now that three structures of class I aaRS are known (Brick et al. 1988; Rould et al. 1989; Brunie et al. 1990), it is possible to define, from sequence alignments guided by the knowledge of three-dimensional structures, three subclasses in class I aaRS (Landes et al. submitted). Inside each of these subclasses, the sequence similarities are much more extended than between members of different subclasses.

Class Ia contains MetRS, CysRS, IleRS, LeuRS, ValRS (see also Hou et al. 1991, for instance, for a map of the relative insertions and deletions in this subclass);

Class Ib contains TyrRS and TrpRS;

Class Ic contains GlnRS, GluRS, ArgRS.

This partition of class I aaRS is based on the systematic use of an alignment program called RELATE (Dayhoff et al. 1983), the comparisons of hydrophobicity patterns and the knowledge of the three-dimensional and therefore secondary structure of three members of this class. Basically, the sequence similarities between class I aaRS can be extended to an entire structural domain of about 170 amino acids, starting from the first consensus sequence and finishing at the second consensus sequence. There can be large insertions in this region, called the connective peptides 1 and 2 (see Fig. 3).

2.2 Class II aaRS

The ten remaining aaRS were recently shown to share three motifs, which proved to be specific for this set of aaRS: a search through the entire data base with profiles derived from an alignment of these motifs proved to detect selectively only class II aaRS (Eriani et al. 1990b). It is the *concatenation* of the three motifs that is specific for class II aaRS: each one of them individually is not enough to detect specifically aaRS. Omitting one of the ten different class II aaRS in the construction of the profile still led to the detection of this aaRS in the search through the entire data bank. To define the motifs, we found it essential to align first closely related aaRS (including all aaRS from different origins) to assess the regions of high homology and the location of gaps. Then we tried to match conserved regions in different subclasses, not paying too much attention to peptides well conserved between subclasses if they were not well conserved inside each subclass (see Argos 1987). We also tried to maintain conserved residues in the same environment. Three regions of high homology were identified and called motifs 1, 2 and 3 (Fig. 2). It turned out that some of these motifs (motifs 2 and 3) had already been detected for some of the aaRS: a shorter version of motif 2 had been described for AsnRS, AspRS, HisRS and LysRS (Anselme and Haertlein 1989) and of motif 3 for the same enzymes along with PheRS and AlaRS (Molina et al. 1989; Leveque et al. 1990). We showed that motifs 2 and 3 were in fact common to HisRS, PheRS, AspRS, AsnRS, LysRS, SerRS, ThrRS and ProRS and that they shared an additional motif called motif 1, invariably located about 50 residues upstream of motif 2. Only motif 3 could be detected for AlaRS and GlyRS. Motifs 2 and 3 showed strictly conserved arginine residues followed by a loop of variable length and other charged residues, in the case of motif 2, or by five hydrophobic residues in a row, in the case of motif 3. The distance between motifs 2 and 3 was found to vary between 150 and 250 residues (Fig. 2).

ALIGNMENT OF THE THREE MOTIFS OF CLASS II tRNA-SYNTHETASES

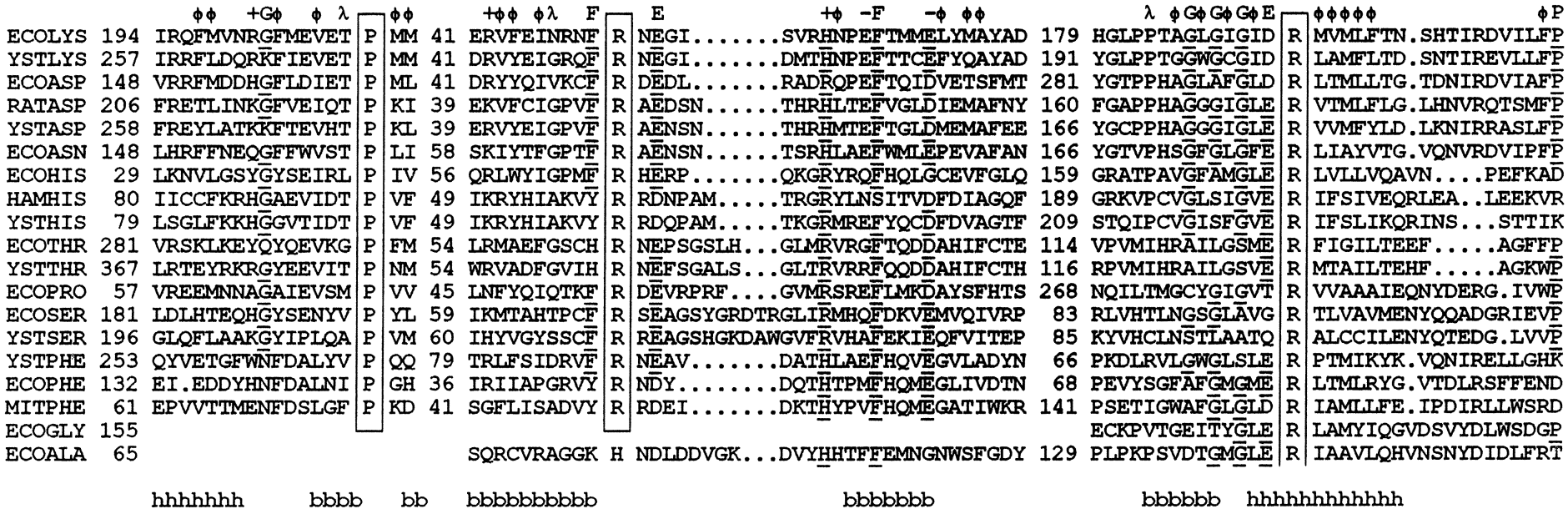


Fig. 2. Alignment of the three motifs of class II aaRS as originally proposed by Eriani et al. (1990b). The position of the first amino acid of motif 1 is shown on the *left*; the *numbers* that appear between motifs are the distances between them. The *bottom line* indicates the secondary structure, as observed in both AspRS and SerRS. The *top line* contains symbols for strongly conserved residues: λ (lambda) for small amino acids (P, A, G, S, T), ϕ (phi) for hydrophobic residues (M, I, L, V, C, F, Y, W), + for positively charged residues (H, K, R) and – for negatively charged residues (D, N, E, Q). A small error in motif 3 for yeast AspRS has been corrected. Cusack et al. (1991) have convincingly argued that motif 2 can be found in AlaRS (included here) and that motif 1 for PheRS may be wrong

As is usual in this situation, these sequence alignments were improved and refined by examination of the three-dimensional structure of two members of the family, extending the homology (which is very weak outside the motifs) to an entire domain of about 250 amino acids (Cusack et al. 1991; Ruff et al. 1991). As for class I, class II aaRS can also be subdivided into three subclasses:

Class IIa contains SerRS, ThrRS, ProRS (Cusack et al. 1990; Eriani et al. 1990b) and HisRS (Cusack et al. 1991); the assignment of HisRS to this subclass is mainly based on sequence homologies in the C-terminus part of ThrRS, ProRS and HisRS, but not SerRS.

Class IIb contains AspRS, AsnRS and LysRS, with also an additional sequence homology, as noted by others, in the N-terminus part of the proteins (Anselme and Haertlein 1989; Gampel and Tzagoloff 1989; Eriani et al. 1990a; Leveque et al. 1990).

Class IIc includes the remaining three enzymes, GlyRS, PheRS and AlaRS which show distinctive features, as compared to the other seven class II enzymes: while all class IIa and class IIb aaRS are dimers, GlyRS in *E. coli* is an $\alpha_2\beta_2$ heterodimer, AlaRS is a homotetramer α_4 in *E. coli* but a monomer in *Bombyx mori*, and PheRS is also an $\alpha_2\beta_2$ heterodimer in *E. coli*. Cusack et al. (1991) have been able to identify a convincing motif 2 in AlaRS; they also argue that our motif 1 for PheRS is unlikely to be correct, because in the SerRS structure (as in the AspRS structure), motif 1 is involved in the dimerization of the protein and can be extended in a

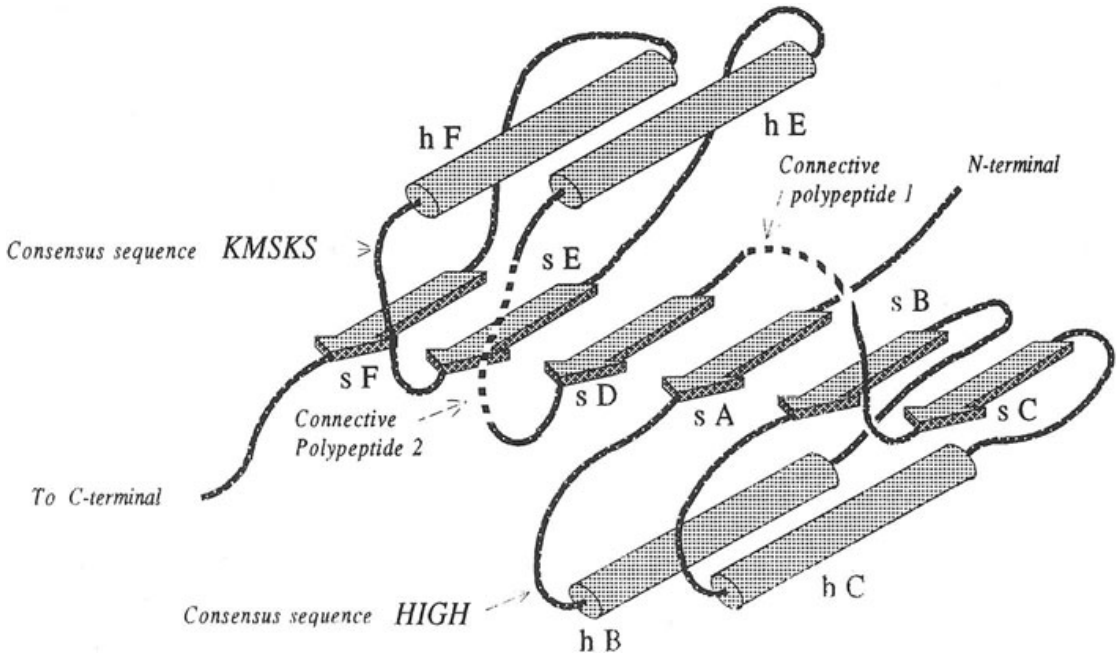


Fig. 3. Schematic drawing of the Rossmann fold, as seen in class I aminoacyl-tRNA synthetases (redrawn from Burbaum et al. 1990). The location of insertion (*connective peptides*) is also indicated

Table 1. Classification of all aaRS into two classes on the basis of sequence similarities (redrawn from Eriani et al. 1990b) with a few modifications: subclasses were taken into account; PheRS has been moved to the column "Motifs 2 and 3 only", to which AlaRS, and GlyRS also belong, even though motif 2 remains undetectable for GlyRS. For those enzymes for which the structure is known, the presence (RF) or absence (no RF) of the Rossmann fold is shown. The primary site (2'OH or 3'OH) of aminoacylation is also displayed (? is used for enzymes where the situation is different in different species and ?? for enzymes where both situations seem possible), as well as the quarternary structure of the enzyme

Class II synthetases							Class I synthetases			
Motifs 2 and 3 only			Motifs 1, 2 and 3				HIGH + KMSKS			
Gly	($\alpha 2\beta 2$)	3'OH								
Ala	($\alpha 4$)	3'OH								
			Pro	($\alpha 2$)	3'OH					
			Ser	($\alpha 2$)	3'OH	No RF				
			Thr	($\alpha 2$)	3'OH					
			His	($\alpha 2$)	3'OH					
			Asp		??	No RF				
			Asn	($\alpha 2$)	3'OH					
							Glu	(α)	2'OH	
							Gln	(α)	?	RF
			Lys	($\alpha 2$)	3'OH					
							Arg	(α)	2'OH	
							Cys	($\alpha 2$)	??	
							Met	($\alpha 2$)	2'OH	RF
							Val	(α)	2'OH	
							Ile	(α)	2'OH	
							Leu	(α)	2'OH	
Phe	($\alpha 2\beta 2$)	2'OH					Tyr	($\alpha 2$)	??	RF
							Trp	($\alpha 2$)	?	

straightforward manner to comprise a full and long α-helix followed by a β-strand (see Fig. 3) in which the sequence homology in PheRS is not good. We agree with these observations on the basis of purely crystallographical results obtained on a MIR map at 4 Å resolution of *Th. thermophilus* PheRS (M. Safro and coll., unpubl. results); furthermore, the small subunit (which contains both motifs 2 and 3) of PheRS of yeast mitochondria has recenlty been shown to be active as a monomer (Sanni et al. 1991); therefore, it makes sense that only the true dimers of class IIa and class IIb contain motif 1. We propose to create a new subclass, class IIc (see Table 1) to describe all those class II aaRS that are not true dimers (absence of motif 1).

The only remaining puzzle is the absence of motif 2 in GlyRS, because structural results on SerRS and AspRS indicate that both motifs 2 and 3 are

part of the active site. However, only one sequence of GlyRS, the one of *E. coli* (Webster et al. 1983), is available, so far; it may well be that more sequences of this enzyme in different organisms will provide the solution to this puzzle.

The resulting partition of all aaRS on the basis of sequence comparisons is shown in Table 1, along with different structural and functional aspects, which will now be discussed.

3 Structural Aspects

This section is based on the recent determination of the three-dimensional structure of five aminoacyl tRNA-synthetases: MetRS from *E. coli* (Brunie et al. 1990), TyrRS from *B. stearothermophilus* (Brick et al. 1988) and GlnRS from *E. coli*, as a complex with its cognate tRNA (Rould and Steitz, this Vol.) for class I aaRS; SerRS from *E. coli* (Cusack et al. 1990) and AspRS from baker's yeast (Ruff et al. 1991) for class II aaRS. There is one structure known for each subclass except for one member of class IIc, but this should be available soon with structural studies on *Th. thermophilus* PheRS well on the way (Chernaya et al. 1987; M. Safro et al. in preparation).

3.1 The ATP-Binding Domain (To Have or Not to Have the Rossmann Fold)

It is striking to observe that the partition into two classes corresponds to the presence of two very different ATP-binding domains. The aim of this section is to stress the structural importance of the "consensus sequences" of class I aaRS, as well as the one of the three motifs of class II aaRS in the definition of this domain, which forms the core of the active site in aaRS.

Class I aaRS bind ATP through an α/β domain called the Rossmann fold, as characterized 15 years ago from the structural analysis of enzymes of the glycolytic pathway (Rossmann et al. 1974). This domain contains (see Fig. 3) a parallel β-sheet, with a $\beta\alpha\beta$ structural motif as the repeating unit. The HIGH consensus sequence is located at the turn between the first β-strand and the first α-helix (strand A and helix B of Fig. 3), with the second histidine residue forming a hydrogen bond with the α- and γ-phosphates of ATP. The KMSKS is located in the loop connecting the fifth β-strand and the fifth α-helix (strand E and helix E). There is also a conserved acidic residue at the end of the fourth β-strand (strand D).

For class II aaRS, the catalytic domain which binds ATP is an anti-parallel β-sheet (see Fig. 4 for a schematic drawing of this domain, where only the features common to AspRS and SerRS have been included). Motif 2 contains a very large and flexible loop of variable length between two antiparallel and adjacent strands. Motif 3 is the central strand of the β-sheet,

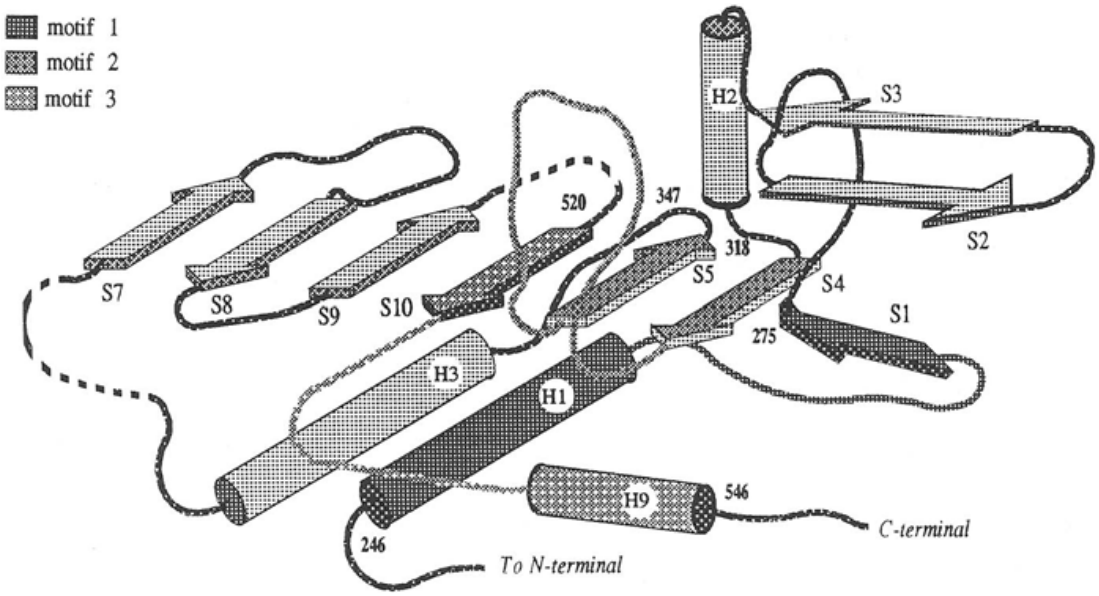


Fig. 4. Schematic drawing of the antiparallel β-sheet that forms the catalytic domain of class II aminoacyl-tRNA synthetases (redrawn from Ruff et al. 1991). Only the secondary structures common to both SerRS and AspRS are displayed. The presence of insertions is shown by *dotted* connections. The location of the different sequence motifs is also indicated

followed by a hydrophobic helix. With the help of the structure, it can be seen that motif 1, at the dimer interface, should be extended to contain the entire helix 1 (see Fig. 4), and also the β-strand S1. Motifs 2 and 3 have their strictly conserved arginine residues pointing into a cavity that is thought to be the active site of the molecule. More detailed studies of the binding mode of ATP to this domain should be available soon, with data already collected for *Th. thermophilus* AspRS in the presence of ATP (M. Delarue, S. Nikonov, J.-C. Thierry in preparation), as well as with the yeast AspRS-tRNA^{Asp} complex (D. Moras et al. in preparation).

3.2 tRNA Positioning (Right Side, Left Side: a Variation in Minor/Major Grooves)

Careful structural comparisons of MetRS and GlnRS in *E. coli* recently allowed the identification of two other regions of (structural) homology between these two enzymes, which could not be detected at the sequence level (Perona et al. 1991). One of these regions encompasses a 23 amino acid long α-helix – turn – β-strand motif inserted between the two halves of the Rossmann fold; the other is made of two α-helices connected by a large loop and a β-strand which bind into the corner of the L-shaped tRNA. This alone seems to be enough to position the tRNA, mainly on the right side of the acceptor stem, with major interactions with tRNA through the minor

groove, as described for the GlnRS-tRNA^{Gln} complex (Rould and Steitz, this Vol.). Also, the CCA arm is very much distorted, as compared to its conformation in free tRNA. In contrast, as revealed by the AspRS-tRNA^{Asp} complex (Ruff et al. 1991), the situation is quite reversed for class II aaRS: the CCA has a normal helical conformation and only the left side of the tRNA is in contact with the protein, making direct interactions through the major groove (see Fig. 6).

3.3 tRNA Recognition (Where Idiosyncrasy Makes a Strong, But Not Unexpected Comeback)

Since the structure of only two complexes (GlnRS, a class I aaRS, and AspRS, a class II aaRS) of an aaRS with its cognate tRNA has been determined, it is somewhat hazardous to draw general conclusions; however, the statement made above about the positioning of the protein relative to the acceptor stem is convincing enough to be accepted as general. There are many other contacts between the protein and tRNA, as stressed by T. Steitz and colleagues for the GlnRS complex (Rould et al. 1989, 1991), for instance. What can be said about these other zones of tRNA-protein contacts?

In the GlnRS-tRNA^{Gln} complex, an additional domain, β-barrel like, binds the anticodon, an obvious element of the identity of tRNAs. The binding induces a conformational change that extends the anticodon stem so that the three bases of the anticodon itself are completely unstacked, as compared to the well-ordered helical loop of the free tRNA, only to hide inside hydrophobic pockets of the β-barrel domain of the protein (Rould et al. 1991; Rould and Steitz, this Vol.). Perona et al. (1991) have derived a model for the interaction of the MetRS enzyme with its cognate tRNA, by superimposing the two Rossmann fold domains of MetRS and GlnRS; it can then be seen that the anticodon binding domain in MetRS is mostly α-helical and totally different from the one of GlnRS.

For Class II aaRS, the yeast AspRS-tRNA^{Asp} complex structure also provides information on the interactions of the protein with the anticodon; this also involves a β-barrel-like protein domain (Ruff et al. 1991), but with a different fold than GlnRS; however, the structural details of this interaction should await further refinement of the structure. This domain, located at the N-terminus of the sequence, seems to be reasonably conserved in AspRS, AsnRS and LysRS (Anselme and Haertlein 1989; Gampel and Tzagoloff 1989; Eriani et al. 1990a; Leveque et al. 1990). Interestingly, the superposition of the common ATP-binding domain of SerRS and AspRS allows a prediction to be made on the probable interaction of SerRS with its cognate tRNA: it was found that there is interaction not with the anticodon of the tRNA, but rather with the extra loop of tRNA^{Ser}, through a very peculiar structural feature of SerRS, namely the very long coiled-coil helices of the N-terminal part of the protein (Cusack et al. 1990). This model of

interaction has also been proposed on the basis of "docking experiments" and fits experimental data on protection of tRNA (Dock-Bregeon et al. 1990; Schatz et al. 1991). It is important to note that the helices of, say, monomer A, are in contact with the extra loop of the tRNA whose CCA is in contact with monomer B. This interaction with the extra loop of $tRNA^{Ser}$ instead of its anticodon is particularly attractive since Ser has six codons at quite different positions in the genetic code and many isoacceptors are present in the cell: in this case, nature has found a way to achieve specificity without resorting to anticodon recognition. Interestingly, we note that the very long coiled-coil helices of SerRS have recently been predicted, through an especially devised algorithm working on sequences only, to also be present in *E. coli* AlaRS (Lupas et al. 1991).

Recently, a considerable effort has been devoted to the definition of the so-called identity of tRNA. This was made possible by the use of molecular biology techniques which allow the production and isolation of many mutants of tRNA. This has been done on $tRNA^{Lys}$, $tRNA^{Gln}$, $tRNA^{Asp}$, $tRNA^{Gly}$, $tRNA^{Leu}$, $tRNA^{Phe}$, $tRNA^{Arg}$, $tRNA^{Met}$, $tRNA^{Ala}$ and $tRNA^{Tyr}$ (see Normanly and Abelson 1989 for a review and, for instance, McClain et al. 1991 for an introduction to recent literature). There again, diversity seems to be the rule, but it will probably be interesting to analyze the results in more detail with the help of the structures of tRNA-aaRS complexes of classes I and II.

3.4 Amino-Acid Binding (Dark Passage)

The only system for which the amino acid binding site has been described at atomic resolution is TyrRS of *B. stearothermophilus*. There is clearly a specificity for the hydroxyl part of tyrosine at the bottom of the cleft that forms the active site, through hydrogen bonding with Asp 76 and Tyr 34, and van der Waals contacts on the phenyl ring, especially on one side of it (Brick and Blow 1987). However, the cavity seems to be able to accomodate larger substrates, as demonstrated experimentally using analogues of tyrosine with ortho and meta large substituents. The α-amino group of the amino acid is tightly bound, probably to correctly orient the α-carboxyl group, which is free of any hydrogen bond, ready for the aminocylation reaction. Clearly, the respective locations of the amino acid and ATP binding sites imply that the kinetic reaction has to be an ordered one: the amino acid has to bind before ATP, otherwise its binding site is blocked.

4 Functional Aspects

AaRS are built from functionally defined and separated domains sequentially distributed along the polypeptide chain: this was clearly demonstrated by genetic experiments on AlaRS (Jasin et al. 1983) and subsequently general-

ized for other systems, like virtually all members of class Ia. The relative position of the active site (ATP binding domain) and the other functional domains on the primary sequence varies from one class to the other, with a trend toward similarity of distribution within each of the two classes, i.e. active site on the NH_2 terminus end for class I (see Schimmel 1987) and on the COOH end for class II.

4.1 Catalytic Domain (2′OH vs 3′OH: the Fundamental Ambiguity of RNA Revisited)

For class I aaRS, both consensus sequences are involved in the binding of the adenylate and for class II aaRS, the two strictly conserved arginine residues of motifs 2 and 3 have been mutated into lysine, resulting in a complete loss of activity (Eriani et al. in preparation), confirming their role in ATP binding: the functional role of the sequences mentioned above begins to be well established and documented. However, there is a more profound functional and chemical sense to the partition of aaRS into two classes.

Twenty years ago, enzymatic studies on aaRS revealed that the primary site of aminoacylation on the last adenosine of tRNA can be either its 3′OH or 2′OH end (for a review, see Hecht 1979; Schimmel and Söll 1979). If a 2′OH adenylate is formed, rapid isomerization occurs to the 3′OH after release of the enzyme. These studies were essentially made using chemically or enzymatically modified tRNAs, with either the 2′OH or 3′OH removed or changed to an amino group. Strikingly, aaRS partition into two groups, depending on this primary site of aminoacylation, and the partition derived from sequence comparisons are found to coincide almost exactly (Hecht 1979; Eriani et al. 1990b; Table 1). Later, this partition was refined by studying the kinetic characteristics of aaRS with different classes of ATP analogues; the results confirm the earlier studies (Freist et al. 1981).

4.2 The Role of the Other Domains (Anatomy of a Border)

A clear result of the five known three-dimensional structures of aaRS is the existence of additional domains, physically distinct from the catalytic site. In AspRS and GlnRS, some of these extra domains are associated with the binding of the anticodon of the tRNA.

Apart from providing additional binding energy and/or specificity, is there another functional role for some of the other additional domains of the aaRS? For instance, in monomers like GlnRS, is there a transmission of information, along the protein backbone, from the anticodon binding domain to the catalytic domain and triggered by the binding of tRNA, as suggested by T. Steitz and colleagues (Perona et al. 1991)? For dimeric

enzymes like AspRS, is there anticooperativity for the binding of ATP? Are the monomers equivalent? The answers to these questions will have to await further structural and functional studies dealing, in particular, with the three-dimensional structure of the enzyme alone.

AaRS exhibit different behaviours toward substrate binding. For instance, GlnRS and ArgRS need to bind tRNA prior to ATP binding, whereas most of the other aaRS do not. Some synthetases have a proofreading mechanism (e.g. PheRS, ValRS, IleRS), others exhibit strong (TyrRS, CysRS) or significant (MetRS) specificity. These properties contribute to the diversity of the family and are certainly associated with different structural features. Whether these are due to additional domains or more subtle interplay and cooperation of existing and already known domains remains to be seen.

4.3 Some Unanswered Questions

While it is clear that recent structural results contribute greatly to a better understanding of how aaRS work, there are still many unanswered questions. Here is a short list of some of them.

Synthetases have been shown to participate in other steps of the cell life. For instance, some aaRS exhibit a maturase activity and catalyze the excision of type I introns (Lambowitz and Perlman 1990, see below). Others have been invoked in regulatory functions, like ThrRS which regulates its own expression via binding to its own mRNA (Springer et al. 1989). The (evolutionary) meaning of these other roles of aaRS is still unclear.

Where does the specificity for the amino acid come from? Are there general rules that could explain the extraordinary accuracy of the translation apparatus; does the chemical proofreading mechanism proposed by von der Haar and Cramer (1976) have any structural basis?

How can one describe in atomic detail the aminoacylation reaction? Is it possible to have a dynamic picture of the entire process? Time-resolved Laue crystallography should help answer that question.

Why are there sometimes two different aaRS specific for the same amino acid in the same organism (see Leveque et al. 1990 for LysRS)? One of the genes is thermoinducible and regulated differently than the other; is this related to the ability of some synthetases to promote the synthesis of $A(p)_4A$, produced in some extreme (stress, heat shock) conditions?

5 **Evolutionary Aspects** (Desperately Seeking a Scenario)

The evolution of aaRS is very puzzling. They seem to have evolved from two different ancestors with two very different scaffolds: one is the Rossmann fold, present in many other enzymes that bind ATP, the other is an anti-

parallel β-sheet with no equivalent in any other structure known up to now. What is the reason for the partition of aaRS into two groups of ten members each?

5.1 Proofreading Mechanisms

One striking feature of the partition given in Table 1, as well as the identification of subclasses inside each class, is that each subclass corresponds to amino acids with similar physico-chemical properties (in the following discussion, residues, if not charged, are called polar if they are 95% or more buried in less than 40% of the cases, hydrophobic otherwise; see Chothia 1974).

Class Ia contains enzymes specific for hydrophobic residues (Met, Cys, Leu, Ile, Val), class Ib has aromatic residues (Tyr and Trp) and class Ic has charged and large residues (Gln, Glu, Arg); class IIa has small and polar residues (Ser, His, Pro, Thr), while class IIb has charged and small (compared to class Ic) residues (Asp, Asn, Lys). For Gly, Ala and Phe, more structural results are needed to assess their relationship with other class II members. Ala and Gly can be classified as polar, while Phe is much more hydrophobic than Tyr and Trp. The general trend is that class II amino acids are smaller and more polar than class I amino acids. This probably indicates a divergent evolution of aaRS from two different ancestors; how long these two ancestors with a broad specificity for amino acids have lasted before diverging is difficult to answer, but it seems certain that they coexisted and coevolved in a parallel fashion because both types of amino acids (classes I and II) are needed for folded and functional proteins, i.e. compact, stable heteropolymers, with an hydrophobic core and a backbone able to make sharp turns, with functional residues at critical places, defining an active site with a precise geometry. The fact that a mechanism exists for correction of misaminoacylation for hydrophobic residues is probably not fortuitous (why should nature bother building a complicated active site to aminoacylate on 2'OH, while the ultimate target is the 3'OH?) and can be tentitavely explained as follows: the aaRS specific for hydrophobic residues are the ones which need this correction mechanism more, because their active site is bound to be less specific than the one for polar residues: the former can use hydrogen bond mediated interactions, the latter has to rely entirely on van der Waals interactions, which are less specific. This means that small hydrophobic amino acids can readily bind to the active site of aaRS specific for larger hydrophobic residues (see Igloi et al. 1978, for instance). As this was probably troublesome for proper folding, correction mechanisms were developed. Some years ago, Fersht proposed the so-called double sieve filtering mechanism in which larger amino acids are first discarded on the basis of their too large volume, preventing the correct positioning of the α-amino group in the active site; smaller residues are then rejected because

they allow the binding of a water molecule, activated by the free 3′OH (Fersht and Kaethner 1976; von der Haar and Cramer 1976). Polar residues probably do not need this correction mechanism because their active site is more specific in the first place, through directional hydrogen bonds. It is interesting to note in this context that those enzymes that do not have proofreading mechanisms (CysRS and TyrRS) seem to have no preference for the primary site of aminoacylation (see Table 1).

5.2 Similarities of aaRS with Proteins with Other Functions as a Clue to Their Evolution

Another hypothesis is that this partition is the fossil of the very primordial translation apparatus of the first living organisms. Following this hypothesis, some attempts have been made to link the evolution of aaRS with the emergence of the genetic code and also with the establishment of biosynthetic pathways (Wong 1975), but these early attempts are not compatible with the classification of aaRS described here. Anyway, these biosynthetic pathways are probably more ancient than the selection of the 20 amino acids, as they are now known, as the building blocks of proteins. The genetic code itself may have an ancestor, more symmetrical and specifying less amino acids, because it is probably the result of an evolution with a precise pressure selection: to minimize the effect of random mutations on the translated message, i.e. replace an amino acid with one with similar physical and chemical properties (Haig and Hurst 1991). It is therefore impossible to speculate too much, not knowing the intermediates in the evolution.

At this point, it is worth mentioning that some sequence similarities between some aaRS and enzymes that perform different tasks in the cell have been noted by other authors. They include: LeuRS and a leucine binding protein (Williamson and Oxender 1990), TrpRS and a chain release factor (Lee et al. 1990), HisRS and GCN2 (Wek et al. 1989), which takes part in the regulation of the biosynthetic pathways of amino acids, AspRS (or AsnRS) with aspartate ammonia ligase (Gatti and Tzagoloff 1991). Actually, the aaRS-like domain of GNC2 is probably used to monitor the amount of free and loaded tRNA in the cell. It is important to note that motifs 2 and 3 of GCN2 and aspartate ammonia ligase are quite degenerate and could not have been picked up by our profile searches (Eriani et al. 1990b); they did, however, show up using motif 1 alone as a profile. Other metabolic enzymes are known to require tRNA as a cofactor, for instance in the biosynthetic pathway of chlorophyll (Schoen et al. 1986); tRNA is also known to be necessary to initiate the transcription of some retroviral genomes by reverse transcriptases (see Jacobo-Molinas and Arnold 1991 for a review), but this probably reflects a very ancient role of tRNA which is also known to be involved in the degradation pathway of proteins (Ferber

and Ciechanover 1987). On the whole, the information available to date is too sparse and incomplete to make any definite conclusion, and no convincing correlation between the genetic code and the partition of aaRS can at present (and to our knowledge) be made.

However, an additional function has been found for aaRS, which might be relevant to the 2′OH/3′OH specificity of different aaRS, namely the fact that a maturase function was found in some aaRS (Lambowitz and Perlman 1990): LeuRS in yeast (Herbert et al. 1988), and, in a separate study, TyrRS of *Neurospora crassa* (Kittle et al. 1991). It would be interesting to determine whether a class II intron splicing activity could be associated with an aaRS, because the excision of class II introns goes through a lariat intermediate and involves special chemistry at both the 2′OH and 3′OH of the same ribose.

5.3 The RNA World Hypothesis

One possible scenario for the apparition of the two different ancestors of present-day aaRS is the following: at the beginning was RNA, a folded RNA, which has been shown in numerous examples to be able to perform catalysis in several chemical reactions (Cech 1987). Somehow this RNA developed two separate binding sites for amino acids (see Yarus 1988), say, one site for large and hydrophobic residues, close to the 2′OH of one nucleotide, and one site for small and polar residues, close to the 3′OH of the same nucleotide. Binding of, for instance, a class I amino acid to its proper site and its attachment to the 2′OH would have to be followed by the binding of a class II amino acid to its site (the only one available) and attachment to the 3′OH of the ribose. Formation of the peptide bond would automatically make the class I site free again, ready to bind another amino acid (see Schimmel and Söll 1979; Fig. 5). The reaction could go on to leave, after another condensation, a class II amino acid site free again to give, finally, a polypeptide characterized by a sequence of alternate amino acids of class I, class II, class I, class II . . . (binary code). A specificity not too restricted for each site would allow sufficient variability to ensure, eventually, the synthesis of useful (i.e. capable of folding and displaying some catalytic activity or some structural role) polypeptides. The system would want to keep track of the blueprints necessary to reproduce the useful products, which means a code and machinery to read it; and then, probably, the ribosome would have evolved drastically, eventually dropping off the synthesis of alternate copolymers on the same ribose, overcoming the strict alternation of hydrophobic/hydrophilic amino acids so that more varied and useful structures could then appear. Note that proteins of alternate hydrophobic and hydrophilic amino acids could readily fold into β-sheets and make membrane-spanning channels.

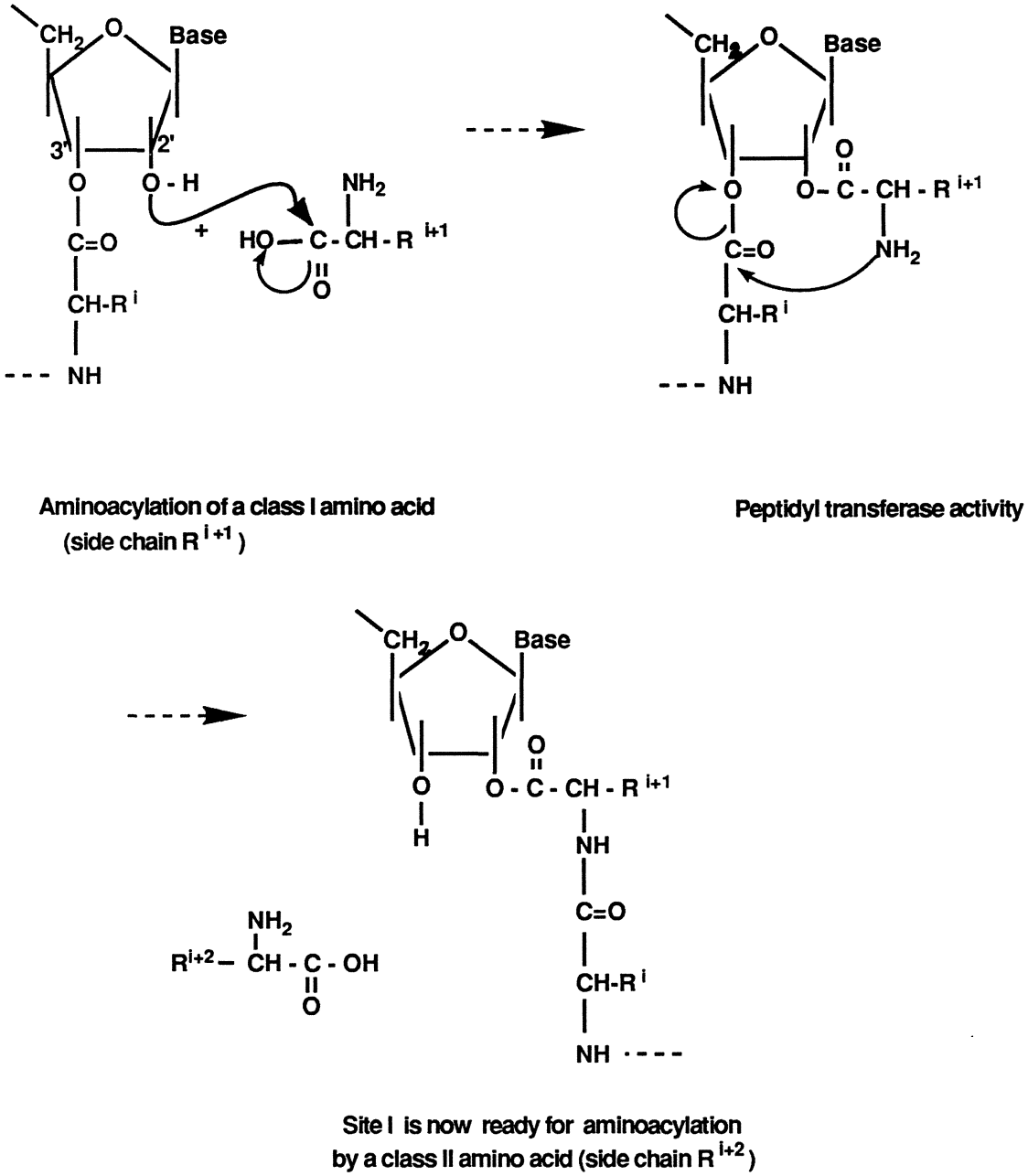
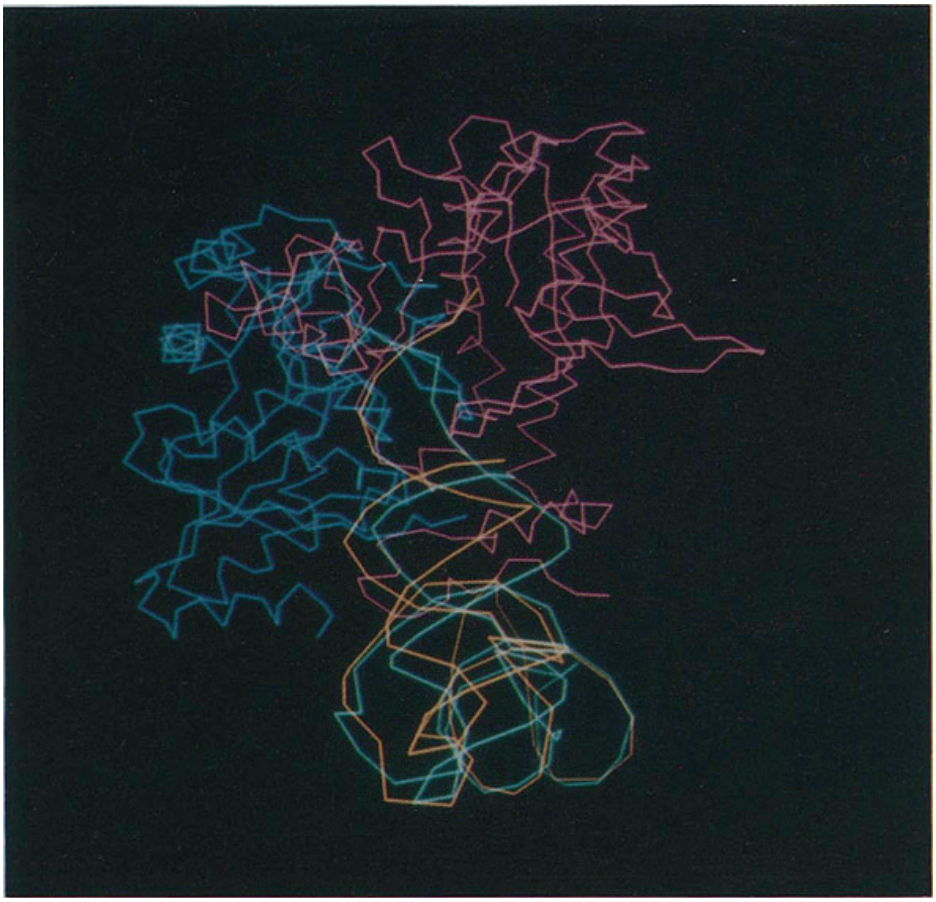


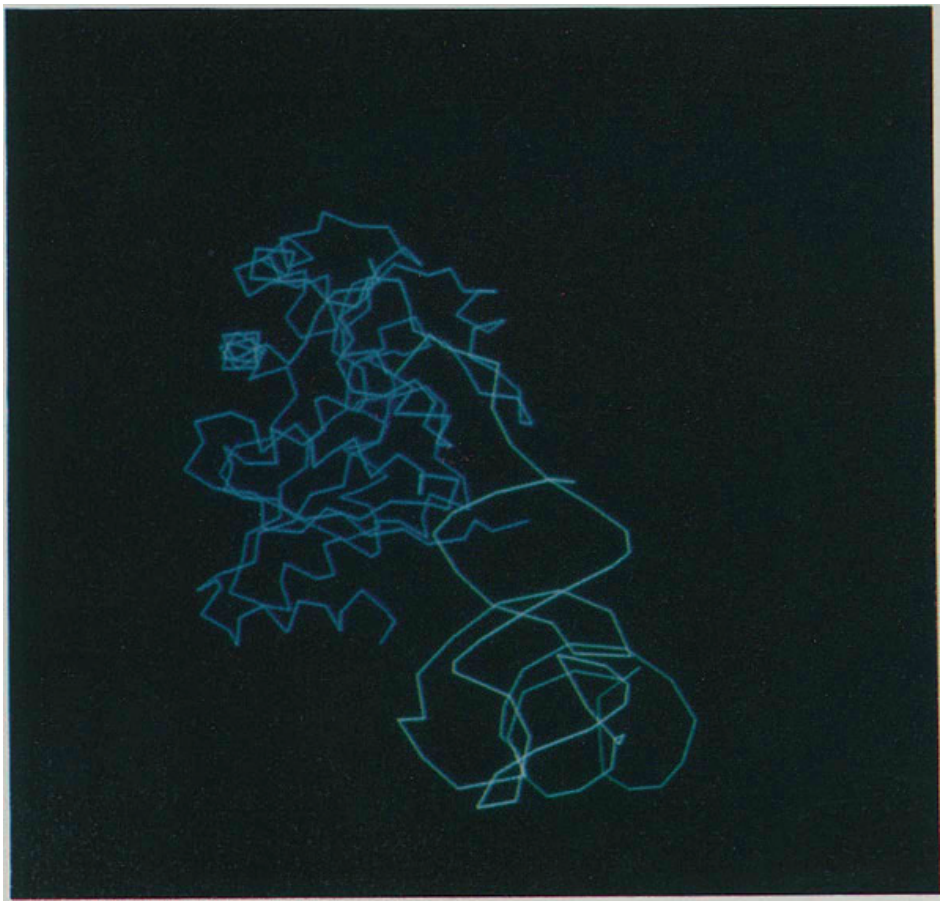
Fig. 5. Possible mechanism for a primitive peptidyl transferase activity and translocation of amino acids between the 2′OH and 3′OH of the same ribose

6 Conclusion and Perspectives

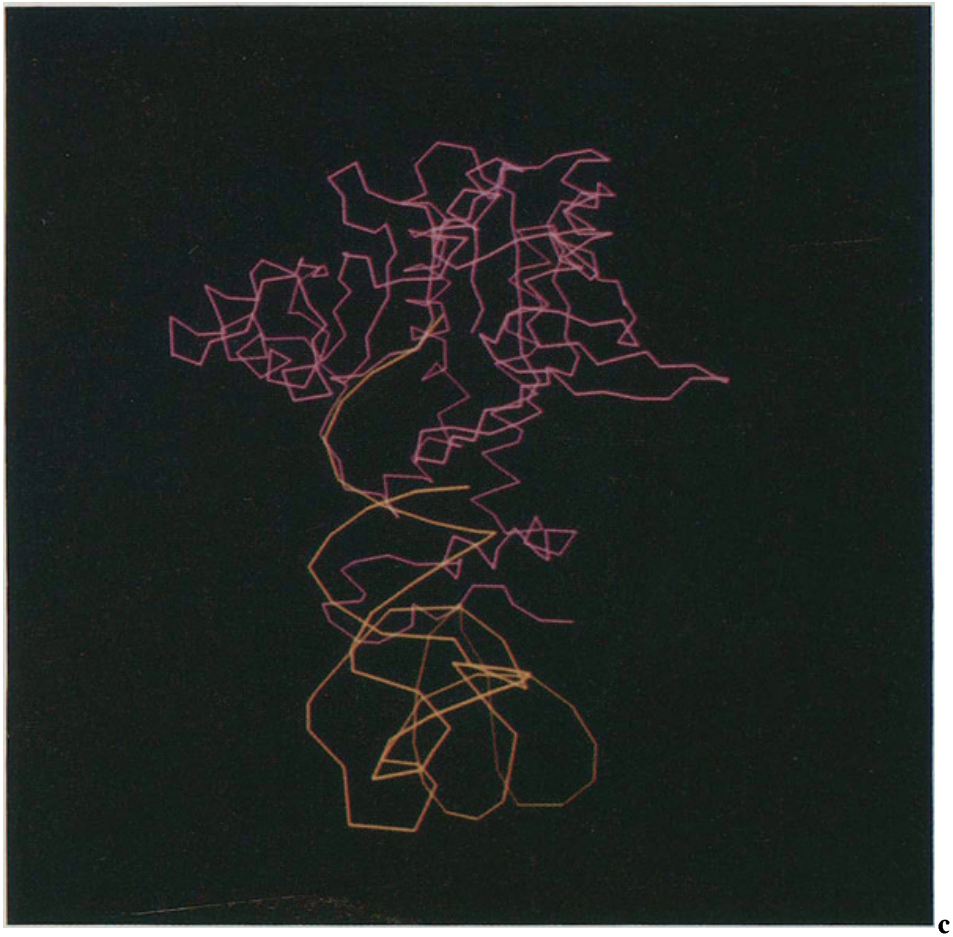
In conclusion, with the recent partitioning of the 20 aaRS into two classes and the almost simultaneous structure determinations of the complexes between representative members of each class with their cognate tRNA, it has become possible to form new ideas about how aminoacylation of tRNAs works. While this falls short of explaining how the genetic code appeared, it



a



b



c

Fig. 6. Superposition of a class I aaRS (GlnRS) and class II aaRS (AspRS) onto the same tRNA. The CCA (phosphate backbone) stem is seen from *above*, the Cα backbone of GlnRS (residues 8 to 262) is shown in *blue* (tRNA is *green*) on the left of the tRNA, the COOH-terminus part of yeast AspRS (residues 205 to 507) is *red* (tRNA is *yellow*) and on the right of the tRNA; the anticodon binding domains of both proteins have been omitted. **a** GlnRS and AspRS with their cognate tRNA superposed; **b** and **c** same view, but for GlnRS and AspRS separately, respectively

is possible to postulate a few mechanisms. For instance, the translocation of two activated amino acids bound to the same ribose is an attractive idea. It is, however, impossible to find any trace of such a mechanism in the present-day coding sequences. Also, the synthesis of random copolymers without any selection mechanism is not a very Darwinian one. In any case, an unexpected application of this observation would be to try to construct a chimera of two catalytic domains of class I and class II aaRS that would bind to the same tRNA. Indeed, since the modes of binding of these two domains to the tRNA are so different, there would be in fact almost no steric clashes (see Fig. 6) of the same tRNA. Simultaneous charging of two different amino acids to the 2′OH and the 3′OH may require some engineering of the protein to accomodate their volumes, and facilitating the formation of the peptide bond may require imaginative ideas in enzymology. However, some

solution studies on model compounds showed this to be possible in the absence of aaRS (Lacey et al. 1991), and that such a mechanism favours L-amino acid polymerization.

More reasonably, protein engineering designed to change the specificity of a given aaRS into another one, even trying to force it to incorporate unnatural amino acids, should now be possible, given the recent accumulation of structural results on aaRS and their different complexes.

Acknowledgements. We thank J.-L. Risler and A. Sanni for sending preliminary versions of their manuscripts, M. Ruff for drawing Figs. 3 and 4, E. Westhof for help with the bibliography, J.C. Thierry, B. Rees and all members of the laboratory involved in structural and functional studies of aminoacyl-tRNA synthetases for helpful discussions and sharing of results.

References

Anselme J, Haertlein M (1989) AsnRS from *E. coli* has significant sequence homology with yeast AspRS. Gene 84:481–485

Argos P (1987) Analysis of sequence-similar pentapeptides in unrelated protein tertiary structures. J Mol Biol 197:331–348

Blow DM, Bhat TN, Metcalfe A, Risler JL, Brunie S, Zelwer C (1983) Structural homology in the amino-terminal domains of two aaRS. J Mol Biol 171:571–576

Brick P, Blow DM (1987) Crystal structure of a deletion mutant of a TyrRS compexed with tyrosine. J Mol Biol 194:287–297

Brick P, Bhat TN, Blow DM (1988) Structure of TyrRS refined at 2.3 Å resolution: interaction of the enzyme with the tyrosyl adenylate intermediate. J Mol Biol 208:83–98

Brunie S, Zelwer C, Risler J-L (1990) Crystallographic studies at 2.5 Å resolution of the interaction of MetRS from *E. coli* with ATP. J Mol Biol 216:411–424

Burbaum JJ, Starzyk RM, Schimmel P (1990) Understanding structural relationships in proteins of unsolved three-dimensional structures. Proteins 7:99–111

Cech TR (1987) The chemistry of self-splicing RNA and RNA enzymes. Science 236:1532–1539

Chernaya MM, Korolev SV, Reshetnikova LS, Safro MG (1987) Preliminary crystallographic study of PheRS from *Th. thermophilus* HB8. J Mol Biol 198:557–559

Chothia C (1974) Nature of accessible and buried residues in proteins. J Mol Biol 105:1–14

Cusack S, Berthet-Colominas C, Haertlein M, Nassar N, Leberman R (1990) A second class of synthetase structure revealed by X-ray analysis of *E. coli* SerRS at 2.5 Å. Nature (London) 347:249–255

Cusack S, Haertlein M, Leberman R (1991) Sequence, structural and evolutionary relationships between class II aaRS. Nucleic Acids Res 19:3489–3498

Dayhoff MO, Barker WC, Hunt LT (1983) Establishing homologies in protein sequences. Methods Enzymol 91:524–545

Dock-Bregeon AC, Garcia A, Giege R, Moras D (1990) The contacts of yeast $tRNA^{Ser}$ with SerRS studied by footprinting experiments. Eur J Biochem 188:283–290

Eriani G, Dirheimer G, Gangloff J (1990a) Primary structure of *E. coli* AspRS. Nucleic Acids Res 18:7109–7117

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990b) Partition of aaRS into two classes on the basis of two mutually exclusive sets of sequence motifs. Nature (London) 347:203–206

Eriani G, Dirheimer G, Gangloff J (1991) CysRS: determination of the last *E. coli* aaRS primary structure. Nucleic Acids Res 19:265–269

Ferber S, Ciechanover A (1987) A role of $tRNA^{Arg}$ in protein degradation by the ubiquitin pathway. Nature (London) 326:808–811

Fehrst AR, Kaethner MM (1976) Enzyme hyperspecificity: rejection of threonine by ValRS by misacylation and hydrolytic editing. Biochemistry 15:3342–3346

Freist W, Sternbach H, Cramer F (1981) Survey on substrate specificity with regard to ATP analogs of aaRS from *E. coli* and baker's yeast. Hoppe-Seyler's Z Physiol Chem 362:1247–1254

Gampel A, Tzagoloff A (1989) Homology of aspartyl and Lysyl-tRNA synthetases. Proc Natl Acad Sci USA 86:6023–6027

Gatti DL, Tzagoloff A (1991) Structure and evolution of a group of related aaRS. J Mol Biol 218:557–568

Gribskov M, MacLachlan AD, Eisenberg D (1987) Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci USA 84:4355–4358

Haig D, Hurst LD (1991) A quantitative measure of error minimization in the genetic code. J Mol Evol 33:412–417

Hecht SM (1979) 2′OH vs 3′OH specificity in tRNA aminoacylation. In: Schimmel P, Soll D, Abelson JN (eds) tRNA structure, properties and recognition. Cold Spring Harbor Laboratory, NY, pp 345–360

Herbert C, Labouesse M, Dujardin G, Slonimski PP (1988) The NAM2 proteins from *S. cerevisiae* and *S. douglasii* are mitochondrial LeuRS and are involved in mRNA splicing. EMBO J 7:473–483

Hou Y-M, Shiba K, Mottes C, Schimmel P (1991) Sequence determination and modeling of structural motifs for the smallest monomeric aaRS. Proc Natl Acad Sci USA 88:976–980

Hountoudji C, Dessen P, Blanquet S (1986) Sequence similarities among the family of aaRS. Biochimie 68:1071–1078

Igloi GL, von der Haar F, Cramer F (1978). AaRS from yeast: generality of chemical proofreading in the prevention of misaminoacylation of tRNA. Biochemistry 17:3459–3468

Jacobo-Molina A, Arnold E (1991) HIV reverse transcriptase structure-function relationship. Biochemistry 30:351–6361

Jasin M, Regan L, Schimmel P (1983) Modular arrangement of functional domains along the sequence of AlaRS. Nature (London) 306:441–447

Kittle JD, Mohr G, Gianelos JA, Wang H, Lambowitz AM (1991) The *Neurospora* mitochondrial TyrRS is sufficient for group I intron splicing in vitro and uses the carboxy-terminal tRNA-binding domain along with other regions. Genes Dev 5:1009–1021

Lacey JC, Staves MP, Thomas RD (1991) Ribonucleic acids may be catalysts for the preferential synthesis of L-amino acid peptides: a minireview. J Mol Evol 31:244–248

Lambowitz AM, Perlman PS (1990) Involvement of aaRS and other proteins in group I and group II intron splicing. TIBS 15:440–444

Landes C, Perona JJ, Brunie S, Rould MA, Zelwer C, Steitz TA, Risler JL (1991) Primary sequence and tertiary structure similarities as evidence for a dinucleotide binding fold in class I aaRS. (submitted)

Lee CC, Craigen WJ, Muzny DM, Harlow E, Caskey CT (1990) Cloning and expression of a mammalian peptide chain release factor with sequence similarity to TrpRS. Proc Natl Acad Sci USA 87:3508–3512

Leveque F, Plateau P, Dessen P, Blanquet S (1990) Homology of LysS and LysU, the two *E. coli* genes encoding distinct LysRS species. Nucleic Acids Res 18:305–312

Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252:1162–1165

McClain WH, Foss K, Jenkins RA, Schneider J (1991) Rapid determination of nucleotides that define $tRNA^{Gly}$ identity. Proc Natl Acad Sci USA 88:6147–6151

Molina AJ, Peterson R, Yang DCH (1989) cDNA, predicted primary structure and evolving amphiphatic helix in human AspRS. J Biol Chem 264:16608–16612

Normanly J, Abelson J (1989) tRNA identity. Annu Rev Biochem 58:1029–1049
Perona JJ, Rould MA, Steitz TA, Risler JL, Zelwer C, Brunie S (1991) Structural similarities on Gln- and MetRSs suggest a common overall orientation of tRNA binding. Proc Natl Acad Sci USA 88:2903–2907
Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of a nucleotide binding domain. Nature (London) 250:194–199
Rould MA, Perona JJ, Soll D, Steitz TA (1989) Structure of the *E. coli* GlnRS-tRNA^{Gln} complex. Science 246:1135–1142
Rould MA, Perona JJ, Steitz TA S (1991) Structural basis of anticodon loop recognition by GlnRS. Nature (London) 352:213–218
Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, Rees B, Thierry J-C, Moras D (1991) Class II aaRS: crystal structure of yeast AspRS complexed with tRNA^{Asp}. Science 252:1682–1689
Sanni A, Walter P, Boulanger Y, Ebel JP, Fasiolo F (1991) Evolution of aaRS quaternary structure and activity: *S. cerevisiae* mitochondrial PheRS. Proc Natl Acad Sci USA 88:8387–8391
Schatz D, Leberman R, Eckstein F (1991) Interaction of *E. coli* tRNA^{Ser} with its cognate aaRS as determined by footprinting with phosphorothioate-containing tRNA transcripts. Proc Natl Acad Sci USA 88:6132–6136
Schimmel P (1987) AaRS: general scheme of structure-function relationships in the polypeptides and tRNA recognition. Annu Rev Biochem 56:125–158
Schimmel P, Söll D (1979) AaRS: general features and tRNA recognition. Annu Rev Biochem 48:601–648
Schoen A, Krupp G, Gough S, Berry-Lowe S, Kannangara CG, Soll D (1986) The RNA required in the first step of chlorophyll biosynthesis is a chloroplast tRNA^{Glu}. Nature (London) 332:281–284
Springer M, Graffe M, Dondon J, Grunberg-Manago M (1989) tRNA-like structures and gene regulation at the translational level: a case of molecular mimicry in *E. coli*. EMBO J 8:2417–2427
Von der Haar F, Cramer F (1976) Hydrolytic action of aaRS from baker's yeast: chemical proofreading preventing acylation of tRNA^{Leu} with misactivated Valine. Biochemistry 26:4131–4138
Webster TA, Gibson BW, Keng T, Biemann K, Schimmel P (1983) Primary structure of *E. coli* GlyRS. J Biol Chem 258:10637–10641
Webster TA, Tsai H, Kula M, Mackie G, Schimmel P (1984) Specific homology and three-dimensional structure of an aaRS. Science 226:1315–1317
Wek RC, Jackson BM, Hinnenbusch AG (1989) Juxtaposition of domains homologous to protein kinases and HisRS in GCN2 protein suggests a mechanism for coupling gene expression to amino acid availability. Proc Natl Acad Sci USA 86:4579–4583
Williamson RM, Oxender DL (1990) Sequence and structural similarities between the leucine-specific binding protein and LeuRS from *E. coli*. Proc Natl Acad Sci USA 87:4561–4565
Wong JT-F (1975) A coevolution theory of the genetic code. Proc Natl Acad Sci USA 72:1909–1912
Yarus M (1988) A specific amino acid binding site composed of RNA. Science 240:1751–1758