J Mol Evol (1995) 40:499–508

JOURNAL OF MOLECULAR EVOLUTION

The Class II Aminoacyl-tRNA Synthetases and Their Active Site: Evolutionary Conservation of an ATP Binding Site

Gilbert Eriani,[1] Jean Cavarelli,[2] Franck Martin,[1] Laurent Ador,[1] Bernard Rees,[2] Jean-Claude Thierry,[2] Jean Gangloff,[1] Dino Moras[2]

[1] UPR 9002, Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance, Laboratoire de Biologie Structurale, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15, rue René Descartes, 67084 Strasbourg, France
[2] UPR 9004, Laboratoire de Biologie Structurale, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15, rue René Descartes, 67084 Strasbourg, France

Received: 18 July 1994

Abstract. Previous sequence analyses have suggested the existence of two distinct classes of aminoacyl-tRNA synthetase. The partition was established on the basis of exclusive sets of sequence motifs (Eriani et al. [1990] *Nature* 347:203–306). X-ray studies have now well defined the structural basis of the two classes: the class I enzymes share with dehydrogenases and kinases the classic nucleotide binding fold called the Rossmann fold, whereas the class II enzymes possess a different fold, not found elsewhere, built around a six-stranded antiparallel β-sheet. The two classes of synthetases catalyze the same global reaction that is the attachment of an amino acid to the tRNA, but differ as to where on the terminal adenosine of the tRNA the amino acid is placed: class I enzymes act on the 2' hydroxyl whereas the class II enzymes prefer the 3' hydroxyl group. The three-dimensional structure of aspartyl-tRNA synthetase from yeast, a typical class II enzyme, is described here, in relation to its function. The crucial role of the sequence motifs in substrate binding and enzyme structure is highlighted. Overall these results underline the existence of an intimate evolutionary link between the aminoacyl-tRNA synthetases, despite their actual structural diversity.

Key words: Two classes of aminoacyl-tRNA synthetases — Sequence comparisons — Homology — X-ray structure — Structure–function relationships — Origin of aminoacyl-tRNA synthetases — Two ancestral molecules — Genetic code

Based on a presentation made at a workshop—"Aminoacyl-tRNA Synthetases and the Evolution of the Genetic Code"—held at Berkeley, CA, July 17–20, 1994

Correspondence to: G. Eriani

Introduction

Aminoacylation of tRNA is a crucial step in the process of translation of the genetic message. This reaction, the esterification or charging of a specific amino acid onto the corresponding tRNA, is catalyzed by a family of enzymes, the aminoacyl-tRNA synthetases (aaRS), and occurs in two steps:

$$\mathrm{ATP} + \mathrm{aa} + \mathrm{aaRS} \overset{\mathrm{Mg}^{2+}}{\leftrightharpoons} \mathrm{aaRS} \cdot (\mathrm{aa\text{-}AMP}) + \mathrm{PPi} \quad (1)$$

$$\mathrm{aaRS} \cdot (\mathrm{aa\text{-}AMP}) + \mathrm{PPi} \leftrightharpoons \mathrm{aa\text{-}tRNA} + \mathrm{aaRS} + \mathrm{AMP} \quad (2)$$

The first step of the reaction (1) consists of the amino acid activation by ATP-Mg^{2+} with formation aminoacyl-adenylate and release of PPi. The second step of the reaction (2) leads to tRNA aminoacylation and release of AMP.

Twenty of these enzymes, one per amino acid, constitute the minimum set for the protein biosynthesis. The ATP substrate is common to all aaRS whereas each enzyme is specific for the amino acid and the tRNA. More than one tRNA isoacceptor is often recognized by the enzyme; seryl-tRNA synthetase, for example, recognizes six different $tRNA^{Ser}$ isoacceptors.

A CLASS I

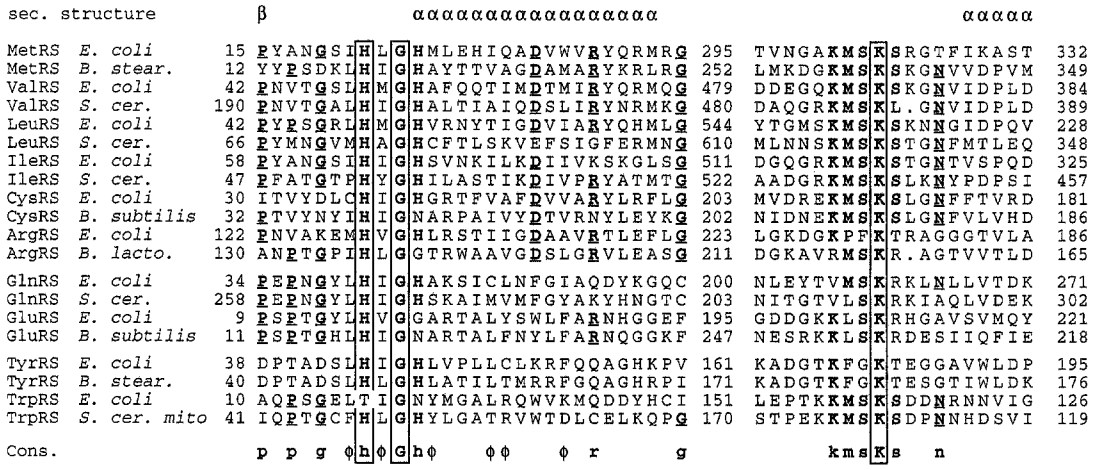

Fig. 1. Sequence alignments of **(A)** the two signature sequences of the class I synthetases and **(B)** of the three peptidic motifs found in the class II aaRS. The position of the beginning of each sequence is indicated as well as residue number between the different stretches and of the end of the protein. Invariant residues are *boxed;* strongly conserved are in *bold* and *underlined.* The *top lines* indicate the secondary structure; the *bottom lines* indicate the strongly conserved residues according to: *l* for small residues (Ala, Gly, Pro, Ser, Thr), 0 for hydrophobic residues (Cys, Val, Leu, Ile Met, Phe, Tyr, Trp), + for positively charged and polar residues (His, Lys, Arg), – for negatively charged and polar residues (Asp, Asn, Glu, Gln). The alignment **(A)** is derived of Eriani et al. (1991). The alignment **(B)** proposed is derived of Eriani et al. (1990) with the following modifications: motif 2 of AlaRS, proposed by Cusack et al. (1991), has been added after correction by Lu and Hill (1994). Motif 1 of the same enzyme has been added following Ribas de Pouplana et al. (1993). Motif 1 of PheRS and GlyRS have been corrected and added, respectively, according to the structural arguments of Mosyak and Saffro (1993). HisRS has been classified in the SerRS subgroup by virtue of significant homologies in the carboxy terminal part of the molecules (Cusack et al. 1991).

Like the RNA replicases, which were probably the first ribozymes in the RNA world, the aminoacyl-tRNA synthetases which arose early in the evolution of protein synthesis appeared to increase fidelity of protein biosynthesis by charging specifically one tRNA. They may have been the first set of functional proteins. The importance of the aminoacylation reaction can be easily understood since at the ribosome level the transfer of information used for the protein synthesis depends only on the recognition anticodon-codon of mRNA.

Because of their universal conservation we can assume that aminoacyl-tRNA synthetases arose very early, taking into account the job of the RNA catalysts affected at the same function in a primitive biosynthetic system and genetic code. Once the bases of the specificity were well established we can expect a more rapid evolution of the code. Thus, studying the aaRS relationships can bring crucial information on the mechanism of appearance, development, and expression of the genetic code.

The Apparent Structural Diversity of Aminoacyl-tRNA Synthetases

The fact that aaRS catalyze the same type of reaction and use a common strategy to activate the amino acid led the first investigators to expect some structural homogeneity in this class of proteins. However, structural studies underlined an important structural diversity with molecular weights ranging from 51,000 (CysRS, monomer) to 384,000 (AlaRS, homotetramer) (Schimmel and Söll 1979), suggesting a polymorphic adaptation of the enzymes to the specific type of recognition needed.

In the last 4 years important discoveries in the field of tRNAs and aminoacyl-tRNA synthetases have pointed the way to an understanding of the relationships between the synthetase family and the origin of the genetic code and the translation apparatus. The most significant advances come from aaRS gene-sequence analysis, crystallographic structure determination of the enzymes complexed with their substrates, and studies of the rules which govern specific recognition of tRNAs. From these different approaches has emerged a new view of the aminoacyl-tRNA synthetase family, its substrates, and the genetic code.

The Two Classes of Aminoacyl-tRNA Synthetases

For a long time the synthetase family was thought to be a unique class of enzymes whose active site was based on the Rossmann fold (Rossmann et al. 1974). This tertiary structure was observed in the first three enzymes whose crystallographic structures were solved: TyrRS, MetRS,

B

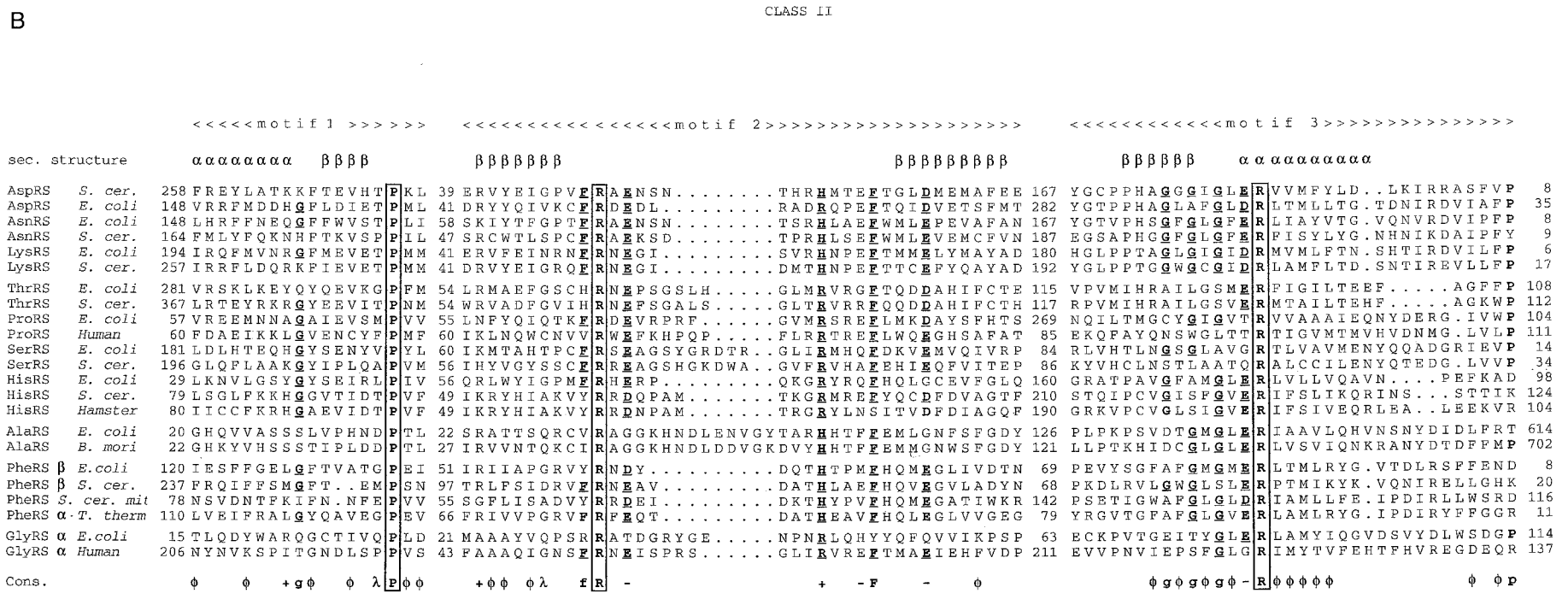

Fig. 1. Continued.

and GlnRS (Rubin and Blow 1981; Risler et al. 1981; Rould et al. 1989). These enzymes share with other aaRS two consensus functional peptides which were considered the signature sequences of the active site (HIGH and KMSKS sequences) (Webster et al. 1984; Hountondji et al. 1986) (Fig. 1, class I enzymes).

However, all attempts to unify the synthetase family were unsuccessful until the comprehensive classification of aminoacyl-tRNA synthetases into two classes of ten members each (Eriani et al. 1990), which takes into account the existence of sequence homologies in three peptidic regions of the enzymes (called motifs 1, 2, and 3 for the class II enzymes) (Fig. 1), as well as the new and specific tertiary fold observed in the crystallographic structures of seryl- and aspartyl-tRNA synthetases (Cusack et al. 1990; Ruff et al. 1991). Moreover, a functional correlation was established with the primary site of attachment of the amino acid on the 3′ terminus of the tRNA. The class I enzymes aminoacylate the 2′ hydroxyl of the terminal adenosine whereas the class II enzymes charge the 3′ hydroxyl, as measured earlier (Fraser and Rich 1975; Sprinzl and Cramer 1975). In each of the two families, the synthetases can be divided into subgroups of enzymes where homologies can be extended beyond the motifs (Eriani et al. 1990; Cusack et al. 1991; Carter, 1993) (Fig. 1).

Another point which has strengthened the classification is the oligomeric distribution of the enzymes. Most of the class I enzymes are monomeric, with the exception of MetRS (dimer in prokaryotes but monomer in eukaryotes), and TrpRS and TyrRS, two dimers which display allosteric properties. We can also find the three monomeric aaRS—ArgRS, GluRS, and GlnRS—which need tRNA to ensure the activation step of the amino acid. In contrast, most of the class II aaRS are dimers, heterotetramers, or tetramers with the exception of yeast mitochondrial PheRS (monomer) and eukaryotic AlaRS (monomer).

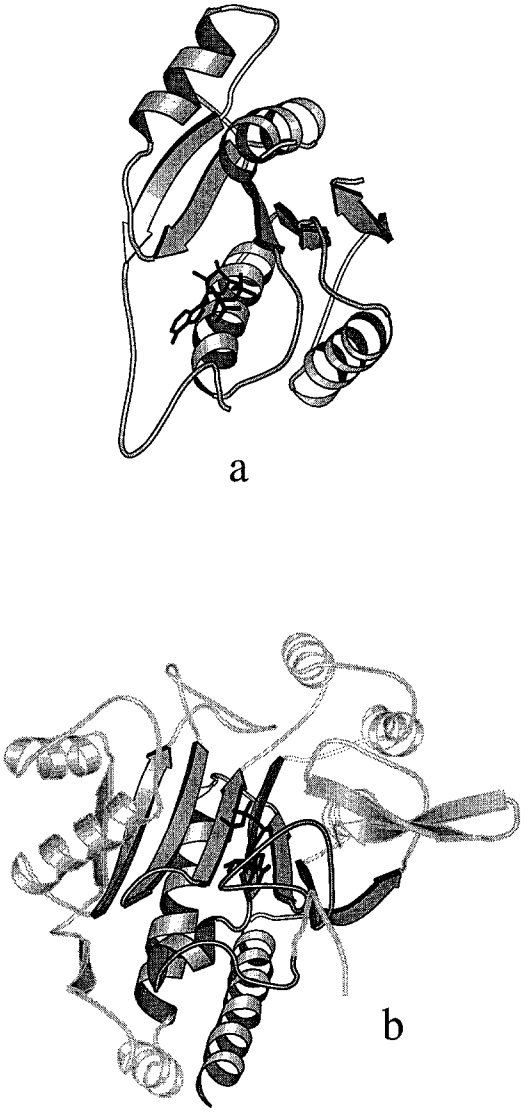

Fig. 2. The aaRS active sites with bound ATP. **a** In a class I aaRS: GlnRS. The active site is built around a parallel β-sheet surrounded by α-helices with a topology reminiscent of the nucleotide binding domain first discovered in the dehydrogenases and known as the Rossmann fold. **b** In a class II aaRS: AspRS. The core of the active site is built around a six-stranded antiparallel β-sheet partly closed by loops and helices. This and all other images were created using the program MOLSCRIPT (Kraulis 1991).

Structural Basis that Defines the Two Classes of Aminoacyl-tRNA Synthetases

The resolution of the first three crystallographic structures of aaRS (MetRS, TyrRS, GlnRS) provided the structural basis for the HIGH and KMSKS consensus sequences of the class I family. These peptides are located in the nucleotide binding fold (called the Rossmann fold) constituted of alternating α-helices and β-strands (Fig. 2). This type of structure is similar to the fold found in enzymes that bind an adenine coenzyme, such as NADH in dehydrogenases and ATP in kinases. In the three enzymes the consensus peptides are located in similar tridimensional positions and together they form the ATP binding site.

The following three structures to be solved were SerRS, AspRS, and, more recently, PheRS (Mosyak and Saffro 1993), of the class II family. These enzymes share a structure different from the Rossmann fold of class I aaRS, which is built around a central sheet of six antiparallel β-strands and two long helices (Fig. 2). The three sequence motifs are found in crucial positions of the structure: motif 1 constitutes in part the long helix located at the interface and a distorded strand, whereas motifs 2 and 3 are spatially close to form the active site of the enzyme (Cavarelli et al. 1994).

Thus, the consensus sequences which served to classify the 20 synthetases are located in defined and conserved places of the two types of tertiary structure. During evolution, additional domains were transplanted on the two basic folds, to give the actual observed diversity

in the aminoacyl-tRNA synthetase family. The main function of these domains is to select and bind the specific tRNA. Thus, GlnRS uses four globular domains to achieve binding and discrimination of $tRNA^{Gln}$, and AspRS uses one domain arranged as a β-barrel to bind the anticodon region of $tRNA^{Asp}$, whereas SerRS uses a hairpin formed of two long helices to bind the extra arm of $tRNA^{Ser}$ (Rould et al. 1989; Ruff et al. 1991; Cusack et al. 1994). Another role of these extra domains can be dimerization, as shown for *Escherichia coli* MetRS (Cassio and Waller 1971). The best illustration of the modular character of the aaRS family is certainly the eukaryotic *Drosophila melanogaster* gluprolyl-tRNA synthetase, which displays both glutamyl- and prolyl-tRNA synthetase activities as the result of the fusion of the two gene sequences (Cerini et al. 1991). Additional functions of aaRS such as splicing of RNAs can be assumed by these extra domains (Akins and Lambowitz 1987).

An Example of Class II Active Site: Yeast Aspartyl-tRNA Synthetase

Cytoplasmic AspRS from yeast is an α_2 homodimeric enzyme (Sellami et al. 1986). Its crystallographic structure has been solved in its complexed form with tRNA, aspartic acid, ATP, and AMPPcP (Ruff et al. 1991; Cavarelli et al. 1994). The overall structure of this enzyme clearly illustrates the modularity of aaRS, assigning one function to each particular domain (Fig. 3). Each monomer contains 557 amino acids, binds 1 tRNA, and can be structurally divided into 3 domains.

The N-terminal domain (residues 1–207) is built around a five-stranded β-barrel and an α-helix inserted between the third and the fourth strands. This module recognizes the anticodon loop of $tRNA^{Asp}$.

The second domain is a small globular hinge module (residues 207–241) connecting the C-terminal and the N-terminal domains and composed of four short helices. This region anchors $tRNA^{Asp}$ at the D stem level (base pair G10:U25).

The last module is the C-terminal domain formed by a six-stranded antiparallel β-sheet, partly closed by loops and helices. This domain, or core enzyme, is expected to be conserved in all class II synthetases. The three conserved motifs characteristic of class II synthetases are located in this domain. (See below.) The function of the core enzyme is to bind the acceptor stem of the tRNA and to perform the catalysis.

General Binding of $tRNA^{Asp}$

The binding of $tRNA^{Asp}$ on AspRS proceeds in a different way than that of $tRNA^{Gln}$ to GlnRS (Rould et al. 1989). The aspartyl-tRNA synthetase approaches the tRNA on the variable loop side, thus interacting with the

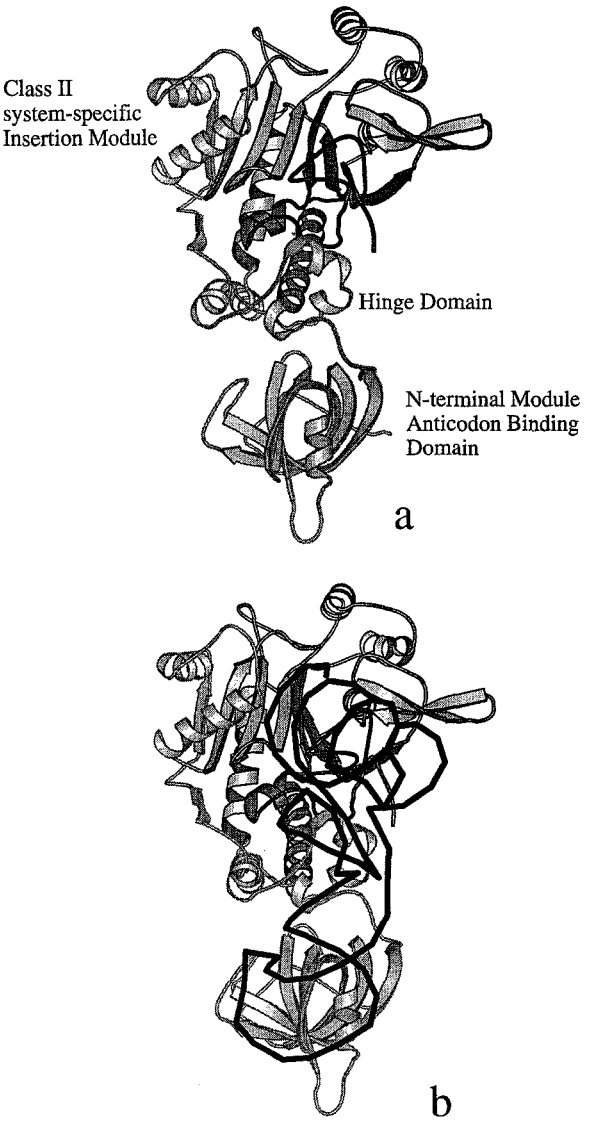

Fig. 3. Yeast AspRS structure: Representation of one monomer **(a)**, in interaction with $tRNA^{Asp}$ **(b)**. Motifs 1, 2, and 3 are shown in *dark*.

major groove side of the helical amino acid acceptor stem (Ruff et al. 1991; Cavarelli et al. 1993). Four loops clamp the acceptor stem and arm of the tRNA in the active site pocket. The variable loop of motif 2 (mainly residues 326–335) is responsible for the specific interaction in the major groove, in particular with the discriminator base G73. The correct positioning of the tRNA is assumed by contacts with the connecting domain (residues 207–241) and, as a consequence, of the anticodon stem and loop with the N-terminal domain. As opposed to SerRS, which binds its tRNA across the subunits, each monomer of AspRS binds independently a molecule of tRNA.

The anticodon binding of $tRNA^{Asp}$ is essential for specific recognition by AspRS. The anticodon loop contains four out of the six identity elements of this tRNA, the two other being located in the acceptor arm (discriminator base G73) and D stem (base pair G10:U25). It interacts with the N-terminal domain (residues 90–201)

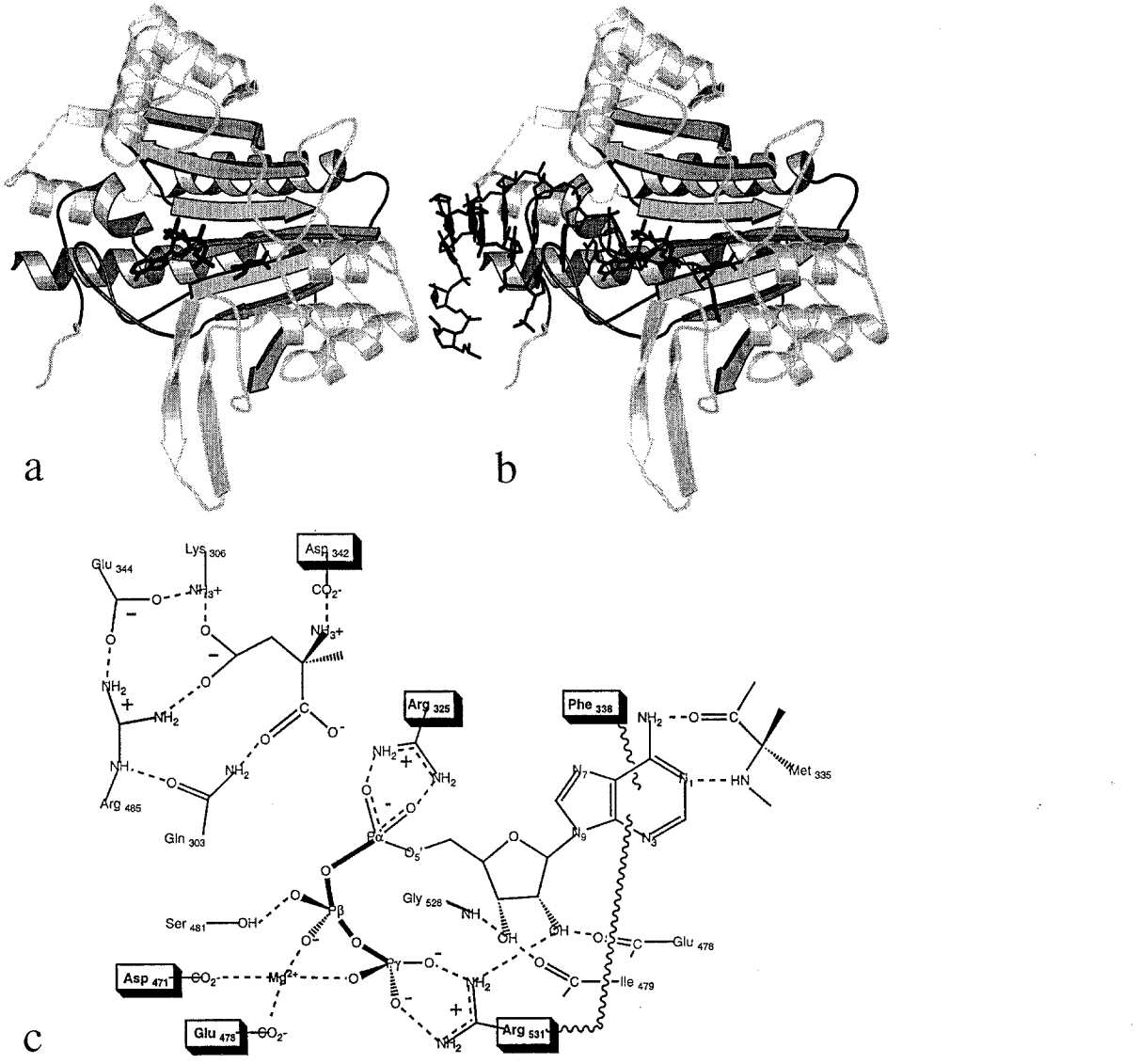

Fig. 4. **a** View of the active site of yeast AspRS, showing the ATP and aspartic acid substrate. **b** The acceptor end of $tRNA^{Asp}$ lies on top of the ATP. **c** Schematic representation of the molecular interactions of ATP and aspartic acid in their binding sites.

of the synthetase. The most characteristic feature of this interaction is the complete disruption of the anticodon loop conformation. The three bases of the anticodon protrude out in order to maximize contacts with the protein. The backbone forms a bulge at the modified purine 37 (m1G37) which is stabilized by two intramolecular H-bonds in the tRNA and hydrophobic interactions with the protein (Cavarelli et al. 1993). In this domain, five amino acids responsible for the specific interaction with the tRNA are common to the subgroup containing AspRS, AsnRS, and LysRS, three enzymes which recognize proximal anticodons (XUX) (Eriani et al. 1990; Anselme and Härtlein 1989; Lévêque et al. 1990).

While $tRNA^{Asp}$ conserves its overall topology, as seen in the crystal structure of free $tRNA^{Asp}$, the two crystal structures have demonstrated the great flexibility of the tRNA and its ability to adopt noncanonical local structures. This is illustrated by the large conformational changes in the anticodon loop and in the amino acid acceptor end. The deformations seen in the two molecules at the level of the tRNA were certainly not predictable.

Binding of ATP, Aspartic Acid, and the Acceptor end of the tRNA in the Active Site

The central β-sheet of six antiparallel strands binds the small substrates (ATP, aspartic acid) and the acceptor end of the tRNA (Fig. 4). The three motifs defined originally by sequence comparison are located in this part of the molecule. Motif 1 (residues 258–275) is located at the interface of the subunits and forms an α-helix and a distorted β-strand. The invariant proline residue of this

motif is located in this strand at an apparently insignificant position. However, substitution of this residue shows its crucial function in catalysis, through interface structure and communication between the monomers (Eriani et al. 1993). The crucial function of this residue and its invariance underline the importance of the oligomerization of the class II aaRS, all multimeric enzymes with the sole exceptions of eukaryotic AlaRS and mitochondrial PheRS. The two other conserved motifs are implied in ATP binding. Motif 2 (residues 316–349 with invariant Arg325) is composed of two antiparallel strands connected by a large loop. Motif 3 (residues 517–548, with invariant Arg531) is constituted by a strand followed by an α-helix. We must also note the presence of two strongly conserved residues of motif 2 (Phe324 and Glu337) at the dimer interface. Like Pro273 of motif 1 they can structure in a functional way the interface of the dimeric class II enzymes.

The crystallographic results show that the ATP molecule lies on the β-sheet (Fig. 4). The adenine ring is firmly held by π-electron interactions with Phe338 on one side and Arg531 on the other side. Arg325, a strictly invariant class II residue, forms a salt-bridge interaction with the α-phosphate of the ATP. The ATP molecule adopts a bent conformation, stabilized by the interaction of Arg531 with γ-phosphate and the presence of a magnesium ion located between the β-phosphate and the γ-phosphate. Two acidic residues, invariant in class II synthetase—namely, Glu478 and Asp471—are part of the magnesium binding site. Substitution of anyone of these residues leads to important decreases of the catalytical parameters with important losses of ATP affinity. (See Table 2 and Cavarelli et al. 1994 for more details.)

The aspartic acid binding site is formed by a pocket mapped by amino acids conserved in aspartyl-tRNA synthetase enzymes from different origins. The side-chain functional group of aspartic acid is recognized by a network of interactions specific to the aspartic system (Cavarelli et al. 1994). The only conserved class II-type interaction involves the residue Asp342 which should mediate the α-amino group recognition of the substrate.

The interaction of the acceptor stem of the tRNA and the fine positioning of the CCA require interactions specific to the aspartic system. However, two residues of motif 2 (Glu327 and His334) may be involved in tRNA binding as well as ATP binding. In its reactive position the CCA arm of $tRNA^{Asp}$ presents the 3′ hydroxyl to the aminoacyl adenylate. This final positioning is in large part due to the mode of entry of the CCA end in the active site of the enzyme, and this could be responsible of the distinction 2′ and 3′ of the two synthetase classes.

Class II aminoacyl-tRNA synthetases are built around a central active site made up by a six-stranded antiparallel β-sheet which is the framework for a new ATP binding domain. The dimeric organization of this domain seems to be necessary, as suggested by the residue conservation of motif 1 and some of motif 2. The amino acids involved in ATP binding are highly conserved residues belonging to motifs 2 and 3, common to all class II aaRS. Attached to the central core, the class II aaRS have extra or inserted domains which perform specific functions in each system: anticodon binding or recognition of other parts of the tRNA molecule.

Functional and Evolutionary Considerations

The 20 actual aminoacyl-tRNA synthetases probably evolved from two ancestral molecules, the ancestors of the Rossman fold (class I) and of the antiparallel β-sheet (class II). The equal distribution into two families of ten enzymes suggests a parallel evolution.

The fact that the enzymes have conserved their specific aminoacylation site (except Phe) suggests a functional importance for the site. However, once the synthesis of the aa-tRNA is performed, a rapid isomerization between the two sites occurs with a half-time for migration of approximately two-tenths of millisecond (Griffin et al. 1966). During protein synthesis, only the 3′ species are carried to the ribosome by the elongation factor Tu, which seems to block the isomerization in the 3′ position.

The existence of two different sites of aminoacylation may be related to the proofreading mechanisms observed in some cases. Some synthetases are able to hydrolyze the incorrect aa-tRNA when fixed on the hydroxyl which doesn't serve as a primary binding site (von der Haar and Cramer 1976). This hydrolytic function is called posttransfer proofreading and is used by the enzyme to correct the nonspecifically aminoacylated tRNA once the isomerization has occurred. The existence of a second binding site for the correction remains to be demonstrated. Proofreading has been well studied for some enzymes of class I which aminoacylate hydrophobic amino acids—ValRS and IleRS—where it plays a pivotal role in accuracy of the aminoacylation reaction (Igloi et al. 1978; Freist et al. 1987). For other enzymes like ArgRS and TyrRS such reaction seems to be marginal, the pretransfer proofreading reaction being the main correction step (Freist et al. 1989; Freist and Sternbach 1988).

The conservation of the two distinct aminoacylation sites may support the necessity of having a correction mechanism of the acylated product. This distinction certainly appeared very early since the distribution of the sites matches perfectly the partition of the two classes (PheRS excepted). Early in the development of the genetic code and protein biosynthesis we can suppose that the primordial aaRS were not perfectly adapted to the amino acid recognition, and then used more often the posttransfer proofreading to increase fidelity. Now, the biosynthetic apparatus uses synthetases with very evolved structures which can sometimes discriminate almost perfectly for the correct amino acid (example, TyrRS; Freist and Sternbach 1988).

The amino acid specificity distribution (Table 1) re-

Table 1. Classification of the 20 aminoacyl-tRNA synthetases into two classes of ten members[a]

Class I synthetases (HIGH + KMSKS sequences charge of 2′ OH)				Class II synthetases (3 Peptidic motifs, charge of 3′ OH)			
CysRA	α		X G X	AspRS	α2		X A X
ValRS	α		X U X	AsnRS	α2		X A X
LeuRS	α		X U X	LysRS	α2		X A X
IleRS	α		X U X	SerRS	α2		X C X
MetRS	α2	(α)	X U X	ThrRS	α2		X C/ G X
ArgRS	α		X G X	ProRS	α2		X C X
TrpRS	α2		X G X	HisRS	α2		X A X
TyrRS	α2		X U X	AlaRS	α4 (α)		X C X
GluRS	α		X A X	GlyRS	α2β2	(α2)	X G X
GlnRS	α		X A X	PheRS[b]	α2β2	(α, α2)	X U X

[a] The partition takes into account the existence of sequence homologies and site of attachment of the amino acid on the acceptor hydroxyl of the tRNA. The oligomeric structures of the enzymes as well as the second base of the codon are mentioned.

[b] PheRS aminoacylate on the 2′ OH of $tRNA^{Phe}$

mains also an interesting puzzle of diversity. No absolute correlation between this partition and the genetic code and tRNA mode of recognition has been detected; however, concerning the physico-chemical properties of the amino acids and their distribution in Table 1, some observations can be made:

1. The hydrophobic amino acids are mainly charged by the class I enzymes. This could reflect a specialization of the active site of the class I enzymes on hydrophobic recognition which involves hydrophobic and van de Waals interactions. On the other hand, the small polar residues are preferred by the class II enzymes which are in the same subgroup of enzymes.
2. The charged residues are equally distributed in the two classes, but with a preference of class I enzymes for the largest amino acid when homologous enzymes are considered (example, Glu by class I whereas Asp by class II; Arg by class I whereas His by class II). Finally, the recognition of Tyr is assumed by a class I enzyme whereas Phe (the smaller) is assumed by a class II enzyme.

Thus it seems that the specialization of the classes for some homologous amino acids is not a general rule since an equal distribution in the two classes is sometimes observed. Nature could have selected two different binding sites in order to improve discrimination between related substrates differing only by size.

The actual observed partition suggests an ordered distribution of the specificities whose mechanism could have begun at the origin of life. Considering the first prebiotic amino acids we can see that Val, Leu, and Ile could have been selected by class I synthetases (XUX codons) whereas Ala, Gly, Asp, and Ser could have been selected by class II synthetases (XAX, XCX, and XGX codons); aaRS probably appeared before the existence of the 20 amino acids. When the new biosynthetic pathways appeared, new aaRS specificities had to be generated, probably by divergence from the first enzymes, as shown by the existence of only two aaRS classes. From what enzyme they originated remains an important question. The Rossman fold has been found in numerous enzymes which bind adenine coenzymes, the central sheet of the class II enzymes has only been found recently in the structure of another protein, the biotine synthetase/repressor protein (BirA), an enzyme which proceeds through acyl-adenylate intermediate (Artymiuk et al. 1994).

The amino acid allocation in the genetic code raises some questions and remarks on its parallel evolution with the aaRS family. What were the relationships between amino acid appearances, codons, and aaRS assignment? To be incorporated in the genetic code, the new amino acids had to take codons (and corresponding tRNAs) and to acquire an aminoacyl-tRNA synthetase specificity. To minimize the effect of such misincorporations in the biosynthesis the codons for the new amino acids were probably taken to the more related one (Sonneborn 1965; Woese 1965), which in turn could favor synthetase recognition since the physicochemical properties of the amino acids would have been similar. The subgroup of hydrophobic residues (codons XUX) and small polar residues (codons XCX) would support, in part, this theory. However, this phenomenon of distance minimization between physicochemical properties and codons would have played only a subsidiary role in the evolution of the genetic code.

Another hypothesis, called coevolution of the genetic code and biosynthetic apparatus, suggests that the organization of the genetic code was primarily determined by the mechanism of codon concession by the precursor to product amino acid (Wong 1975). In this model the first prebiotic amino acids could have served as precursors for the formation of other amino acids along prebiotic pathways which became the amino acid biosynthetic pathways of the present-day organisms. Thus the genetic code would represent a map of the biosynthetic relationships between amino acids. Concession of the codons would have required competition of the new product for tRNA of the precursor and this would have linked codons and aaRS. This theory finds support in the subgroup Asp, Asn, and Lys, which share proximal codons and biosynthetic origins (Gatti and Tzagoloff 1991) as well as in the Glu and Gln subgroup (Guildo 1993).

However, these hypotheses only partially explain codon and synthetase distribution. The first attribution of codons to amino acids remains a vast subject of discussion (as well as the attribution to one of the two synthe-

Table 2. Summary of the most significant results of mutation of the conserved residues of the three motifs of yeast aspartyl-tRNA synthetase[a]

Residue mutated	Effect on charging rate	Effect on activation rate	Effect on ATP affinity	Effect on Asp affinity	Effect on tRNA affinity	Interaction observed function
Motif 1		258 F R E Y L A T K K F T E V H T **P** K L				
				273		
Pro 273	+	+	no	no	no	Structure of the interface Communication between subunits
Motif 2		315 E R V Y E I G P V **F** **R** A **E** N S N K L . . . . . T H R **H** M T **E** **F** T G L **D** M E M A F E				
			324 325	327		334 337 338 342
Phe 324	+	+	+	no	nd	Structure of the interface
Arg 325	+++	+++	+++	+++	no	Binding of α-phosphate of ATP
Glu 327	++	++	+	++	no	Interaction with ATP and tRNA
His 334	+	+	+	+	+	Interaction with tRNA and ATP
Glu 337	+	+	+	+	no	Structure of the interface
Phe 338	++	++	+++	++	no	Stacking on the adenine part of ATP
Asp 342	+++	++	+	+++	no	Binding of amino part of Asp
Motif 3		517 Y G C P P H A **G** G **G** I **G** L **E** **R** V V M F Y L D . . L K I R R A S F V P				
			524 526 528	530 531		
Gly 524	inactive	inactive	nd	nd	nd	Asp and ATP binding surface
Gly 526	inactive	++	+++	+++	no	Asp and ATP binding surface
Gly 528	+	++	+++	+	nd	Interaction with ATP
Glu 530	+	+	+	+	nd	Structure of the active site
Arg 531	inactive	inactive	nd	nd	no	Binding of γ-phosphate of ATP Stacking on the adenine moiety of ATP

[a] The effects on the $tRNA^{Asp}$ charging rate and ATP-dependent aspartic acid activation rate are expressed in the following terms: + moderate decrease, ++ important decrease, +++ drastic decrease. The effects on substrate affinities were estimated by Km measurements (ATP and aspartic acid) and Kd measurements ($tRNA^{Asp}$). The three motifs are described according to Fig. 1; nd, not determined. Most of the data compiled in this table are from Cavarelli et al. (1994) and Eriani et al. (1993)

tase families). The primitive enzymes were probably polyspecific and adapted more to a physicochemical type than to specific side chains. From these ancestral enzymes the actual synthetases probably diverged by a cascade of pressures and selections that we are trying now to understand. At this step, study of the origins of aaRS converges with the study of the origins of the genetic code. Understanding this evolution is the next challenge for molecular biology.

Acknowledgments. We are grateful to John Arnez for correction of the manuscript.

References

Anselme J, Härtlein M (1989) Asparaginyl-tRNA synthetase from Escherichia coli has significant sequence homologies with yeast aspartyl-tRNA synthetase. Gene 84:481–485

Akins RA, Lambowitz AM (1987) A protein required for splicing group I introns in Neurospora crassa mitochondria is mitochondrial tyrosyl-tRNA synthetase or a derivative thereof. Cell 50:331–345

Artymiuk PJ, Rice DW, Poirrette AR, Willet P (1994) A tale of two synthetases. Nature Structural Biology 1:758–760

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

Cassio D, Waller JP (1971) Modification of methionyl-tRNA synthetase by proteolytic cleavage and properties of the trypsin-modified enzyme. Eur J Biochem 20:283–300

Cavarelli J, Eriani G, Rees B, Ruff M, Boeglin M, Mitschler A, Martin F, Gangloff J, Thierry JC, Moras D (1994) The active site of yeast aspartyl-tRNA synthetase: structural and functional aspects of the aminoacylation reaction. EMBO J 13:327–337

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

Carter G (1993) Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases. Annu Rev Biochem 62:715–748

Cerini C, Kerjan P, Astier M, Gratecos D, Mirande M, Semeria MA (1991) Component of the multisynthetase complex is a multifunctional aminoacyl-tRNA synthetase. EMBO J 10:4267–4277

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

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

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

Eriani G, Dirheimer G, Gangloff J (1991) Cysteinyl-tRNA synthetase: determination of the last E. coli aminoacyl-tRNA synthetase primary structure. Nucleic Acids Res 19:265–269

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

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

Freist W, Sternbach H, Cramer F (1987) Isoleucyl-tRNA synthetase from baker's yeast and from Escherichia coli MRE600. Eur J Biochem 169:33–39

Freist W, Sternbach H (1988) Tyrosyl-tRNA synthetase from baker's yeast. Order of substrate addition, discrimination of 20 amino acids in aminoacylation of $tRNA^{Tyr}CCA$ and $tRNA^{Tyr}CCA(3'NH2)$. Eur J Biochem 177:425–433

Freist W, Sternbach H, Cramer F (1989) Arginyl-tRNA synthetase from yeast. Discrimination between 20 amino acids in aminoacylation of $tRNA^{Arg}CCA$ and $tRNA^{Arg}CCA(3'NH2)$. Eur J Biochem 186:535–541

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

Griffin BE, Jarmen M, Reese CB, Sulston JE, Trentham DR (1966) Some observations relating to acyl mobility in aminoacyl soluble ribonucleic acids. Biochemistry 5:3638–3649

Guildo MD (1993) Origin glutaminyl-tRNA synthetase: an example of palimpset? J Mol Evol 37:5–10

Hountondji C, Dessen P, Blanquet S (1986) Sequence similarities among the family of aminoacyl-tRNA synthetases. Biochimie 68: 1071–1078

Igloi GL, von der Haar F, Cramer F (1978) Aminoacyl-tRNA synthetases from yeast: generality of chemical proofreading in the prevention of misaminoacylation of tRNA. Biochemistry 17:3459–3467

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

Lévêque F, Plateau P, Dessen P, Blanquet S (1990) Homology of lysS and lysU, the two Escherichia coli genes encoding distinct lysyl-tRNA synthetase species. Nucleic Acids Res 18:305–312

Lu Y, Hill KAW (1994) The invariant arginine of motif 2 of Escherichia coli alanyl-tRNA synthetase is important for catalysis but not for substrate binding. J Biol Chem 269:12137–12141

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

Ribas de Pouplana L, Buechter DD, Davis MW, Schimmel P (1993) Idiographic representation of conserved domain of a class II tRNA synthetase of unknown structure. Protein Sci 2:2259–2262

Risler JL, Zelver C, Brunie S (1981) Methionyl-tRNA synthetase shows the nucleotide binding fold observed in dehydrogenases. Nature 292:385–386

Rossmann MG, Moras D, Osen KW (1974) Chemical and biological evolution of a nucleotide binding domain. Nature 250:194–199

Rould MA, Perona JJ, Söll D, Steitz TA (1989) Structure of E. coli glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2,8 Å resolution. Science 246:1135–1142

Rubin J, Blow D (1981) Amino acid activation in crystalline tyrosyl-tRNA synthetase from Bacillus stearothermophilus. J Mol Biol 145:489–500

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

Schimmel P, Söll D (1979) Aminoacyl-tRNA synthetases: general features and recognition of transfer RNAs. Ann Rev. Biochem 48: 601–648

Sellami M, Chatton B, Fasiolo F, Dirheimer G, Ebel JP, Gangloff J (1986) Nucleotide sequence of the gene coding for yeast cytoplasmic aspartyl-tRNA synthetase (APS); mapping of the 5′ and 3′ termini of AspRS mRNA. Nucleic Acids Res 14:1657–1666

Sonneborn TM (1965) In: Bryson V, Vogel HJ (eds) Evolving genes and proteins. Academic Press, New York, pp 337–397

Sprinzl M, Cramer F (1975) Site of aminoacylation of tRNAs from Escherichia coli with respect to the 2′- or 3′-hydroxyl group of the terminal adenosine. Proc Natl Acad Sci USA 72:3049–3053

von der Haar F, Cramer F (1976) Hydrolytic action of aminoacyl-tRNA synthetases from baker's yeast: "chemical proofreading" preventing acylation of $tRNA^{Ile}$ with misactivated valine. Biochemistry 15:4131–4138

Webster TA, Tsai H, Kula M, Mackie GA, Schimmel P (1984) Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. Science 226:1315–1317

Woese C (1965) The genetic code. Harper and Row, New York, pp 156–160

Wong JTF (1975) A co-evolution theory of the genetic code. Proc Natl Acad Sci USA 72:1909–1912