4
tRNA and Synthetases

4.1
tRNA: Structure and Function

Viter Marquéz and Knud H. Nierhaus

4.1.1
Introduction

Even before deciphering the genetic code during the 1960s, Francis Crick had postulated, in 1956, that protein synthesis is mediated by "adaptor" RNA molecules [1]. Two years later, Hoagland et al. [2] discovered a nucleic acid fraction of low molecular weight that served as a carrier for amino acids, transporting them to the place where polypeptides are synthesized. This fraction was termed soluble RNA (sRNA) and was described as mixture of components, each with an adaptor ability for a particular amino acid [3]. Nowadays, we know that the sRNA or adaptor molecules are the transfer RNAs (tRNAs) and they are the linking factors between the RNA world, deciphering the triplet code or codon encoded in the messenger RNA (mRNA), and the protein world, because they carry amino acids to the ribosomes where they are linked together to form proteins. The specificity of deciphering results from the fact that tRNAs contain at one tip of their L-shaped tertiary structure an anticodon complementary to a specific codon and at the other tip the corresponding aminoacyl residue linked by an energy-rich ester bond ($\Delta G^{\circ\prime} = \sim$ -6 kcal mol^{-1}). It follows that the charging of a tRNA with its amino acid is the true translation step, whereas it is the astounding task of the ribosome to translate the sequence of codons of an mRNA into the corresponding protein sequence in a fast and accurate fashion (see Chaps. 8.2 and 8.3). Charging of tRNAs is performed by synthetases (aaRS), and all tRNAs that can carry the same amino acid (isoacceptor tRNAs) are usually recognized by one and the same enzyme (see the next chapter).

Many milestones in molecular biology were achieved as a product of tRNA research:

(1) binding assays of tRNA to ribosomes were essential for deciphering the genetic code [4];
(2) $tRNA^{Ala}$ from yeast was the first nucleic acid molecule, whose complete sequence was determined [5];

Protein Synthesis and Ribosome Structure. Edited by K. H. Nierhaus and D. N. Wilson

ISBN 3-527-30638-2

(3) $tRNA^{Phe}$ from yeast was the first RNA molecule, for which the crystal structure could be resolved [6, 7];
(4) a tRNA gene ($tRNA^{Ala}$ from yeast) was the first chemically synthesized gene that showed activity *in vivo* [8];
(5) the structural motif called a pseudoknot was first described in studies of tRNA-like structures.

In fact, for a long time tRNAs were the only RNA molecules that could be produced in large amounts and be obtained in homogenous form; thus many of the techniques used nowadays in the study of RNA and RNA–protein interactions were developed using tRNAs [9]. Additionally, the tRNA primary sequence carries information about the age of the genetic code. By comparison of sequence alignments of tRNAs, applying a method called "statistical geometry in sequence space", it was possible to draw conclusions about phylogeny (common progenitor of the kingdoms) and the origin of the genetic code. With the assumption that the kingdom separation occurred around 2.5±0.5 billion years ago, the age of the genetic code was determined to 3.8 (±0.6) billion years [10]. Now the rRNAs have replaced the tRNAs as the preferential tool to study evolutionary relationships.

The central role of tRNAs is to carry amino acids to the ribosome, but even beyond this adaptor function, tRNAs also dictate the functional states of the ribosome. The tRNA locations on the ribosome define the pre- and post-translocational states, namely the two main conformations between which ribosomes oscillate during the elongation cycle (see Chap. 8).

4.1.2 Secondary Structure

Three main species of RNA molecules exist in all living cells: rRNA, mRNA and tRNA. tRNA constitutes only 10–15% of the total RNA in *Escherichia coli*, each tRNA has a molecular weight of about 25 kDa and a relative sedimentation coefficient of 4S giving rise to the original name "4S RNA". The size of tRNAs is variable, but on average they have a length of 76 nucleotides (nt), the longest tRNA identified so far is $tRNA^{Ser}$ from *E. coli* having 93 nt whereas in nematode mitochondria very short cripple tRNAs are found lacking either the D or the TΨC stem loop (about 56 nt, see Ref. [11]).

Holley et al. [5] proposed several possible secondary structures, when the first tRNA sequence of $tRNA^{Ala}$ from yeast was obtained. However, when the yeast tRNA sequences of $tRNA^{Ser}$ [12], $tRNA^{Tyr}$ [13], and $tRNA^{Phe}$ [14] also became available, a planar cloverleaf secondary structure was the only one that best satisfied all sequences (Fig. 4.1-1). Sequence comparison of other tRNAs revealed that all tRNAs adopt the cloverleaf secondary structure. Figure 4.1-2(a) shows a graphic representation of the distribution of conserved bases in elongator tRNAs [15] on the basis of 932 sequences. Figure 4.1-2(b) shows the generally accepted numbering of the tRNA nucleotides [16].

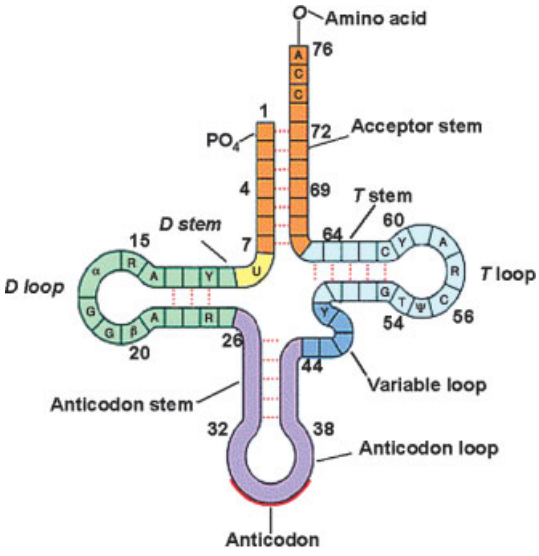

Figure 4.1-1 Cloverleaf secondary structure of tRNA. The various secondary structure motifs are shown in different colors. Abbreviations: R, purine base (A or G); Y, pyrimidine base (C or U); T, ribothymidine; Y, pseudouridine. Taken from Ref. [52] with permission.

The cloverleaf structure is characterized by three stem loops and four helices. The acceptor helix (termed acceptor stem) is formed by seven base pairs combining the 5′ and 3′-end of the molecule. In all elongator tRNAs, the first base pair of the acceptor stem is a G1-C72 Watson–Crick interaction. However, in eubacterial initiator $tRNA^{fMet}$, a mismatch between residue C1 and A72 is observed. This stem contains, at the 3′-end, a single-strand sequence N^{73}CCA-3, where N73 could be any nucleotide representing a "discriminator" base important for the specificity and efficiency of tRNA synthetase activity [17]. The universally conserved CCA sequence carries at its 3′-end the aminoacyl residue via an ester link between the ribose 2′- or 3′-OH group of the carboxyl group of the amino acid. Ten synthetases are linking "their" amino acids to the 2′-end of the ultimate A, the other 10 at the 3′-end (see Chap. 4.2). This has no direct consequence to protein synthesis, since the aminoacyl residue can trans-esterify between 2′- and 3′-OH groups with a rate of 1–10 s^{-1} [18], which might be increased and fixed at the 3′ position by elongation factor EF-Tu·GTP (M. Sprinzl, pers. comm.). This factor carries an aminoacyl-tRNA to the A-site of a ribosome (see Chap. 8). On the ribosome, the aminoacyl and peptidyl residues are linked to the 3′-OH group via an ester bond; the 2′-OH group is essential at least for the translocation of tRNAs on the ribosome from A- to the P-sites as well as from the P- to the E-sites [19, 20].

The CCA ends are the docking sites of tRNAs at the A- and P-site regions of the ribosomal peptidyl-transferase center on the 50S subunit (see Chap. 8.3). Eubacteria

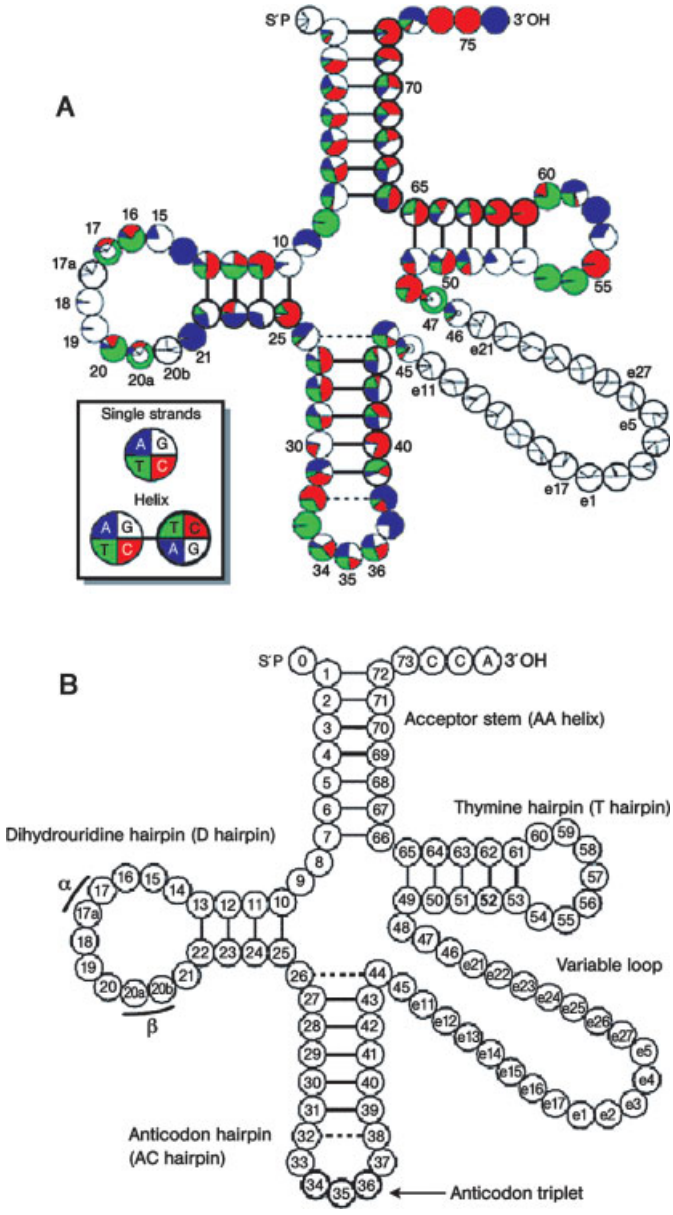

Figure 4.1-2 Features of the tRNA secondary structure. (A) Cloverleaf secondary structure of tRNA showing the distribution and conserved position of nucleotides found in all tRNA sequences known. The nucleotides are colored (A, blue; G, white; T(U), green; C, red) and their occurrence are indicated by the fractions of the circle area. According to Ref. [53] with permission, modified. (B) Conventional numbering of nucleotides according to Ref. [16], with permission.

such as *E. coli* often encode the 3′-terminal CCA in the tRNA genes, whereas in most organisms the CCA end is added post-transcriptional with a ATP (CTP): tRNA nucleotidyl-transferase or CCase [21], thus representing the most common editing mechanism. In all cells, the ATP (CTP):tRNA nucleotidyl-transferase is an essential enzyme, since it functions also in the repair of damaged CCA ends.

The second stem-loop structure is the D stem loop, where the helical region consists of 3–4 base pairs and the loop, between 8 and 11 nucleotides. The loop contains two hydrouridine bases, hence the name dihydrouridine-stem-loop for this substructure.

The anticodon stem loop at the opposite end of the molecule to the acceptor stem contains the anticodon in the middle of its loop. The loop has a universal length of 7 nt with a consensus sequence Py_{32}-U_{33}-XYZ-Pu (modified)-N_{38}, where Py represents a pyrimidine, XYZ is the anticodon, Pu a purine base and N any nucleotide. The stem always contains 5 bps and the nucleotide at position 33 of the anticodon loop is a universally conserved U in all tRNAs.

Like the anticodon stem loop, the fourth helix also comprises 5 bps with a 7 nt loop. The loop contains the sequence TΨC that gives rise to the name "TΨC-loop", where T stands for ribose-thymidine and Ψ for pseudouridine.

In addition to the defined substructures described above, a variable region exists between the T-loop and the anticodon loop, which can be anywhere between 4 and 24 nts. According to the length of this variable loop, the tRNAs have been classified (not extremely useful) as class I (4–5 nts, the vast majority) and class II (10–24 nts, $tRNA^{Leu}$, $tRNA^{Ser}$, $tRNA^{Tyr}$ in eubacteria and some organelles).

4.1.3 Tertiary Structure

The crystal structures at 3 Å resolution of yeast $tRNA^{Phe}$ [6, 7] and later $tRNA^{Asp}$ [22] confirmed that the tRNA molecule adopts an L shape. It is product of a double coaxial stacking between the acceptor stem with its CCA end and the T stem loop forming the short arm of the L arm, and the anticodon stem loop and the D stem-loop forming the long arm (Fig. 4.1-3A). In this way, the cloverleaf structure of the tRNA is transformed into two main domains: the acceptor and anticodon arms, respectively (Fig. 4.1-3B), enclosing an angle of about 90°. The extremities of both domains represent the functional "hot spots" of tRNAs: at the tip of the acceptor arm the amino acid is covalently attached, whereas at the tip of the anticodon arm, the anticodon is located, and are thus separated from each other by a distance of 75–80 Å. This is precisely the distance from the decoding center on the 30S ribosomal subunit to the peptidyl-transferase center on the 50S ribosomal subunit. Since a rigid and straight rod-shaped molecule could also fulfill the distance requirement, it raises the question as to why tRNAs have a universally conserved L shape?

The answer becomes clear when we consider other functions of the tRNAs on the ribosome. We will see in the translocation chapter (Sect. 8.4), where there is good evidence that during translocation the tRNAs are the handle to move the $tRNA_2$•mRNA complex, and thus there is a need to link the mRNAs tightly to the tRNAs via two

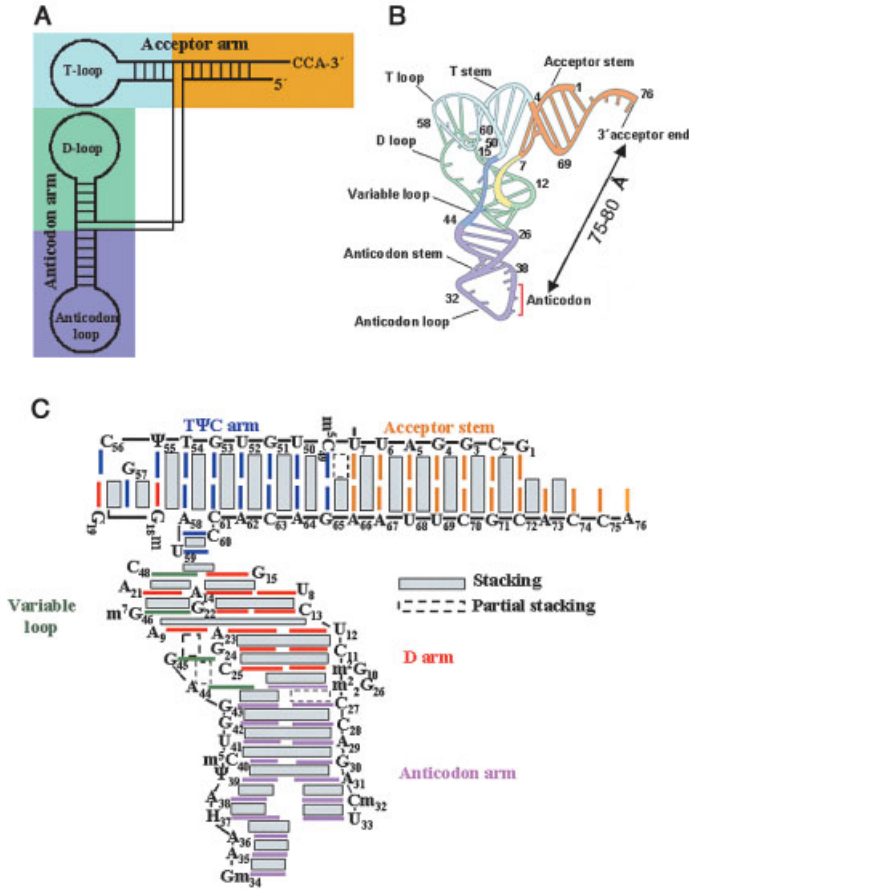

Figure 4.1-3 Features of the tRNA tertiary structure. (A) Double coaxial stackings between the anticodon stem loop and D-stem loop (anticodon arm) and between the acceptor stem and the T-stem loop (acceptor arm) transform the cloverleaf secondary structure of a tRNA into a two domain structure that include an angle of about 90°. According to Ref. [30] with permission, modified.
(B) The functional hot spots of a tRNA are the anti-codon and the CCA-3′-end that are separated by 75–80 Å in almost all canonical L-shape tRNAs. The aminoacyl or peptidyl residue is linked to the CCA-3′-end via an ester bond. This precise architecture allows codon–anticodon interaction at the decoding center of the small ribosomal subunit and contacts of the aminoacyl-(peptidyl) residue with the peptidyl transferase center on the large ribosomal subunit. From Ref. [47] with permission, modified. The colors correspond to those in (A).
(C) Base stacking (gray blocks) is an important element for the stability of the tertiary L-shaped tRNA. Colored lines indicate base interactions (mostly base pairs), the colors indicate secondary-structure motifs such as acceptor stem or D arm. Four bases (16, 17, 20, and 47) that are not involved in base stacking are not included in the figure for the sake of clarity. According to Ref. [54] with permission, modified.

adjacent codon–anticodon interactions. However, since a tRNA has a diameter of 20 Å (that of a double helix) but a codon length of only ~10 Å, it is immediately apparent that it is the L shape that prevents a steric clash of the tRNA bodies and allows simultaneous codon–anticodon interaction of both tRNAs at one end and at the same time a neighborhood of the CCA ends at the A- site and P-site regions of the peptidyl-transferase center at the other end. In fact, an angle of about 40° has been detected between the planes defined by the two L-shaped tRNAs at A- and P-sites [23–25] and about 140° between the A- and P-site codons [26].

A tRNA in solution is rigid and stable – features that are more typical for a protein than for an RNA of 76 nt. Aside from the Watson–Crick interactions seen in the cloverleaf structure and the base-stacking effects of the helices, a multitude of unusual interactions are involved in establishing the tertiary structure of a tRNA. One important feature at the tRNA elbow is the interspersed base stacking between the two main domains of the tertiary structure of tRNA: the D-arm is participating with two bases G18 and G19 in the coaxial stacking of the acceptor and the TΨC arms, and the reverse is also true: the TΨC arm participates with C60 and U59 ($tRNA^{Phe}$) in the coaxial stacking of the anticodon and D-arms (Fig. 4.1-3C). Other unusual features are listed below:

(1) At the junction of T- and D-loops, the elbow of the tRNA, a cross-strand stacking of four purines form a "base zipper structure" [27], which is important for stabilizing the "elbow architecture" of a tRNA (see Fig. 4.1-4A).

(2) Non-canonical Watson–Crick base pairs such as G:U are quite common in RNA helices (see Fig. 4.1-4B), but also more rare interactions such as the iminoG26:A44 base pair are observed (see Fig. 4.1-4C). Hoogsteen base pairs are formed when the nitrogen N7 of the imidazole ring of the adenine base is involved in the hydrogen-bonding interaction instead of the pyrimidine edge as in the normal Watson–Crick base pairing. These non-canonical base pairs are observed between the conserved residues U8 and A14 organizing the sharp turn of the D-loop (see Fig. 4.1-4D). Another example is the base pairing between the universally conserved ribothymidine residue T54 and A58 (see Fig. 4.1-4E).

(3) Non-Watson–Crick base pairs from residues located in a single-stranded region with a base pair within a helix, form base triplet interactions. For example, residue G45 of the variable loop interacts with the base 10 of base pair m^2 G10-C25 of $tRNA^{Phe}$ from yeast (Fig. 4.1-4F).

(4) The sugar-phosphate backbone participates in complex interactions to hold the tRNA compact. Single-stranded regions in the tRNA can adopt the C2′-endo configuration, although the C3′-endo configuration is found in the helical regions (A-form) [28]. Furthermore, the 2′-OH of the ribose and the oxygen of the anionic phosphates groups are involved in a series of hydrogen bonds, which contribute to the stability of the tRNA tertiary structure. For instance, the N1 of the conserved base A21 interacts with the 2′-OH of the highly conserved U8 and the 2′-OH of the residue A58 donates a proton to an anionic oxygen of the phosphate at position 60 (Fig. 4.1-4G).

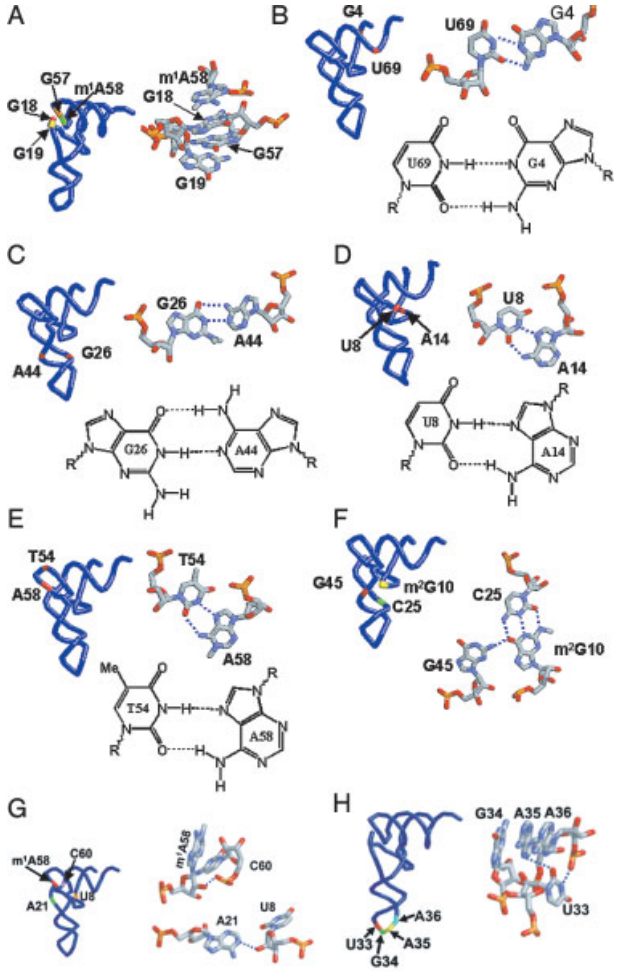

Figure 4.1-4 Unusual interactions involved in establishing the tertiary structure of tRNA. (A) "Base-zipper structure" characterized by the stacking of four purines coming from different secondary-structure motifs, viz. the D-loop (G18 and G19) and the T-loop (G57 and m¹A58); (B) cis Watson–Crick G4–U69 pair; (C) cis Watson–Crick G26–A44 inter-action (imino G-A pair); (D) trans Hoogsteen U8–A14 pair; (E) trans Hoogsteen T54-A58 pair; (F) base triplet interaction m²G10-C25-G45; (G) hydrogen-bonding formation between sugar phosphate backbone and a nitrogen base or an anionic oxygen of a phosphate. Two examples are shown, the U8(2′OH)-A21(N1) and C60(PO)-m¹A58(2′OH) hydrogen bonds. H, the U turn around U33 is essential for an optimal presentation of the bases of the anticodon loop, the 2′OH is important for an effective translocation of the tRNA from the P- to the E-Site (see text for details). All figures were prepared using 6tna pdb files processed by RasMol software.

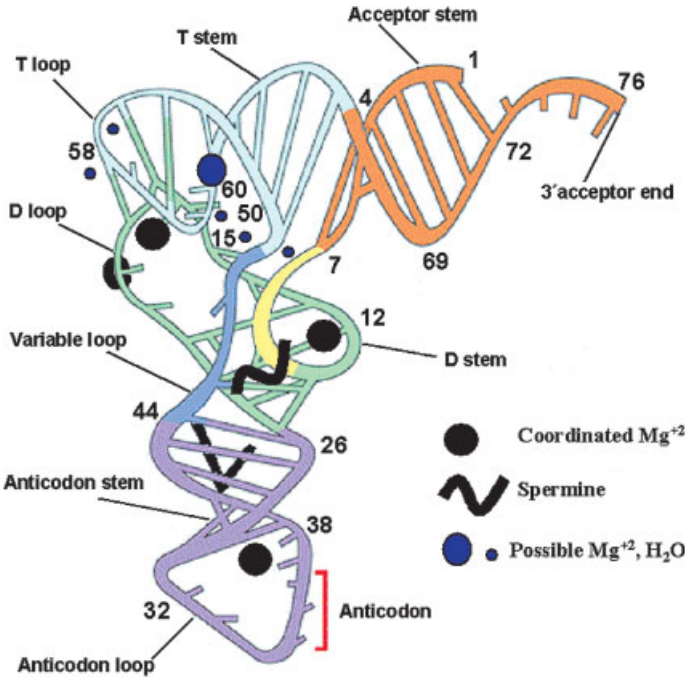

Figure 4.1-5 Coordinated Mg^{2+} ions and polyamines interact with the tRNA sugar-phosphate backbone where unusual folding of tRNA might be stabilized by these interactions. According to Ref. [55], modified.

(5) Finally, Mg^{2+} and polyamines interact with the phosphate backbone of RNA through electrostatic interactions to stabilize the folded backbone structure of the tRNA. The binding of the Mg^{2+} ion is coordinated by six molecules of water, some of which form hydrogen bonds with phosphate oxygens. The magnesium ions can also be directly coordinated by one or two phosphate oxygens and the rest of the sites can be occupied by water molecules which participate in hydrogen bonding with nitrogens or oxygens of the bases (Fig. 4.1-5).

As mentioned already, U33 in the anticodon loop is universally conserved, and this residue is of prominent importance for protein synthesis, including the reactions of tRNA binding to the P and A sites and translocation. U33 is instrumental for the "U-turn", a sharp 180° turn within a short stretch of three nucleotides in the anticodon loop that was in fact the first structural motif recognized in this loop (Figure 4.1-4H). The U-turn is characterized by two hydrogen bonds and a stacking interaction of U33 with O-P of residue 35 leading to a sharp turn in the phosphodiester bond following U33. The first hydrogen bond goes from U33 N3-H to O-P of residue 36 and the second from O2′ of U33 to N7 of A35 (or O4 of uracil or N4 of cytosine, if residue 35 is a U or C, respectively; [29]). The 2′-OH group of U33 is of outstanding

importance for P site binding and translocation, and less important for A site binding. Replacing the 2′-OH with a hydrogen or a 2′-O-methyl group reduces the P site binding by 50- and 100-fold, respectively [30, 31]. Similarly, a 2′-deoxy substitution impairs the translocation reaction by 50-fold, whereas the A site binding is reduced only by a factor of 7 [32]. U33 is also instrumental for the 3′-stack conformation of the anticodon loop, where the anticodon is stacked on the 3′-side of the loop. This seems to be the anticodon conformation being present under all binding conditions of a tRNA (see Ref. [34] for review).

4.1.4
tRNA Modifications

tRNAs are the most modified RNA molecules; almost 25% of the nucleotides of a tRNA are modified. Eighty nucleoside modifications out of a total of 95 reported modifications in all RNA molecules have been observed in tRNAs. Specific enzymes modify tRNAs during their maturation. Table 4.1-1 lists all types of known RNA modifications, and those that are found in tRNAs are indicated [34].

Usually the modification reaction is an alteration of, or addition to, existing bases in the tRNA, an exception being the base queuosine. This base is found 5′ to the anticodon at position 34 of tRNAs that read NAU or NAC codons (where N is any nucleotide), and the modification requires an enzyme that exchanges free queuosine with guanosine. Many examples of tRNA modifications include ribose/base methylations (Gm, Cm / m^5C), base isomerization (U to pseudouridine Ψ), base reduction (U to D; dihydro-uridine), base thiolation (s^2C, s^2U, s^4U) and base deamination (inosine). Some modifications are conserved features of all tRNA molecules (D residues that give rise to the name of the D-arm, Ψ found in the TΨC sequence).

Several functions have been attributed to tRNA modifications such as tertiary structure stabilization, increase in the specificity of the tRNA for its cognate tRNA synthetase, increase in the surface exposure of the tRNA, increase of interaction with initiation factors and elongation factors and involvement in the decoding of the RNA on the ribosome. The most direct effect of modification is seen in the anti-codon. Inosine, which is generated by deamination of adenine (I, instead of A), is often present at the first position of the anticodon (pairs with third position of the codon), where it is capable of pairing with any one of three bases: U, C, and A (wobble position). Curiously, although the inosine base is derived from adenine, its behavior is most similar to that of guanine in terms of potential base-pairing formation.

4.1.5
Recognition of tRNA by tRNA synthetase: Identity Elements

An important feature warranting the transfer of the genetic information is seen in the bridge function that a tRNA has, i.e., serving as linker between the RNA world and the protein world. Although codon:anticodon is the key interaction that directs the translational process, the correct recognition and aminoacylation of the tRNA by its cognate tRNA synthetase is of comparable importance. Mis-charging of tRNAs

can lead to the incorporation of the wrong amino acid into the polypeptide, thus impairing the fidelity of the translational process, which can lead to production of inactive or worse, toxic protein products.

Table 4.1-1 Abbreviations and chemical names of modified nucleosides found in RNA
Data taken from Ref. [34].

	Symbol	Name	Found in tRNA
1	Am	2′-*O*-methyladenosine	+
2	m^2A	2-Methyladenosine	+
3	m^6A	N^6-methyladenosine	+
4	$m^6{}_2A$	N^6,N^6-dimethyladenosine	
5	m^6Am	N^6,2′-O-dimethyladenosine	
6	$m^6{}_2Am$	N^6, N^6,2′-O-trimethyladenosine	
7	ms^2m^6A	2-Methylthio-N^6-methyladenosine	+
8	i^6A	N^6-isopentenyladenosine	+
9	ms^2i^6A	2-Methylthio-N^6-isopentenyladenosine	+
10	io^6A	N^6-(*cis*-hydroxyisopentenyl)-adenosine	+
11	ms^2io^6A	2-Methylthio-N^6-(*cis*-hydroxyisopentenyl)-adenosine	+
12	g^6A	N^6-glycinylcarbamoyladenosine	+
13	t^6A	N^6-threonylcarbamoyladenosine	+
14	m^6t^6A	N^6-methyl-N^6-threonylcarbamoyladenosine	+
15	ms^2t^6A	2-Methylthio-N^6-threonylcarbamoyladenosine	+
16	hn^6A	N^6-hydroxynorvalylcarbamoyladenosine	+
17	ms^2hn^6A	2-Methylthio-N^6-hydroxynorvaylcarbamoyladenosine	+
18	Ar(p)	2′-*O*-ribosyladenosine (phosphate)	+
19	m^1A	1-Methyladenosine	+
20	I	Inosine	+
21	Im	2′-*O*-methylinosine	
22	m^1I	2′-*O*-methylinosine	+
23	m^1Im	1,2′-*O*-dimethylinosine	+
24	Um	2′-*O*-methyluridine	+
25	s^2U	2-Thiouridine	+
26	s^2Um	2-Thio-2′-*O*-methyluridine	+
27	m^3U	3-Methyluridine	
28	m^3Um	3,2′-*O*-dimethyluridine	
29	acp^3U	3-(3-Amino-3-carboxypropyl)uridine	+
30	s^4U	4-Thiouridine	+
31	m^5U	Ribosylthymine	+
32	m^5Um	5,2′-*O*-dimethyluridine	+
33	m^5s^2U	5-Methyl-2-thiouridine	+

	Symbol	Name	Found in tRNA
34	ho^5U	5-Hydroxyuridine	+
35	mo^5U	5-Methoxyuridine	+
36	cmo^5U	Uridine 5-oxyacetic acid	+
37	$mcmo^5U$	Uridine 5-oxyacetic acid methyl ester	+
38	cm^5U	5-Caboxymethyluridine	
39	mcm^5U	5-Methoxycarbonylmethyluridine	+
40	mcm^5Um	5-Methoxycarbonylmethyl-2′-*O*-methyluridine	+
41	mcm^5s^2U	5-Methoxycarbonylmethyl-2-thiouridine	+
42	ncm^5U	5-Carbamoylmethyluridine	+
43	ncm^5Um	5-Carbamoylmethyl-2′-*O*-methyluridine	+
44	chm^5U	5-(Carboxyhydroxymethyl)uridine	+
45	$mchm^5U$	5-(Carboxyhydroxymethyl)uridinemethyl ester	+
46	nm^5s^2U	5-Aminomethyl-2-thiouridine	+
47	mnm^5U	5-Methylaminomethyluridine	+
48	mnm^5s^2U	5-Methylaminomethyl-2-thiouridine	+
49	mnm^5se^2U	5-Methylaminomethyl-2-selenouridine	+
50	$cmnm^5U$	5-Carboxymethylaminomethyluridine	+
51	$cmnm^5Um$	5-Carboxymethylaminomethyl-2′-*O*-methyluridine	+
52	$cmnm^5s^2U$	5-Carboxymethylaminomethyl-2-thiouridine	+
53	D	Dihydrouridine	+
54	m^5D	Dihydroribosylthymine	
55	Ψ	Pseudouridine	+
56	Ψm	2′-*O*-methylpseudouridine	+
57	$m^1\Psi$	1-Methylpseudouridine	+
58	$m^3\Psi$	3-Methylpseudouridine	
59	m^1acp Ψ	1-Methyl-3-(3-amino-3-carboxypropyl)pseudouridine	
60	Gm	2′-*O*-methylguanosine	+
61	m^1G	1-Methylguanosine	+
62	m^2G	N^2-methylguanosine	+
63	$m^2{}_2G$	N^2,N^2-dimethylguanosine	+
64	m^2Gm	N^2,2′-*O*-dimethylguanosine	+
65	$m^2{}_2Gm$	N^2,N^2,2′-*O*-trimethylguanosine	+
66	Gr(p)	2′-*O*-ribosylguanosine (phosphate)	+
67	m^7G	7-Methylguanosine	+
68	$m^{2,7}G$	N^2,7-dimethylguanosine	
69	$m^{2,2,7}G$	N^2,N^2,7-trimethylguanosine	
70	imG	Wyosine	+
71	mimG	Methylwyosine	+
72	OHyW*	Undermodified hydroxywybutosine	+

	Symbol	Name	Found in tRNA
73	yW	Wybutosine	+
74	OHyW	Hydroxywybutosine	+
75	o_2yW	Peroxywybutosine	+
76	Q	Queuosine	+
77	oQ	Epoxyqueuosine	+
78	GalQ	Galactosyl-queuosine	+
79	manQ	Mannosyl-queuosine	+
80	$PreQ_0$	7-Cyano-7-deazaguanosine	+
81	gQ(G+)	Archaeosine (alternate name 7-formamidino-7-deazaguanosine)	+
82	$PreQ_1$	7-Aminomethyl-7-deazaguanosine	+
83	Cm	2′-*O*-methylcytidine	+
84	m^4C	*N4*-methylcytidine	
85	m^4Cm	*N4*,2′-*O*-dimethylcitidine	
86	ac^4C	*N4*-acetylcytidine	+
87	ac^4Cm	*N4*-acetyl-2′-*O*-methylcytidine	+
88	m^5C	5-Methylcytidine	+
89	m^5Cm	5,2′-*O*-dimethylcytidine	+
90	hm^5C	5-Hydroxymethylcytidine	
91	f^5C	5-Formylcytidine	+
92	f^5Cm	2′-O-methyl-5-formylcytidine	+
93	m^3C	3-Methylcytidine	+
94	s^2C	2-Thiocytidine	+
95	k^2C	Lysidine	+

There is a distinct synthetase that recognizes every tRNA that participates in the decoding of the same amino acid, and this group of tRNAs are termed “isoaccepting” tRNAs. For such a situation to exist, the isoaccepting tRNAs must carry identical signals for the recognition of their synthetase. These common signals define the recognition identity of the isoacceptor tRNAs and accordingly represent the “identity elements”, which have been discovered in the past 15 years by many groups (see Table 4.1-2). An identity element or positive element is defined as a recognition site on the tRNA that allows the unique aminoacylation by its cognate aaRSs. Table 4.1-2 gives a survey of the currently known identity elements described in Ref. [35].

An interesting case is the so-called “negative determinants”: these modifications do not improve the recognition by the cognate synthetase (i.e., the corresponding, correct synthetase), but prevent or impair recognition by a non-cognate synthetase and thus mis-charging by a non-cognate synthetase [35].

Table 4.1-2 Identity elements in tRNAs aminoacylated by class I (A) and class II (B) synthetases
Data taken from Ref. [35]

	E. coli[a]	***S. cerevisiae***[b]	***T. thermophilus***[c]	**Others**[d]
(A) *Aminoacylated by class I synthetases*				
Val	(a) **A73** **G3:C70**, **U4:A69** (b) A35, C36	***A73*** – ***A35***		
Ile	(a) **A73** **C4:G69** (b)**L**/G34, A35, U36 **t**^{6}**A37**, **A38** (c) **U12:A23**, **C29:G41**	**I**34,**A35**,**U36**		
Leu	(a) **A73** (b) – (c) **U8•A14**	***A73*** ***A35*** ***G37***	– –	***A73*** **C3:G70**, **A4:U69** **G5:C68** ***C20a***
Met (fMet)	(a) **A73** (**G2:C71**,**C3:G70**) *U4:A69,A5:U68* (b) **C34**,**A35**,**U36** (**C32**,**U33**,**A37**) (c)	A73 **C34**,**A35**,**U36** **& the 4 other AC loop nts** **D-arm**		
Cys	(a) U73 *G2:C71,C3:G70* (b) G34,*C35,A36* (c) **G15•G48**,**A13•A22**	***U73***		
Tyr	(a) A73 (b) U35	**A73** *C1:G72* *G34,Ψ35*		**C1:G72**
Trp	(a) G73 A1:U72,*G2:C71* *G3:C70* (b) C34,C35,A36	 C24,C35		G73 A1:U72 G5:C68,A9 C34,C35,C36

	E. coli[a]	*S. cerevisiae*[b]	*T. thermophilus*[c]	Others[d]
Glu	(a) **G1:C72,U2:A71** (b) **s**[4]**U34**,U35 ***A37*** (c) **U11:A24** **U13:G22-A46**, Δ**47**			
Gln	(a) **G73** **U1:A72**, G2:C71 G3:C70 (b) **Y34,U35,G36** **A37,U38** (c) **G10**			
Arg	(a) A/G73 (b) **C35,U/G36**	**C35,U/G36**		
	(c) **A20**			
(B) *Aminoacylated by class II synthetases*				
Ser	(a) *G73* *C72,*G2:C71, A3:U70 *C11:G24,***R4:Y69**			G73
	(c) *C11:G24* Variable loop	Variable loop		Variable loop
Thr	(a) **G1:C72,C2:G71**	**G1:C72**	U73 **G1:C72,U3:A70**	
	(b) **G34,G35,U36**	G35,U36	G35,U36	
Pro	(a) **A73** G72 (b) **G35,G36**			**G35,G36**
	(c) **G15•C48**			
Gly	(a) U73 G1:C72,C2:G71 G3:C70	**A73** **C2:G71, G3:C7O**	**U73** **G1:C72, C2:G71** (**G3:C70**)	**A73** **C2:G71**
	(b) C35,**C36** (c)	**C35,C36**	C35,C36 (**G10:C25**)	
His	(a) C73 **G-1**	**A73** **G-1**		
	(b) anticodon	**G34,U35**		

	E. coli[a]	*S. cerevisiae*[b]	*T. thermophilus*[c]	Others[d]
Asp	(a) **G73** **G2:C71** (b) **G34,U35,C36** **C38** (c) **G10**	**G73** **G34,U35,C36** **C38** **G10•U25**	**G73** **G34,U35,C36** **C38** **G10**	
Lys	(a) A73 (b) **U34**,U35,**U36** (mnm^{5}s^{2}U)34			
Asn	(a) *G73* (b) *G34,U35,U36*			
Phe	(a) *A73* (b) *G34,A35,A36* G27:C43,G28:C42 (c) *U20* *G44,U45,U59,U60*	**A73** **G34,A35,A36** i^{6}A37 **G20**	**A73** **G34,A35,A36**	**A73** **G34,A35,A36** **G30:C40** **A31:U39,G20**
Ala	(a) **A73** G2:C71,G3•U70 G4:C69 (c) **G20**	**G3•U70**		**G3•U70**

The tRNAs are listed according to the synthetase classification in two classes with subclasses. Identity elements are classified according to their location in the amino acid accepting stem (a), anticodon region (b), and other tRNA domains (c). Identity nucleotides in bold were identified by the *in vitro* approach, those in italics by the *in vivo* approach, and those in normal scripts by both approaches; when underlined, the identity element is the modified nucleotide. Numbering of residues is according to Ref. [16] and nomenclature of modified nucleotides according to Ref. [56]. In the case of base pair, (:) denotes WC pair, (:) non-WC pairs, and (-) tertiary pairs; (/) indicates that two residues can be identity elements at the same position. R, purine; Y, pyrimidine.

4.1.6
Is the tRNA Cloverleaf Structure a Pre-requisite for the L-shape?

The answer to this question is "no", since there are a number of structures that mimic the tertiary structure of an L-shaped tRNA but do not contain the canonical secondary cloverleaf structure. These variant tRNA structures can be recognized and aminoacylated by the cognates aaRSs.

Examples are tRNA-like molecules such as bacterial tmRNA (= 10Sa) that contain about 350 nt (cf. 75 nt for a tRNA –>4 × larger!), but 5′- and 3′-regions form a tRNA-like structure without T stem loop and the anticodon loop. The tmRNA contains an mRNA "module" that codes for about 9–30 amino acids; it is charged by alanine-tRNA synthetase (AlaRS) [36]. This RNA displays both tRNA and mRNA functions

(hence the name tmRNA) and plays an important role in recycling 70S ribosomes that are stuck at the 3′-end of fragmented mRNAs lacking a stop codon. The mRNA part of the tmRNA encodes an oligopeptide sequence that is tagged onto the incomplete polypeptide, targeting it for rapid degradation (see Sect. 8.2.5)

Another example is the tRNA-like structure in 5′-untranslated region of *thrS* mRNA (regulatory domain of threonyl-tRNA synthetase gene), which is recognized by ThrRS [37]. The tRNA-like structure at the 3′-end of the RNA from the turnip yellow mosaic virus (TYMV) can be charged by ValRS [38] and seems to play a role during replication of the virus [39]

Here we should also mention the "crippled" tRNAs that are found in the mitochondria of nematodes [40, 11], where the T-loop has been reduced to few base pair or deleted completely, whereas the $tRNA^{Ser}$ in this organelle is lacking the D stem loop instead of the T-loop, and also possesses its own EF-Tu factor. Human $tRNA^{Ser}$ with the anticodon GCU is another example where T-stem is not present at all, similar to nematode $tRNA^{Ser}$ [41].

4.1.7
Other Functions of tRNA outside the Ribosomal Elongation Cycle

In addition to the main role of the tRNAs in protein synthesis during the ribosomal elongation cycle, tRNAs are also involved in a series of other reactions beyond protein synthesis.

1. Viral reverse transcriptase of the human immunodeficiency virus (HIV) uses $tRNA_3^{Lys}$ as a primer for the synthesis of DNA [42].
2. Some tRNAs induce the formation of anti-termination structures of the nontranslated region upstream (UTR) of the structural genes of some amino acid operons (*ilv-leu*, *his, trp*) and of some tRNA synthetase genes (*thrS*, *tyrS, lueR, pheS*). Under starvation conditions, deacylated tRNA seems to base-pair via its anticodon and the NCCA-3′-end with complementary sequences in the leader UTR promoting the formation of anti-termination structures in these systems (reviewed in Ref. [43]).
3. Under nutrient deprivation conditions, bacterial cells down-regulate the transcription of genes that belong to the fields of molecular genetics such as replication, transcription, and translation. This most important regulation circuit in bacteria is called the "stringent response" and is mediated by the synthesis of (p)ppGpp. Binding of a deacylated tRNA to the ribosomal A-site activates via the ribosomal protein L11 the ribosome-bound enzyme, RelA, which synthesizes the signaling molecule (p)ppGpp [44] (see Chapter 11.2.3).
4. Glu-$tRNA^{Glu}$ is an activated intermediate in the biosynthetic pathway of δ-aminolevulinate (ALA), a tetrapyrrole precursor of porphyrins in plants and bacteria. The ALA biosynthesis starts with the aminoacylation of $tRNA^{Glu}$ by GluRS, then a NADPH-dependent reduction reaction catalyzed by glutamyl-tRNA reductase occurs on Glu-$tRNA^{Glu}$ to yield glutamate 1-semialdehyde. Finally, the amino group of glutamate 1-semialdehyde is transferred to its terminal carbon by an intramolecular reaction catalyzed by a specific aminotransferase forming ALA [45].

5. Amino acid residues from aminoacyl-tRNAs are used in a cross-linking reaction that occurs during peptidoglycan synthesis of the bacterial cell wall. By this reaction, the pentapeptide moieties attached to the *N*-acetyl muramic acid residue of both the disaccharides, *N*-acetyl muramic acid-*N*-acetyl glucosamine units, become covalently bound [46, 47].

6. RNA polymerase III activity in silkworm depends on several transcription factors. Among these transcription factors, TFIIIR stands out since it contains a $tRNA^{Ile}$ with the anticodon IAU, where I stands for inosine [48].

7. Aminoacyl-tRNAs are involved in a proteolytic pathway, the so-called the "N-end rule" pathway. The N-end pathway governs the half-life of a protein in a cell with respect to the identity of its N-terminal amino acid residue. For instance, arginine-tRNA-protein transferase (R-transferase) is an enzyme that uses Arg-$tRNA^{Arg}$ to "arginylate" polypeptides whose N-terminal residue is Asp or Glu in bacteria and Cys in mammals. This arginylation is the signal for the proteolytic machinery for the protein degradation [49].

8. Many tRNA-like structures are specifically aminoacylated and participate actively in protein synthesis. An example of such structures is found at the 3′-end of the genome of several plant viral RNAs. In the case of 3′-untranslated region of turnip yellow mosaic virus (TYMV), a tRNA-like structure aminoacylated with valine is involved in virus replication and indispensable for virus viability [34]. More recently, a tRNA-like molecule (sRNA85) was identified in the trypanosomatid signal recognition particle in addition to the canonical 7SL RNA homolog. The complex has an *S*-value of ~14S and binds to the ribosomes [50].

4.1.8
Human Neurodegenerative Disorders Associated with Mitochondrial tRNAs

In mammals, many diseases are known that are caused by tRNA defects in the mitochondria. Often they are related to human neurodegenerative disorders. Table 4.1-3 summarizes mutations in human mitochondrial tRNA genes associated with the corresponding disease or phenotype.

Table 4.1-3 Disease-related mutations in human mitochondrial tRNA genes
Data taken from Ref. [51]

	tRNA mutation			
Amino acid specificity	**Gene mutation**	**Domain**	**Position**	**Related pathologies**
Ala	A5628G	AC stem	31–39	CPEO
Asn	A5692G	AC loop	38	CPEO
	C5698T	AC loop	32	PEO
	C5703T	AC stem	27–43	CPEO, MM
Asp	A7543G	AC stem	29–41	MS

tRNA mutation				
Amino acid specificity	**Gene mutation**	**Domain**	**Position**	**Related pathologies**
Cys	A5814G	D stem	13–22	EM, MELAS, PEO
Gln	C4332T	acc. stem	3–70	EM, D
	A4336G	acc. stem	7–66	ADPD
	instT4370	AC loop	After 31	MM, CD
Glu	A14709G	AC loop	37	MM, EM, D
Gly	T9997C	acc. stem	7–66	MHCM
	A10006G	D loop	18	CIPO
	T10010C	D stem	12–23	EM
	A10044G	T loop	59	EM
Ile	A4269G	acc. stem	7–66	FICP, EM
	T4274C	D stem	13–22	CPEO
	T4285C	AC stem	27–43	PEO
	G4298A	AC stem	30–40	CPEO
	G4309A	T stem	51–63	CPEO
	A4317G	T loop	59	FICP
	C4320T	T stem	52–62	ECM
Leu (CUN)	T12297C	AC loop	33	DCM
	G12301A	AC loop	37	AISA
	G12415A	T stem	52–62	CPEO
	A12320G	T loop	57	MM
Leu (UUR)	A3243G	D loop	14	MELAS, DMDF
	A3243T	D loop	14	PEM, MM
	G3249A	D loop	19	KS
	T3250C	D loop	20	MM
	A3251G	D loop	20:01	MM
	A3252G	D loop	21	MELAS
	C3254G	D stem	12–23	MM
	C3256T	D stem	10–25	MERRF-like, MELAS
	T3258C	AC stem	27–43	LA, E1
	A3260G	AC stem	29–41	MMC
	T3264C	AC loop	33	DM
	T3271C	AC stem	30–40	MELAS, DM
	delT3272	AC stem	29–41	PEM
	T3273C	AC stem	28–42	O: EI
	C3275A	Var. region	44	LHON
	A3280G	T stem	49–65	MM
	A3288G	T loop	57	MM

tRNA mutation				
Amino acid specificity	**Gene mutation**	**Domain**	**Position**	**Related pathologies**
	T3291C	T loop	60	MELAS
	A3302G	acc. stem	2–71	MM
	C3303T	acc. stem	1–72	MMC
Lys	A8296G	acc. stem	2–71	DMDF, MERRF
	G8313A	D stem	12–24	MNGIE
	T8316C	AC stem	27–43	MELAS
	G8328A	AC stem	31–39	EM
	G8342A	T stem	53–61	PEO, MS
	A8344G	T loop	55	MERRF
	T8355C	T stem	50–64	PEO, SM
	T8356C	T stem	49–65	MERRF
	T8362G	acc. stem	2–71	SM
	G8363A	acc. stem	1–72	MICM, D, MERRF, LS
Met	T4409C	acc./D stem	8	MM
	G4450A	T stem	53–61	MM
Phe	G583A	acc. Stem	7–66	MELAS
	A606G	AC stem	29–41	M
	T618C	AC stem	29–41	MM
Pro	T15965C	T stem	50–64	ADPD
	G15990A	AC loop	36	MM, O
Ser (AGY)	C12246A	T loop	55	CIPO
	C12258A	acc. stem	7–66	DMDF
Ser (UCN)	insG7472	Var. region	46	PEM
	C7497T	D stem	13–22	MM, PEM, RRF, LA
	A7511G	acc. stem	4–69	DEAF, SNHL
	A7512G	acc. stem	3–70	PEM
Thr	G15915A	AC stem	30–40	MM
	A15923G	AC loop	38	LIMM
	delT15940	T loop	60	MM
	G15950A	acc. stem	3–70	ADPD
Trp	G5521A	D stem	10–25	MM
	insT5537	AC stem	After 27	MILS
	G5540A	AC stem	30–40	PEM, CD
	G5549A	AC stem	31–39	DEMCHO, D, A
Tyr	A5874G	D stem	13–22	EI, LW, CD
Val	G1606A	acc. stem	5–68	AMDF

tRNA mutation

Amino acid specificity	Gene mutation	Domain	Position	Related pathologies
	G1642A	AC stem	27–43	MELAS
	G1644T	Var. region	45	LS

tRNA genes are listed by amino acid specificity in alphabetical order. "Gene mutation" refers to the nucleotide substitution and position of the mutation in human mt genome. "tRNA mutation" refers to the location of the mutation in the gene product. Structural domains affected by the mutations refer to loops and stems, with AC for anticodon; acc., acceptor; Var., variable. Nucleotide numbering is according to classical tRNA numbering [16]. Pathologies are abbreviated as follows:
A, ataxia;
ADPD, Alzeimer's disease and Parkinsons disease;
AISA, acquired idiopathic sideroblastic anemia;
AMDF, ataxia, mental deterioration, deafness;
CD, Cox deficiency;
CIPO, chronic instestinal pseudoobstruction with myopathy;
CPEO, chronic progressive external ophtalmoplegia;
D, diabetes;
DCM, dilated cardiomyopathy;
DEAF, maternally inherited deafness or aminoglycoside-induced deafness;
DEMCHO, DEMentia;
Chorea;
DM, diabetes mellitus;
DMDF, diabetes mellitus, Deafness;
ECM, encephalocardiomyopathy;
EI, exercise intolerance;
EM, encephalomyopathy;
FICP, fatal infantile cardiomyopathy plus a melas-associated cardiomyopathy;
HCM, hypertrophic cardiomyopathy;
KS, Kearns–Sayre syndrome;
LA, lactic acidose,
LHON, leber hereditary
optic neuropathy;
LIMM, lethal infantile mitochondrial myopathy;
LS, leigh syndrome;
LW, limb weakness;
M, myoglobinuria;
MELAS, mitochondrial encephalomyopathy, lactis acidose, Stroke-like episodes;
MERRF, myoclonic epilepsy and ragged red muscle fibers;
MHCM, maternally inherited hypertrophic cardiomyopathy;
MICM, maternal, inherited cardiomyopathy;
MILS, maternal inherited leigh syndrome;
MM, mitochondrial myopathy;
MMC, maternal myopathy and cardiomyopathy;
MNGIE, mitochondrial neurogastrointestinal encephalomyopathy;
MS, myoclonic seizures;
O, ophtalmoplegia;
PEM, progressive encephalomyopathy;
PEO, progressive external ophtalmoplegia;
SM, skeletal myopathy;
SNHL, sensorineural hearing loss.

References

1 F. H. C. Crick, in *The Structure of Nucleic Acids and their Role in Protein Synthesis,* ed. E. M. Crook, Cambridge University Press London 1957.
2 M. B. Hoagland, P. C. Zamecnik, M. L. Stephenson, *BBA* **1957**, *24,* 215.
3 G. Zubay, *J. Mol. Biol.* **1962**, *4.*
4 M. Nirenberg, P. Leder, *Science* **1964**, *145,* 1399–1407.
5 R. Holley, J. Apgar, G. Everett et al., *Science* **1965**, *147,* 1462–1465.
6 S. H. Kim, F. L. Suddath, G. J. Quigley et al., *Science* **1974**, *185,* 435–440.
7 J. D. Robertus, J. E. Ladner, J. T. Finch et al., *Nature* **1974**, *250,* 546–551.
8 H. G. Khorana, *Science* **1979**, *203,* 614–625.
9 U. L. RajBhandary, D. Söll, In: *tRNA: Structure, Biosynthesis, and Functions,* eds D. Söll and U. L. RajBhandary, ASM Washington, DC 1995, 1–4.
10 M. Eigen, B. F. Lindemann, M. Tietze et al., *Science* **1989**, *244,* 673–678.
11 Y. I. Watanabe, H. Tsurui, T. Ueda et al., *J. Biol. Chem.* **1994**, *269,* 22902–22906.
12 H. G. Zachau, H. Dutting, H. Feldmann et al., *Cold Spring Harbor Symp. Quant. Biol.* **1966**, *31,* 417–424.
13 J. T. Madison, G. A. Everett, H. K. Kung, *Cold Spring Harbor Symp. Quant. Biol.* **1966**, *31,* 409–416.
14 U. L. RajBhandary, A. Stuart, R. D. Faulkner et al., *Cold Spring Harbor Symp. Quant. Biol.* **1966**, *31,* 425–434.
15 P. Auffinger, E. Westhof In: *Modification and Editing of RNA,* eds H. Grosjean and R. Benne, ASM, Washington, DC 1998.
16 M. Sprinzl, C. Horn, M. Brown et al., *Nucleic Acids Res.* **1998**, *26,* 148–153.
17 P. Schimmel, L. R. de Pouplana, *Cell* **1995**, *81,* 983–986.
18 M. Taiji, S. Yokoyama, T. Miyazawa, *Biochemistry* **1985**, *24,* 5776–5780.
19 E. G. Wagner, P. C. Jelenc, M. Ehrenberg et al., *Eur. J. Biochem.* **1982**, *122,* 193–197.
20 T. Wagner, M. Sprinzl, *Biochemistry* **1983**, *22,* 94–98.
21 M. A. Augustin, A. S. Reichert, H. Betat et al., *J. Mol. Biol.* **2003**, *328,* 985–994.
22 D. Moras, M. B. Comarmond, J. Fischer et al., *Nature* **1980**, *288,* 669–674.
23 R. K. Agrawal, C. M. T. Spahn, P. Penczek et al., *J. Cell. Biol.* **2000**, *150,* 447–459.
24 K. H. Nierhaus, J. Wadzack, N. Burkhardt et al., *Proc. Natl. Acad. Sci. USA* **1998**, *95,* 945–950.
25 M. M. Yusupov, G. Z. Yusupova, A. Baucom et al., *Science* **2001**, *292,* 883–896.
26 G. Z. Yusupova, M. M. Yusupov, J. H. Cate et al., *Cell* **2001**, *106,* 233–241.
27 T. Hermann, D. J. Patel, *J. Mol. Biol.* **1999**, *294,* 829–849.
28 Z. Sharbarova, A. Bogdanov, In: *Advance Organic Chemistry of Nucleic Acids,* ed. T. Kellersohn VCH Publishers, Weinheim New York, NY 1994, 335–388.
29 G. J. Quigley, A. Rich, *Science* **1976**, *194,* 796–806.
30 S. S. Ashraf, G. Ansari, R. Guenther et al., *RNA* **1999**, *5,* 503–511.
31 U. von Ahsen, R. Green, R. Schroeder et al., *RNA* **1997**, *3,* 49–56.
32 S. S. Phelps, O. Jerinic, S. Joseph, *Mol. Cell* **2002**, *10,* 799–807.
33 S. M. Stagg, M. Valle, R. K. Agrawal, J. Frank et al., *RNA* **2002**, *8,* 1093–1094.
34 Y. Motorin, H. Grosjean, *tRNA Modification* ed. N. P. Group, Nature Publishing Group, London 1999.
35 R. Giegé, M. Sissler, C. Florentz, *Nucleic Acids Res.* **1998**, *26,* 5017–5035.
36 Y. Hou, P. Schimmel, *Nature* **1988**, *333,* 140–145.
37 M. Graffe, J. Dondon, J. Caillet et al., *Science* **1992**, *255,* 994–996.
38 M. Pinck, P. Yot, F. Chapeville et al., *Nature* **1970**, *226,* 954–956.

39 R. Giegé, C. Florentz, T. W. Dreher, *Biochemie* **1993,** *75,* 569–582.
40 T. Ohtsuki, G. Kawai, K. Watanabe *FEBS Lett.* **2002,** *514,* 37–43.
41 G. Dirheimer, G. Keith, P. Dumas et al., In: *tRNA: Structure, Biosynthesis and Function* eds D. Söll and U. L. RajBhandary, ASM, Washington, DC 1995.
42 A. T. Das, B. Klaver, B. Berkhout *J Virol* **1995,** *69,* 3090–3097.
43 P. Gollnick, P. Babitzke, *BBA* **2002,** *1577,* 240–250.
44 T. M. Wendrich, G. Blaha, D. N. Wilson et al., *Mol. Cell* **2002,** *10,* 779–788.
45 C. G. Kannangara, S. P. Gough, P. Bruyant et al., *Trends Biochem. Sci.* **1988,** *13,* 139–143.
46 R. Plapp, J. L. Strominger, *J. Biol. Chem.* **1970,** *245,* 3667–3674.
47 S. S. Hegde, T. E. Shrader, *J. Biol. Chem.* **2001,** *276,* 6998–7003.
48 H. M. Dunstan, L. S. Young, K. U. Sprague, *Mol. Cell. Biol.* **1994,** *14,* 3588–3595.
49 A. Varshavsky, *Proc. Natl Acad. Sci. USA* **1996,** *93,* 12142–12149.
50 L. Liu, H. Ben-Shlomo, Y. Xu et al., *J. Biol. Chem.* **2003,** *278,* 18271–18280.
51 C. Florentz, *Biosci. Rep.* **2002,** *22,* 81–98.
52 E. Goldman, *Transfer RNA,* ed. N. P. Group, Nature Publisher Group London 1999.
53 E. Westhof, P. Auffinger, *Transfer RNA Structure,* ed. N. P. Group, Nature Publishing Group London 1999.
54 S. H. Kim, *Prog. Nucleic Acid Res. Mol. Biol.* **1976,** *17,* 181–216.
55 G. J. Quigley, M. M. Teeter, A. Rich, *Proc. Natl. Acad. Sci. USA,* **1978,** *75,* 64–68.
56 P. A. Limbach, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1994,** *22,* 2183–2196.

4.2
Aminoacylations of tRNAs: Record-keepers for the Genetic Code

Lluís Ribas de Pouplana and Paul Schimmel

4.2.1
Introduction

Aminoacyl-tRNA synthetases (ARS) catalyze aminoacylation reactions and therefore are essential components of the genetic code [1, 2]. These enzymes aminoacylate each transfer RNA with its cognate amino acid, thus establishing the amino acid–trinucle-otide relationships of the code. Each synthetase recognizes its specific amino acid and all its isoacceptor tRNAs. The reaction takes place in two steps:

$$E + AA + ATP \rightarrow E(AA\text{-}AMP) + PPi$$
$$E(AA\text{-}AMP) + tRNA \rightarrow E + AA\text{-}tRNA + AMP$$

First, the enzyme (E) activates the cognate amino acid by condensing it with ATP to form a transient aminoacyl adenylate (AA-AMP) that remains bound to the enzyme's active site. Secondly, the enzyme catalyzes the formation of an ester linkage between the carboxyl group of the amino acid and a hydroxyl of the ribose of the terminal 3′ adenosine of the tRNA. The aminoacylated tRNAs (AA-tRNAs) are then recognized by translation factors that place them in the ribosome's active site, where protein synthesis takes place.

Every cell requires a synthetase for each of the 20 amino acids of the genetic code. Thus, all cells contain at least 20 synthetases (eukaryotic cellular organelles use an additional set of synthetases) [3]. With minor exceptions, all aminoacyl-tRNA synthetases with the same amino acid specificity are orthologs. For example, all extant aspartyl-tRNA synthetases (the enzymes responsible for aminoacylating $tRNA^{Asp}$ with aspartate) are related to a single ancestor, which has been conserved throughout all speciation events since the last universal common ancestor (LUCA) of all organisms [4].

The concept of LUCA refers to the biological entity that constituted the genetic basis of all extant forms of life. That such a common ancestor existed follows from the universal distribution and composition of the genetic code and its components. Any gene that is found in all extant living species plausibly has an origin that precedes LUCA (this assumption would be false in the case of later-appearing genes transferred to all living species through lateral gene transfer). Most individual ARS are universally distributed. In many cases, the phylogenetic tree derived from their sequences coincides with the evolutionary tree derived from 16S RNA sequences [5, 6]. Hence, most individual synthetases probably predate the separation of the three kingdoms of life – archaea, bacteria, and eukarya.

In addition to the individual evolutionary history of each synthetase, the 20 known ARSs are divided into two classes of homologs, each containing 10 enzymes [7–9]. Each class is identified by a common active site fold, and by certain sequence motifs,

shared by all its members [2, 10]. All enzymes of each class evolved from a common ancestor, which gave rise to the extant 10 types through gene duplication events. Since most ARS may be older than LUCA, most of these duplications took place before LUCA.

Interestingly, a few synthetases had not completely evolved at the time of LUCA [11]. These enzymes are the outcome of branches of the synthetase evolutionary tree that were only fixed in evolution after the separation of the three main branches of life. A particularly interesting subset of this late evolution is that related to the endosymbiotic events that generated mitochondria and chloroplasts. The cohabitation of organelle genomes within eukaryotic cells resulted in selection of new recognition mechanisms between tRNAs and ARS [11]. Among the forces behind these selections is the requirement to preserve faithful recognition of two independent sets of tRNAs. Simultaneously, the recognition mechanisms between tRNAs and ARS might have been influenced by the significant reduction in genome size observed in animal mitochondria. The analysis of these exceptional ARSs provides information about the origin of extant cells after the separation of the three branches of life, as well as about the events that determined the selection of their extant phenotypes [12, 13].

Thus, the origin of aminoacyl-tRNA synthetases is ancient, predating the appearance of the last common ancestor to all living species. The intimate functional link between synthetases and the genetic code suggests a common evolutionary pathway. The evolutionary history of these enzymes can be inferred from structural, sequence, and phylogenetic comparisons. In turn, every aspect of ancient ARS evolution that is solved invariably provides information about general aspects of the origin of life. In this regard, aminoacyl-tRNA synthetases can be used as markers of essential transitions in evolution. Current understanding of the evolution of aminoacyl-tRNA synthetases and the relationship to the development of the code is summarized here.

4.2.2 The Operational RNA Code

The aminoacyl-tRNA synthetases are the actual translators of the genetic code. Their faithful recognition of cognate tRNAs ensures the correct coupling of triplet sequences and amino acids. The recognition of the tRNA molecules by these enzymes depends on the specific interactions between the proteins and identity elements present in the tRNA sequences and structures (see Refs. [23, 25] for a review of identity elements of tRNAs). In some instances, however, the identity elements of tRNAs recognized by ARS do not include the anticodon bases of the tRNA, and are located in the tRNA acceptor stem (Fig. 4.2-1).

For instance, a major identity element of $tRNA^{Ala}$ is a single G:U base pair at the 3:70 position in the acceptor stem of that tRNA [14, 15]. This single G3:U70 base pair is necessary and sufficient to convey alanine acceptance upon many tRNA sequences. Alanine-tRNA synthetase (AlaRS) does not recognize the anticodon region of its cognate tRNA [16].

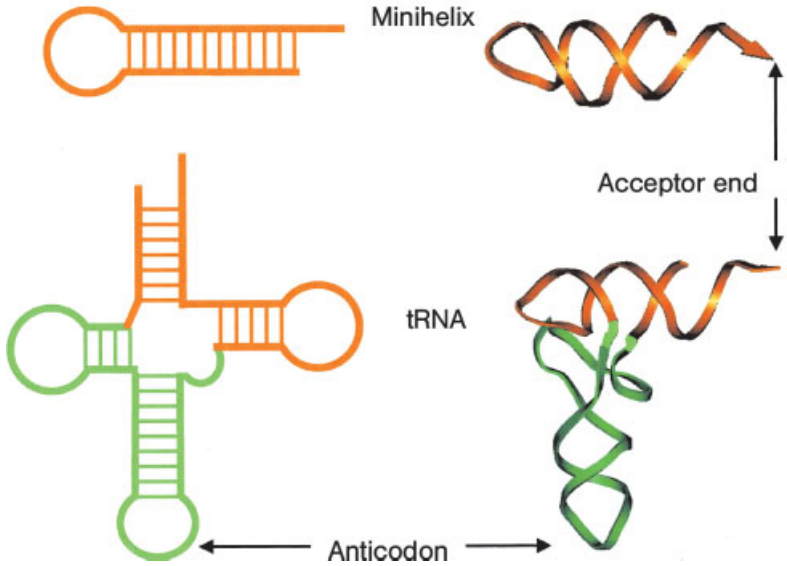

Figure 4.2-1 Secondary and tertiary structure models of an RNA minihelix and a tRNA. The minihelix dumbbell and the corresponding region of the tRNA structure are colored in orange. The rest of the tRNA structure is colored green.

A second example of an ARS that does not recognize the anticodon region is seryl-tRNA synthetase (SerRS). The crystal structure of *Thermus thermophilus* SerRS, complexed with $tRNA^{Ser}$, revealed that this enzyme does not interact with the anticodon triplet [8, 17]. The main identity element in the acceptor stem of $tRNA^{Ser}$ is the discriminator base G73 [18, 19]. The elbow region of the tRNA is also recognized by the enzyme through interactions with an idiosyncratic coiled-coil domain at the N-terminus of SerRS [17].

In these cases, the relationship between an anticodon triplet and an amino acid is indirect, because tRNA recognition is achieved via identity elements embedded in the acceptor stems and not in the anticodons. This set of interactions is known as the operational RNA code [20]. It relates specific RNA sequences/structures in acceptor stems to specific amino acids [20–22]. Indeed, small RNA helices that recapitulate acceptor stems are charged with specific amino acids.

At least 10 synthetases have been shown to aminoacylate specifically RNA minihelices that are based on the acceptor stems of their cognate tRNAs (Fig. 4.2-1) [19, 22–32]. Although the efficiency of aminoacylation of these minimalist structures can be significantly decreased with respect to the full-length tRNA, these minihelices are specifically recognized and aminoacylated by their cognate synthetases. Thus, the set of interactions that constitute the operational RNA code today are possibly the molecular remnants of the set of identity elements that ruled the recognition of the molecular ancestors of tRNA (minihelices).

The operational RNA code might be a testimony of earlier times in genetic code evolution, when RNA stem-loop structures (precursors of modern tRNAs) were aminoacylated by ribozymes in primitive peptide synthesis mechanisms [20, 33, 34]. The modern tRNA shape has been proposed to arise from the fusion of such RNA minihelices, causing in the process the translocation of the early identity signals to the anticodons and acceptor stems of modern tRNAs [34–39]. Thus, identity elements in the acceptor stem and the genetic code are directly linked through the evolutionary history of tRNA [36].

4.2.3 Extant Aminoacyl-tRNA Synthetases

Aminoacyl-tRNA synthetases (ARS) are classified into two distinct structural families: class I and class II [7–9, 40, 41]. Of the twenty aminoacyl-tRNA synthetases, 10 are found in each family (Table 4.2-1) [9]. All the enzymes in each class evolved from a unique single-domain protein that evolved into the active-site characteristic of each class [20, 42]. The only known exception is lysyl-tRNA synthetase (LysRS), which exists as a class I or as a class II enzyme in different organisms [43].

Genetic dissections showed that aminoacyl-tRNA synthetases developed from their ancestral catalytic cores through the addition of domains and insertions [44, 45]. Crystallographic studies confirmed this scheme [2, 10, 46]. The active-site domain recognizes the acceptor stem end of the tRNA, where the amino acid is attached. Most of the class I and II enzymes also recognize the anticodon stem-loop structure of their cognate tRNAs using additional domains that are idiosyncratic to each enzyme [47].

Table 4.2-1 Classes and subclasses of aminoacyl-tRNA synthetases

Subclass	Class I	Class II	Subclass
	LeuRS	AlaRS	
	IleRS	GlyRS	
	ValRS	ThrRS	
Ia	MetRS	SerRS	IIa
	CysRS	ProRS	
	ArgRS	HisRS	
	GluRS	AspRS	
	GlnRS	AsnRS	
Ib	LysRS	LysRS	IIb
	TyrRS	PheRS	
Ic	TrpRS		IIc

a The classification of ARS is based on sequence and structural information.

All members of class I ARS share an active-site domain that forms a Rossmann nucleotide-binding fold. Members of class II have an active-site domain that contains an unusual anti-parallel β-sheet flanked by two long α-helices [2, 10, 46]. The two folds are fundamentally different. Thus, the two classes evolved from two distinct ancestors. Nevertheless, each class is unlikely to have evolved independently, because the composition and tRNA-binding mechanisms of each class are related and complementary (see below).

Within each class, structural classifications further divide the members into three distinct subclasses (Table 4.2-1) [2]. The amino acids recognized by the enzymes in each subclass are chemically related. Each subclass within class I has a matching subclass in class II that recognizes similar amino acids and contains a similar number of enzymes. Class I ARS are subdivided into three subclasses: Ia, Ib, and Ic. Subclass Ia contains enzymes that are specific for the amino acids leucine, isoleucine, valine, methionine, cysteine, and arginine. Subclass Ib enzymes recognize glutamate, glutamine and lysine. Subclass Ic ARS are specific for tyrosine and tryptophan [2].

Similarly, class II enzymes are subdivided into subclasses IIa, IIb, and IIc. Subclass IIa enzymes are specific for serine, threonine, glycine, alanine, proline, and histidine. Subclass IIb enzymes recognize aspartate, asparagine, and lysine. Subclass IIc contains the enzyme specific for phenylalanine [47]. The active-site domains of class I ARS bind the tRNA from the minor groove side of the acceptor stem. In contrast, the active-site domains of class II ARS bind to the tRNA acceptor stem but, in this case, class II enzymes approach the tRNA molecule from its major groove side. These interactions relate acceptor stem sequences/structures to specific amino acids.

These structural observations are consistent with early work showing that most class I ARS (which bind the minor groove of the tRNA acceptor stem) attach their respective amino acid to the 2′–OH group of the terminal ribose of tRNA, whereas most class II ARS (which bind the major groove of the tRNA acceptor stem) attach the amino acid to the 3′–OH [48, 49]. Important exceptions are synthetases that bind aromatic residues. Tyrosyl- and tryptophanyl-tRNA synthetases (TyrRS and TrpRS) are class Ic enzymes, but they bind the tRNA from the major groove side and indistinctly catalyze the attachment of the amino acid to the 2′– or 3′–OH of the tRNA [48, 49]. On the other hand, phenylalanyl-tRNA synthetase (PheRS), a class IIc enzyme, binds the tRNA on the minor groove side and catalyzes the attachment of the amino acid to the 2′–OH [48, 49]. These exceptions give strong support to the 'symmetrical model' for the origin of the two ARS classes (see below).

In addition to their aminoacylation activity, several subclass Ia enzymes possess an editing activity to prevent misacylation of their cognate tRNAs. Valyl-, leucyl-, and isoleucyl-tRNA synthetases activate cognate amino acids that are difficult to discriminate from stereochemically similar ones [50]. In these enzymes, the hydrolysis of noncognate aminoacyl adenylates or misacylated tRNAs is catalyzed by an independent domain [51]. This editing domain is inserted into the catalytic domain for aminoacylation, thereby creating a separate active site [51, 52]. Class II

enzymes glycyl-, alanyl-, prolyl-, and threonyl-tRNA synthetases (GlyRS, AlaRS, ProRS, and ThrRS) also contain editing activities [53–58]. In the cases of AlaRS, ThrRS, and ProRS, these activities are localized to domains that are appended to the catalytic unit of the enzyme, rather than inserted into the active site [56–58]. Those editing domains are completely different in structure than those found in class I enzymes [54].

4.2.4 The Origin of Aminoacyl-tRNA Synthetase Classes: Two Proteins bound to one tRNA

The homology between extant groups of universal ARS implies that several rounds of gene duplication and divergence took place before LUCA to give rise to most of the enzymes that constitute each class. Given that the role of ARS is intrinsically linked to the development of the genetic code, this observation is consistent with the genetic code being completely defined by the time of LUCA.

Because class I and class II synthetases evolved symmetrically to generate two families that display striking similarities and complementarities, their early evolution was probably driven by common constraints. These evolutionary forces shaped the symmetrical nature of the two classes. Based on the analysis of the crystal structures of complexes between ARS and tRNAs, and the structure of the genetic code, a proposal has been put forward that can explain this feature of the ARS classes [59, 60].

The 'symmetry theory' for the origin of the two classes of ARS proposes that the two classes evolved from an ancestral complex where a single tRNA molecule was recognized simultaneously by a class I and a class II ancestor [59]. The extant subclasses would have originated from duplications of the genes coding for these two proteins. This scenario can explain several features displayed by extant ARS. For example, the equivalence in sizes of the two classes, and their subclasses, would result from coupled evolution. Thus, each event of gene duplication and divergence that generated a new tRNA species was followed by the duplication and divergence of the genes coding for the class I- and class II-type active-site domains (Fig. 4.2-4). This process would result in an equivalent numbers of class I and class II ARS. Similarly, the association of a class I and a class II ARS active site with a given tRNA can explain why the synthetases resulting from the evolution of this initial complex recognize sterically similar residues (see below).

The 'symmetry theory' requires that formation of a complex between a single tRNA and two ARS be sterically possible. The association of two extant ARS on a single tRNA would be prevented by steric clashes caused by domains that surround the enzyme's active sites. However, the ARS ancestors were small proteins that contained only the active-site domain [42]. To investigate the possibility that two ancestral ARS active-site domains formed a complex with a single tRNA molecule, the two-synthetase–one-tRNA interactions were modeled using available crystallographic data [59].

The structures of ARS–tRNA complexes were edited to obtain the coordinates of each tRNA bound only to the respective active site domain [52, 54, 61–67]. The

available structures cover at least one representative from each subclass. (Owing of close similarities between enzymes of the same subclass, the mode of binding to the acceptor stem is thought to be the same for each subclass member.) The structures for all possible subclass Ia–c subclass IIa–c pairs bound to tRNA were individually generated. The resulting structures were inspected for steric compatibility of the bound active-site domains. Not all superimpositions generated sterically compatible models. Several pairs, Similar to that of AspRS (subclass IIb) and IleRS (subclass Ia), generated severe steric clashes between large parts of the respective active sites [59].

Several superimpositions generated compatible pairs where two synthetases cover the tRNA acceptor stem without major steric clashes. Remarkably, these pairs link together specific ARS subclasses. In particular, the only combinations that accommodated all enzymes followed exactly a pairing of subclasses. Thus, subclass Ia enzymes (IleRS or ValRS) pair best with subclass IIa enzymes (SerRS or ThrRS). A subclass Ib enzyme (GlnRS) forms a compatible pair with a subclass IIb enzyme (AspRS). Finally, TyrRS (subclass Ic) can only form a compatible pair with PheRS (subclass IIc) [59] (Fig. 4.2-2).

Large translational and rotational differences between the different pairs (with respect to the axis of the tRNA acceptor stem) are an important feature of these complexes. The differences are particularly evident in the Ic–IIc pair (TyrRS and PheRS), which binds the tRNA acceptor stem at a 90º angle of rotation with respect to the other pairs (Fig. 4.2-2). Thus, ancestral ARS pairs have large variations in their orientations around the tRNA acceptor stem [59].

This analysis supported the idea that the two extant classes of synthetases can be interpreted as a consequence of an early interaction of specific synthetase pairs in complex with tRNA. Among the correct predictions derived from the 'symmetry theory' was the assignment of class I LysRS to the subclass Ib prior to the crystallographic evidence. LysRS is an exception among ARS in that it can be found as a class I or class II enzyme (see below). Class I LysRS were discovered recently, in certain archaebacteria and bacteria. Based on the symmetrical pairings of synthetases that recognize similar residues, we predicted that this new enzyme would be a member of the subclass Ib [59]. The crystal structure of the complex between a class I LysRS and $tRNA^{Lys}$ confirmed our prediction, and showed that the complete structures of a class I and class II LysRSs can form a complex around a single $tRNA^{Lys}$ molecule, with almost no steric hindrance [68] (Fig. 4.2-3).

Moreover, the 'symmetry theory' can also explain the uncanny sequence similarities that have been observed between tRNAs that are charged by ARS of opposite classes (i.e., $tRNA^{Tyr}$ and $tRNA^{Phe}$, or $tRNA^{Asp}$ and $tRNA^{Glu}$). The symmetrical pairs may have formed initially to cover and protect the acceptor stem, in a hostile environment where the structure of RNA was susceptible to chemical degradation or denaturation, or where the ester link between the tRNA molecule and its attached amino acid was particularly labile.

Interestingly, the 'symmetry theory' links the evolution of the two ARS families to the development of the genetic code. If the distribution of the two ARS classes

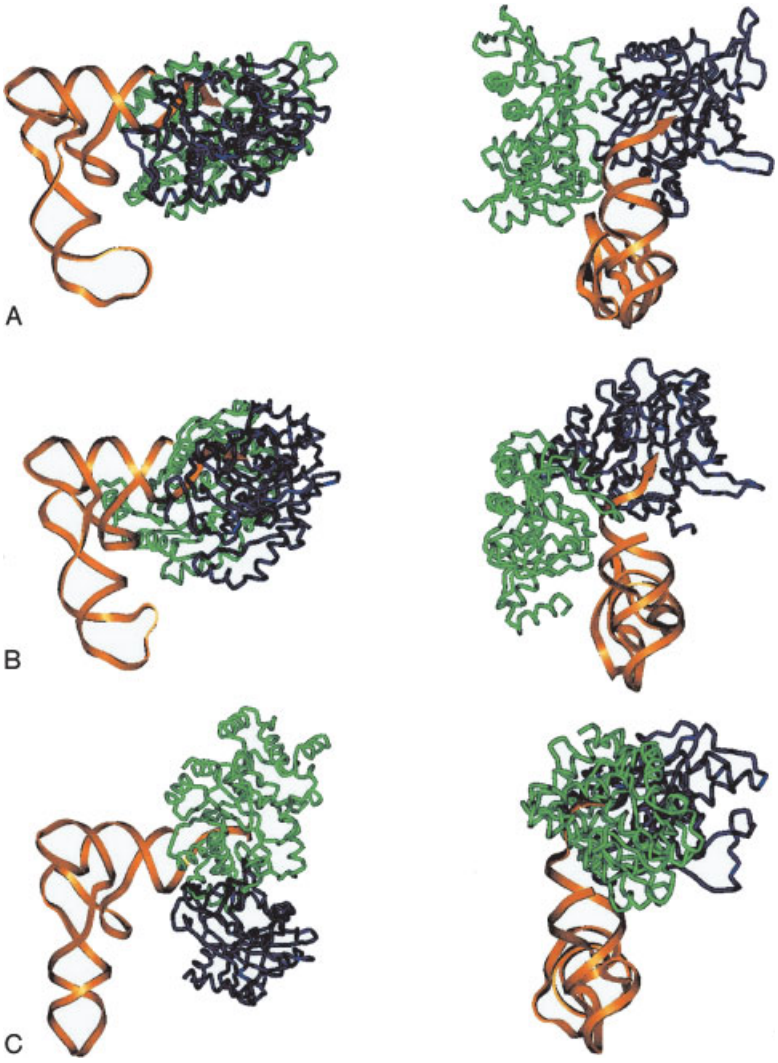

Figure 4.2-2 Graphical representation of complexes formed, respectively, between subclasses Ia–c (green) and IIa–c (blue) synthetase active sites bound simultaneously to a tRNA acceptor stem [59]. The tRNA is depicted in yellow, and each complex is shown in two different orientations. The left complexes are oriented with the plane of the page defined by the axes of the tRNA acceptor stem and anticodon stem helices. The views to the right of the figure show the same molecules along the axis of the anticodon stem-loop, as seen from the acceptor stem side.

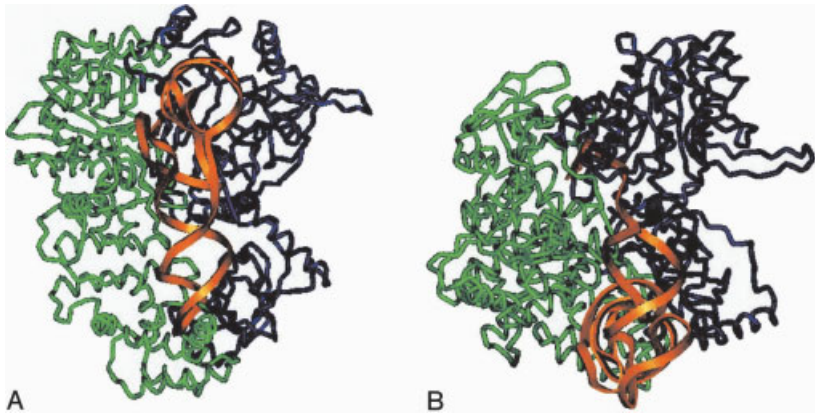

Figure 4.2-3 Graphical representation of the modeled complex between class I and II LysRSs, and $tRNA^{Lys}$ (modified from Ref. [68]). The two proteins were found to be complementary in their mode of binding $tRNA^{Lys}$ [68]. The tRNA molecule is depicted in orange, and the complex is shown in two different orientations as in Fig. 4.2-2.

directly followed the duplications of anticodons, then the ARS paired by the theory are predicted to recognize tRNAs that have related codons. This prediction is largely fulfilled [60], and it produces a general framework within which the growth in complexity of tRNAs and codon families can be examined [60]. The 'symmetrical' theory, however, does not consider the question of the origination of the two class ancestors.

4.2.5
A Common Genetic Origin for all Aminoacyl-tRNA Synthetases ?

Rodin and Ohno [69] first noticed that the coding sequences of class I ARS active sites could be aligned with complementary DNA sequences coding for class II ARS active sites (Fig. 4.2-4). Based on this observation, the genes for the ancestors of the two ARS classes were proposed to be encoded by complementary sequences of RNA or DNA. This proposal offered a mechanistic explanation for the linked duplication of class I and class II ancestral ARS, as stated by the 'symmetry theory'.

More recently, Carter and Duax [70] reported that, in the freshwater mold *Achlya klebsiana,* complementary genes code for proteins having the same folds as class I and II ARS. These genes are complementary to each other in a double-stranded DNA region, with each being transcribed and translated independently. Thus, DNA or RNA complementary strands could have originally coded for the two ARS folds.

The combination of the Rodin–Ohno model and the 'symmetry theory' offers a more detailed explanation for the emergence and evolution of ARS (Fig. 4.2-4). A region of double-stranded RNA or DNA, with complementary coding sequences, would be the starting point. A primitive translation mechanism (possibly RNA-

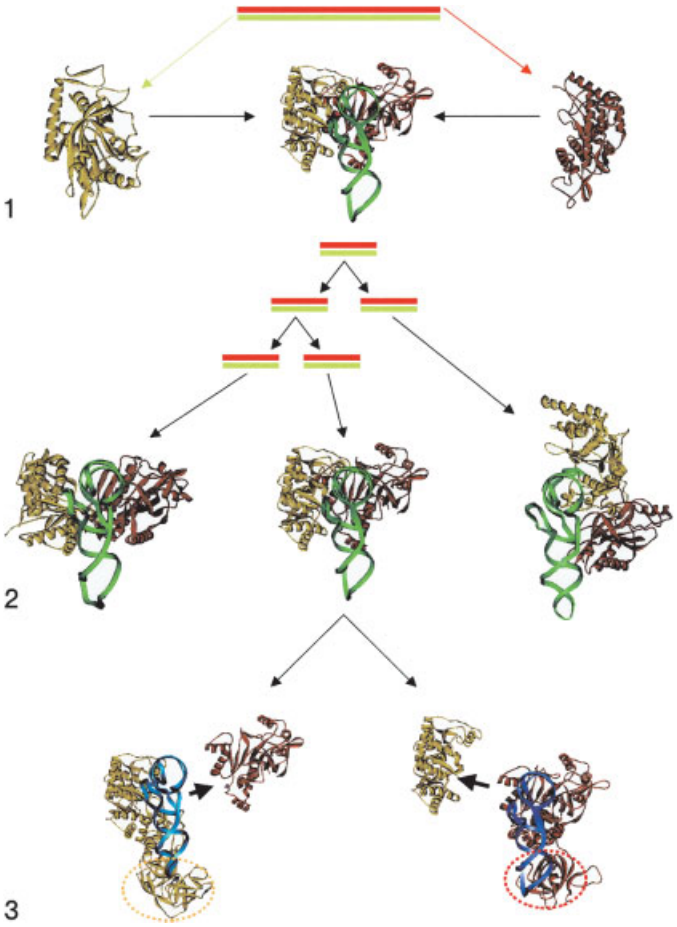

Figure 4.2-4 Evolutionary scheme of the origin and separation of synthetase genes and synthetase pairs, according to the Rodin–Ohno model and the 'symmetry theory' [59, 69]. Panel (1) depicts complementary genes that code for proteins having the same folds as class I and II ARS [70]. These complementary genes code for a pair of ancestral synthetase active sites that bind to a single tRNA (in green) [59]. In this depiction, the class I ancestral domain is shown in yellow on the left side of the complex, and the class II ancestral domain is on the right side in red. During growth of the genetic code, duplication of the complementary genes generates new synthetases that evolve to recognize emerging tRNAs with new identities [59]. Panel (2) represents the idiosyncratic evolution of the new complexes to achieve tRNA and amino acid specificities. Finally, in Panel (3), the ancestral synthetase domain involved in amino acid recognition incorporates other domains (shown as encircled) that allow it to recognize other regions of the tRNA better, giving rise to the modern synthetase structure. The second component of the ancestral pair is lost [60].

based) synthesized two peptides with different folds. These emerging peptides were initially selected for their ability to bind RNA stems simultaneously. Eventually, perhaps through their ability to bring ATP molecules to a pre-existing reaction, these two peptides became functionally involved in the aminoacylation of ancestral tRNA molecules (perhaps minihelix-like structures). The duplication of these tRNA molecules, and divergence of their anticodon sequences, drove the emergence and expansion of the genetic code. Simultaneously, duplication and divergence of the double-stranded region coding for the two ARS ancestors would double the number of available synthetases-like folds, and allow for the evolution of tRNA-specific binding.

It is conceivable that the physical association of the two ARS complementary genes extended to the tRNA genes themselves. Indeed, a comparative sequence analysis of tRNAs suggests that ancestral tRNA genes were coded in pairs by complementary strands of DNA or RNA [36, 69, 71]. This same analysis supported the concept that anticodon sequences arose from duplications of portions of the acceptor stems, thus supporting the idea that the operational RNA code was a precursor to the genetic code.

The proposal that ancestral ARS genes formed complementary DNA or RNA strands has implications for the study of the origin of life. In a primitive RNA world, where metabolic complexity is expected to be lower than in extant organisms, 'double-coding' RNA genomes might have been the norm rather than the exception. If RNA genes evolved to code simultaneously for functional RNA molecules and peptides, a direct physical link could be established between the ancestral machinery for RNA translation (based on ribozymes) and the emerging protein synthesis machinery. More generally, the concept could be extended to other ancient protein families to determine if protein-coding genes can simultaneously code for ribozymes that have the same activities as the proteins they encode. For example, in this scenario, the RNA sequences that code for primitive synthetases might themselves be ribozymes that catalyze aminoacylation.

4.2.5.1 Evolution of Extant Enzymes prior to LUCA

As discussed above, most of the gene duplications that gave rise to extant ARS had been accomplished by the time of appearance of the last universal common ancestor [72–74]. In contrast with phylogenies of whole species, which are not informative about pre-LUCA events, details of pre-LUCA biology might be obtained by analyzing the internal relationships among ARS. Additionally, establishing the order of the duplications that gave rise to the different members of each ARS family would allow us to link this process to the evolution and establishment of the genetic code.

A good example of this type of analysis was provided by studies of sequences of lysyl-tRNA synthetase (LysRS) and $tRNA^{Lys}$. The analysis of the complete genomic sequence of *Methanococcus jannaschii* made apparent that this organism does not contain a gene coding for a canonical class II LysRS [75]. Because all ARSs are essential enzymes, a search was initiated to find the missing activity in related species [43]. The enzyme catalyzing the aminoacylation of $tRNA^{Lys}$ with lysine in

M. maripaludis turned out to be a member of the class I family of ARSs. So far, LysRS represents the only instance of an aminoacylation activity that is catalyzed by a class I, or class II ARS, depending on the species.

The realization that LUCA might have possessed two genes coding for two distinct LysRSs offered, for the first time, the possibility of determining the time of establishment of a tRNA identity relative to the appearance of its cognate ARS. A phylogenetic analysis of the relationships among sequences of $tRNA^{Lys}$ from species bearing class I or class II LysRSs supports the idea that at least one of the extant forms of these enzymes was established in the context of a pre-existing $tRNA^{Lys}$, which remained universally distributed throughout the phylogenetic tree [76]. This prediction subsequently received support from the biochemical analysis of the aminoacylation properties of class I lysyl-tRNA synthetase [77, 78].

Class I LysRSs are mostly limited to archaebacterial species and a small number of eubacteria. On the other hand, class II LysRS are present in all kingdoms of life, including some archaeal species [43, 76]. Explanations for this gene distribution based on a hypothetical late lateral gene transfer of the class I *lysS* genes from bacteria to archaea (or *vice versa*) are not consistent with the phylogenetic analysis of LysRS sequences [43, 76]. More likely, the extant distributions of class I and II LysRSs arose from a situation where an ancestral organism possessed both genes. This redundancy was resolved through the elimination of one of the two genes, either through genetic drift, or by the appearance of selective pressures in favor of one of the two molecules [76].

This situation is, once again, clearly compatible with the 'symmetry' theory. Initially, $tRNA^{Lys}$ was bound by two synthetases, each with the capacity of charging this tRNA with lysine or, at least, of evolving this catalytic activity. As mentioned above, crystallographic studies support the possibility of a complex between $tRNA^{Lys}$ and two LysRSs of opposite classes. The separation of this complex into extant tRNA–ARS interactions took place after LUCA. Most organisms retained the class II fold as LysRS, but some selected the class I LysRS.

4.2.5.2 Changes in Acceptor Stem Identity Elements Correlate with Changes in the Code

The genetic code was first defined as a 'frozen accident' by Francis Crick, who argued that its current structure was due to the fact that its evolution had reached an evolutionary dead-end [79]. Emerging from this cul-de-sac was not possible because the system was incapable of assimilating new changes. This notion of a 'frozen' code has been challenged by the discovery of variations in the code of certain organisms and, more notably, in eukaryotic organelles [80–82]. Nevertheless, the genetic code has remained mostly invariable across the phylogenetic tree. This supports the notion that, for the most part, the code has reached a degree of complexity that does not accept new variations with ease.

As stated above, the 'operational RNA code' for amino acids is the relationship between sequences and structures of acceptor stems and specific amino acids [20, 22, 83]. Through variations of the 'operational RNA code' the genetic code can

change from organism to organism, because any change in the codon–amino acid equivalence has to be adopted by the 'operational RNA code'. Thus, it is conceivable that the frozen state of the genetic code is a consequence of the limitations of the 'operational RNA code'.

Misacylation errors are lethal to cells, and they are prevented through two different mechanisms. On the one hand, potential errors of amino acid recognition (caused by misrecognition of similar residues like isoleucine and valine) are corrected via editing domains contained in the error-prone synthetases [2, 58, 84–88]. On the other hand, potential errors in tRNA recognition are prevented by positive and negative identity elements in each tRNA [31, 32]. But the repertoire of identity elements might have limits. If the capacity of the 'operational RNA code' is limited then new variations in tRNA recognition mechanisms are not possible, because they would result in unacceptable levels of tRNA mischarging by the existing synthetases.

We propose that the fixed state of the genetic code is due to intrinsic limitations of the recognition of tRNAs by aminoacyl-tRNA synthetases. Expansion of the set of tRNAs is restricted because it runs the risk of causing acylation errors. However, incorporation of modifications to the genetic code requires changes in the cellular tRNA set. If the total set of tRNAs in a given organism is reduced, the discrimination problems faced by their cognate synthetases are decreased. This process would facilitate the evolution of tRNA sequences, because the available evolutionary space would increase. In turn, the divergence of tRNA sequences would open the possibility of changes in the genetic code.

Many of the genetic codes found to contain exceptions to the universal codon–amino acid assignments are in animal mitochondria. The first exception to the universal code was detected in the genomes of vertebrate mitochondria, where AUA codes for methionine instead of isoleucine, and UGA codes for tryptophan instead of being a stop triplet. Since that discovery, exceptions to the code have been detected in a large variety of organisms and organelles (reviewed in Ref. [82]). Most of the exceptions, however, are concentrated in metazoans (animals), involving changes of 11 different codons [82].

Additionally, animal mitochondria have experienced a dramatic reduction in their genome size and, in particular, in the number of tRNA genes [89]. If, as we propose, an initial requirement for changes in the code is the relaxation of the recognition constraints between ARS and tRNAs, then the large amount of variations in the genetic code of animal mitochondria should correlate with a large amount of changes in the 'operational RNA code' imbedded in their acceptor stem sequences.

As it can be seen in Table 4.2-2, the percentage of tRNA sequences in mitochondria that contain the recognition elements that are operational in bacteria or eukaryotes is significantly decreased for 16 amino acids. Remarkably, all tRNAs whose identity has been reported to change in mitochondria ($tRNA^{Ile}$, $tRNA^{Arg}$, $tRNA^{Met}$, $tRNA^{Lys}$, and $tRNA^{Ser}$) show important decreases in the conservation of identity elements (Table 4.2-2). Thus, in animal mitochondria, a reduction of tRNA genes is correlated with changes in the mechanisms of recognition between tRNAs and ARS and, simultaneously, with the largest concentration of changes in the genetic code.

Table 4.2-2 Drift of acceptor stem identity elements in animal mitochondrial tRNAs

Amino acid	Identity elements in acceptor stem[a]	Conservation in non-animal mitochondria (%)[b]	Conservation in animal mitochondria (%)[b]
Ala	G3:U70	85	21
Arg	A20	42	5
Asp	G73, G10	92, 80	13, 8
Gln	G2:C71	72	4
Gly	C2:G71, G3:C70	66, 79	46, 0
His	G1	62	0
Ile	C4:C69, C29:G41	47, 52	8, 2
Leu	A73, U8:A14	97, 100	68, 59
Lys	A73, G2:U71	86, 33	70, 10
Met	U4:A69	41	2
Phe	G15:C48	84	2
Pro	G15:C48, A73	37, 88	6, 8
Ser	G73, G2:C71	86, 100	24, 11
Thr	C2:G71	100	38
Tyr	A73	96	86
Val	G3:C70, U4:A69	95, 30	17, 9

a Identity elements in bacterial or eukaryotic tRNAs for review (see Refs. [31, 32]).
b Percentages calculated using all sequences in the Bayreuth tRNA database [90].

This situation was possibly initiated by the reduction of total tRNA genes. This reduction simplified the recognition problem for the mitochondrial synthetases, and allowed the recognition elements to drift into new sequence spaces, thus changing the types of identity elements and 'melting' the genetic code. This process promoted the appearance of new tRNA sequences, because new tRNA variations did not necessarily result in gross aminoacylation errors. These new tRNAs were then capable of acquiring new codon meanings, allowing the genetic code of these organelles to start evolving at a faster pace.

Acknowledgements

This work was supported by grant GM15539 and GM23562 from the National Institutes of Health and by a Fellowship from the National Foundation for Cancer Research.

References

1 P.R. Schimmel, D. Söll, *Ann. Rev. Biochem.* **1979**, *48*, 601–648.
2 S. Cusack, *Curr. Opin. Struct. Biol.* **1997**, *7*, 881–889.
3 A. Tzagoloff, D. Gatti, A. Gampel, *Prog. Nucl. Acid. Res. Mol. Biol.* **1990**, *39*, 129–158.
4 C.R. Woese, G.J. Olsen, M. Ibba et al., *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 202–236.
5 C. Woese, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6854–6859.
6 C.R. Woese, *Microbiol. Rev.* **1987**, *51*, 221–271.
7 T. Webster, H. Tsai, M. Kula et al., *Science* **1984**, *226*, 1315–1317.
8 S. Cusack, C. Berthet-Colominas, M. Hartlein et al., *Nature* **1990**, *347*, 249–255.
9 G. Eriani, M. Delarue, O. Poch et al., *Nature* **1990**, *347*, 203–206.
10 D. Moras, *Trends Biochem. Sci.* **1992**, *17*, 159–164.
11 L. Ribas de Pouplana, P. Schimmel, *J. Biol. Chem.* **2001**, *276*, 6881–6884.
12 T. Hashimoto, L.B. Sanchez, T. Shirakura et al., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6860–6865.
13 J.W. Chihade, J.R. Brown, P.R. Schimmel et al., *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12153–12157.
14 Y.M. Hou, P. Schimmel, *Nature* **1988**, *333*, 140–145.
15 W.H. McClain, K. Foss, *Science* **1988**, *240*, 793–796.
16 S.J. Park, P. Schimmel, *J. Biol. Chem.* **1988**, *263*, 16527–16530.
17 A.D. Yaremchuk, M.A. Tukalo, I.A. Krikliviy et al., *J. Mol. Biol.* **1992**, *224*, 519–522.
18 H. Himeno, T. Hasegawa, T. Ueda et al., *Nucl. Acid. Res.* **1990**, *18*, 6815–6819.
19 M.E. Saks, J.R. Sampson, *EMBO J.* **1996**, *15*, 2843–2849.
20 P. Schimmel, R. Giegé, D. Moras et al., *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8763–8768.
21 C. de Duve, *Nature* **1988**, *333*, 117–118.
22 K. Musier-Forsyth, P. Schimmel, *Acc. Chem. Res.* **1999**, *32*, 368–375.
23 C. Francklyn, P. Schimmel, *Nature* **1989**, *337*, 478–481.
24 M. Frugier, C. Florentz, R. Giegé, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3990–3994.
25 M. Frugier, C. Florentz, R. Giegé, *EMBO J.* **1994**, *13*, 2219–2226.
26 S.A. Martinis, P. Schimmel: in *tRNA, Structure, Biosynthesis, and Function*, eds D. Soll, U.L. RajBhandary, ASM Press, Washington, DC 1994, 349–370.
27 M.E. Saks, J.R. Sampson, J.N. Abelson, *Science* **1994**, *263*, 191–197.
28 O. Nureki et al.: in *The Translational apparatus*, eds K.H. Nierhaus, F. Franceschi, A.R. Subramanian et al., Plenum, New York 1993, 59–66.
29 C.S. Hamann, Y.M. Hou, *Biochemistry* **1995**, *34*, 6527–6532.
30 B. Felden, R. Giegé, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 10431–10436.
31 R. Giegé, M. Sissler, C. Florentz, *Nucleic Acids Res* **1998**, *26*, 5017–5035.
32 P.J. Beuning, K. Musier-Forsyth, *Biopolymers* **1999** *52*, 1–28.
33 P. Schimmel, B. Henderson, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11283– 11286.
34 B.S. Henderson, P. Schimmel, *Bioorg. Med. Chem.* **1997**, *5*, 1071–1079.
35 W. Moller, G.M. Janssen, *J. Mol. Evol.* **1992**, *34*, 471–477.
36 S. Rodin, S. Ohno, A. Rodin, *Orig. Life Evol. Biospys.* **1993**, *23*, 393–418.
37 M. Di Giulio, *Orig. Life Evol. Biosph.* **1994**, *24*, 425–434.
38 T.P. Dick, W.A. Schamel, *J. Mol. Evol.* **1995**, *41*, 1–9.
39 S. Rodin, A. Rodin, S. Ohno, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 4537–4542.
40 C. Hountondji, P. Dessen, S. Blanquet, *Biochimie* **1986**, *68*, 1071–1078.
41 S.W. Ludmerer, P. Schimmel, *J. Biol. Chem.* **1987**, *262*, 10801–10806.
42 P. Schimmel, L. Ribas de Pouplana, *Cell* **1995**, *81*, 983–986.

43 M. Ibba et al., *Science* **1997**, *278*, 1119–1122.
44 M. Jasin, L. Regan, P. Schimmel, *Nature* **1983**, *306*, 441–447.
45 K. Shiba, P. Schimmel, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 1880–1884.
46 C.W. Carter, Jr., *Ann. Rev. Biochem.* **1993**, *62*, 715–748.
47 S. Cusack, *Biochimie* **1993**, *75*, 1077–1081.
48 M. Sprinzl, F. Cramer, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 3049–3053.
49 T.H. Fraser, A. Rich, *Proc. Natl. Acad. Sci. USA* **1975**, *72*, 3044–3048.
50 P. Schimmel, E. Schmidt, *Trends Biochem. Sci.* **1995**, *20*, 1–2.
51 L. Lin, S.P. Hale, P. Schimmel, *Nature* **1996**, *384*, 33–34.
52 O. Nureki et al., *Science* **1998**, *280*, 578–582.
53 W.C. Tsui, A.R. Fersht, *Nucleric. Acid. Res.* **1981**, *9*, 4627–4637.
54 R. Sankaranarayanan et al., *Cell* **1999**, *97*, 371–381.
55 P.J. Beuning, K. Musier-Forsyth, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8916–8920.
56 A. Dock-Bregeon et al., *Cell* **2000**, *103*, 877–884.
57 P.J. Beuning, K. Musier-Forsyth, *J. Biol. Chem.* **2001**, *276*, 30779–30785.
58 K. Beebe, L. Ribas de Pouplana, P. Schimmel, *EMBO J.* **2003**, *22*, 668–675.
59 L. Ribas de Pouplana, P. Schimmel, *Cell* **2001**, *104*.
60 L. Ribas de Pouplana, P. Schimmel, *Trends Biochem. Sci.* **2001**, *26*, 591–596.
61 V. Biou, A. Yaremchuk, M. Tukalo et al., *Science* **1994**, *263*, 1404–1410.
62 S. Fukai, O. Nureki, S. Sekine et al., *Cell* **2000**, *103*, 793–803.
63 H. Bedouelle, *Biochimie* **1990**, *72*, 589–598.
64 L.F. Silvian, J. Wang, T.A. Steitz, *Science* **1999**, *285*, 1074–1077.
65 M.A. Rould, J.J. Perona, D. Söll et al., *Science* **1989**, *246*, 1135–1142.
66 M. Ruff et al., *Science* **1991**, *252*, 1682–1689.
67 Y. Goldgur, L. Mosyak, L. Reshetnikova et al., *Structure* **1997**, *5*, 59–68.
68 T. Terada, O. Nureki, R. Ishitani et al., *Nat. Struct. Biol.* **2002**, *9*, 257–262.
69 S.N. Rodin, S. Ohno, *Orig. Life Evol. Biospys.* **1995**, *25*, 565–589.
70 C.W. Carter, W.L. Duax, *Mol. Cell* **2002**, *10*, 705–708.
71 S. Rodin, S. Ohno, A. Rodin, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4723–4727.
72 G.M. Nagel, R.F. Doolittle, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 8121–8125.
73 G.M. Nagel, R.F. Doolittle, *J. Mol. Evol.* **1995**, *40*, 487–498.
74 J.R. Brown, W.F. Doolittle, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2441–2445.
75 C.J. Bult et al., *Science* **1996**, *273*, 1058–1073.
76 L. Ribas de Pouplana, R.J. Turner, B.A. Steer et al., *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11295–11300.
77 M. Ibba, H.C. Losey, Y. Kawarabayasi et al., *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 418– 423.
78 P. Schimmel, L. Ribas de Pouplana, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 327–328.
79 F.H. Crick, *J. Mol. Biol.* **1968**, *38*, 367–379.
80 S. Osawa, D. Collins, T. Ohama et al., *J. Mol. Evol.* **1990**, *30*, 322–328.
81 M.A. Santos, T. Ueda, K. Watanabe et al., *Mol. Microbiol.* **1997** *26*, 423–431.
82 M.A.S. Santos, M.F. Tuite: in *The Genetic Code and the Origin of Life*, eds L. Ribas de Pouplana, Landes Biosciences, Houston 2003, in press.
83 P. Schimmel, *J. Mol. Evol.* **1995**, *40*, 531–536.
84 E.W. Eldred, P.R. Schimmel, *Biochemistry* **1972**, *11*, 17–23.
85 A.A. Schreier, P.R. Schimmel, *Biochemistry* **1972**, *11*, 1582–1589.
86 A.R. Fersht, C. Dingwall, *Biochemistry* **1979**, *18*, 2627–2631.
87 E. Schmidt, P. Schimmel, *Science* **1994**, *264*, 265–267.
88 L. Lin, S.P. Hale, P. Schimmel, *Nature* **1996**, *384*, 33–34.
89 C.G. Kurland, *Bioessays* **1992**, *14*, 709–714.
90 M. Sprinzl, K.S. Vassilenko, J. Emmerich et al., http://www.uni-bayreuth.de/departments/biochemie/trna/ 1999.