



**HAL**  
open science

# Transfer RNA Recognition and Aminoacylation by Synthetases

Richard Giegé, Gilbert Eriani

► **To cite this version:**

Richard Giegé, Gilbert Eriani. Transfer RNA Recognition and Aminoacylation by Synthetases. Encyclopedia of Life Sciences, 1, Wiley, 2021, 10.1002/9780470015902.a0029242 . hal-03394741

**HAL Id: hal-03394741**

**<https://hal.science/hal-03394741v1>**

Submitted on 22 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Transfer RNA Recognition and Aminoacylation by Synthetases

Richard Giegé and Gilbert Eriani

Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire, Architecture et Réactivité de l'ARN, CNRS UPR9002, 2, allée Konrad Roentgen, F-67084 Strasbourg (France)

**Abstract:** Fidelity of transfer ribonucleic acid (tRNA) charging by amino acids ensures correct translation of the genetic code into proteins. Charging is catalysed by a set of enzymes known as aminoacyl-tRNA synthetases. Owing to the degeneracy of the genetic code, some of the different tRNAs have the same amino acid attached to them. Specificity of charging obeys universal rules and is ensured by positive elements, the identity determinants unique to each tRNA and responsible for its recognition by the cognate synthetase, and negative elements, the antideterminants that prevent false recognitions. To fulfil the aminoacylation specificity and prevent noncognate aminoacyl-tRNA delivery to the ribosome, some synthetases also mediate proofreading reactions that increase fidelity of the tRNA charging. In such reactions, misactivated amino acids or mischarged tRNAs are checked in specific sites and noncognate products are hydrolysed. However, mischarging is beneficial under certain stress circumstances or when catalysed by nondiscriminatory synthetases, and represents a driving force in evolution.

### Key Concepts:

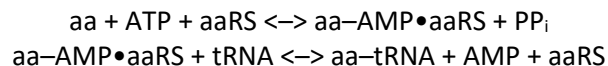
- Translational expression of the genetic code refers to aminoacyl-tRNA- and ribosome-dependent decoding of genes into proteins, a process highly dependent on fidelity of tRNA aminoacylation by synthetases.
- The rules that account for the aminoacylation identity of tRNAs are referred to as the second genetic code.
- The RNA operational code is encoded in the acceptor stem of tRNA and is crucial for recognition by aminoacyl-tRNA synthetases and specific aminoacylation.
- Allostery in tRNA-synthetase systems concerns long-range transfer of chemical information (up to 75 Å) to the synthetase catalytic site (through the body of tRNA and/or synthetase) triggered by contacts of tRNA identity determinants with the synthetase.
- Engineering the identity of tRNA-synthetase systems allows reprogramming the genetic code.

### Keywords:

Allostery, aminoacyl-tRNA synthetase, genetic code, identity antideterminant, identity determinants, protein synthesis, RNA recognition, translation, tRNA, tRNA post-transcriptional modification.

## Introduction

In translation of the genetic information from nucleic acids to proteins, transfer ribonucleic acids (tRNAs) occupy a key position. They are the link between the genetic information encoded by the messenger ribonucleic acids (mRNAs) and their functional peptidic products. On the ribosome, each mRNA reads a given codon on the small subunit and dictates incorporation of the appropriate amino acid into the growing protein chains on the peptidyltransferase site on the large subunit. Because of the degeneracy of the genetic code there is, in general, more than one tRNA for each amino acid. In the cell, the amino acids are charged on the corresponding tRNAs by a family of enzymes called aminoacyl-tRNA synthetases (aaRSs) (Giegé and Springer, 2016). With some exceptions there is only one aaRS per isoacceptor tRNA family (here isoacceptor is used in the sense of tRNA which carry the same amino acid but can have different anticodons). Generally, there are 21 isoacceptor tRNA families, one per standard amino acid plus one for selenocysteine. In addition, some methanogenic archaea and bacteria contain a 22<sup>nd</sup> isoacceptor tRNA for pyrrolysine. AaRSs must recognise, activate and charge the correct amino acid on the cognate tRNA(s). This implies specific recognition within a group of molecules presenting sufficiently similar primary, secondary and tertiary structures in order to fit interchangeably into the ribosome-dependent translation apparatus. This constraint limits the recognition strategies to discrimination between bases at specific positions within the tRNA framework and led to the concept of tRNA identity stipulating that a limited number of signals in tRNA govern recognition by aaRSs and aminoacylation specificity (see Giegé *et al.*, 1998) and more generally to the ideas of the 'RNA operational code' (Schimmel *et al.*, 1993). The fidelity of tRNA in translation depends on anticodon–codon interactions, but more important, it depends on the interaction between tRNAs and aaRSs. This interaction is not rigorously specific since tRNA mischarging is possible and is even a biological necessity in organisms lacking AsnRS and/or GlnRS. Charging of a tRNA by a wrong amino acid can lead to the erroneous incorporation of this amino acid into a protein, therefore some aaRSs can have editing activity to hydrolyse acyl-tRNA bond. Ultimately, the genetic code is established by the correctness of tRNA charging, however a certain amount of mischarging is tolerate and may be beneficial under certain stress circumstances. Aminoacylation of tRNA is achieved in two-step reactions in which aaRSs activate first the appropriate amino acid with adenosine triphosphate (ATP) in the mandatory presence of Mg<sup>2+</sup> before to transfer the activated amino acid to the tRNA



Specificity is determined by protein–RNA interactions and kinetic features. Recognition of ATP and catalysis also implies specific mechanistic schemes in the 20 aminoacylation systems. Structural similarities and differences in the catalytic domains of aaRSs and the presence of exclusive consensus motifs have led to the classification of these enzymes into two groups, the class I and class II aaRSs (Cusack *et al.*, 1990; Eriani *et al.*, 1990).

## A. Identity for tRNA Recognition by Synthetases

### Towards the concept of tRNA identity

#### **Functional background**

How synthetases achieve their specificity was perhaps the first protein–RNA recognition problem to be seriously investigated. Early data in the 1970s pinpointed the importance of the tRNA accepting end and of the anticodon for recognition by aaRSs. It was also realised that specificity of aaRSs for amino acid activation and tRNA aminoacylation is rather low and that accurate tRNA charging relies more on kinetic effects (reduction of catalytic efficiency  $k_{cat}/K_m$  by factors of up to  $10^5$ -fold and more) than on discrimination among cognate and noncognate tRNAs through binding affinity. This led to the proposal of kinetic specificity of tRNA aminoacylation (Ebel *et al.*, 1973) and implies mutual tRNA–aaRS adaptation that is dependent on sequence and three-dimensional features of the interacting molecules. These features were eagerly searched and became the structural basis of tRNA identity.

Binding of tRNAs on aaRSs and proper catalysis require ‘identity elements’ that are involved in the specific recognition by the cognate aaRS, plus additional signals that prevent recognition of the tRNA by noncognate aaRSs. Identity elements in a given tRNA are defined as the set of signals, unique to that tRNA, required for its efficient aminoacylation by the cognate aaRS. Identity elements can be transferred to noncognate tRNAs, allowing them to be charged with the corresponding amino acid. Loss of identity elements in a tRNA leads to a decrease in aminoacylation specificity that can be sufficient to impair cell growth. Identity is known to require only a small number of specific nucleotides. Experimental advances and theoretical considerations on the origin of tRNA aminoacylation systems (Schimmel *et al.*, 1993) indicate that universal rules govern tRNA identity, with an initial role of the accepting branch that contacts the catalytic site of aaRSs and subsequent participation of the tRNA anticodon domain (but not mandatory), as well as of other structural features, including the overall shape of tRNA. Idiosyncratic features fine-tune specificities of the different systems and also distinguish identities for a same amino acid specificity in different biological species.

#### **Structural background**

Early investigations on tRNA identity in the 1980s were based on limited structural data on tRNAs and synthetases, except a rather large panel of tRNA sequences and the first crystallographic structure of a tRNA, that of yeast tRNA<sup>Phe</sup>. Today structural data for both tRNA and aaRSs cover the Tree of Life. Crystallographic structures are known for tRNAs of most amino acid identities, either free or in interaction with macromolecular partners. Altogether, a generalised cloverleaf structure of tRNA that highlights location of the constant nucleotides accounting for the tRNA L-shaped architecture could be proposed (Figure 1).

Importantly, the molecular basis for tRNA recognition by aaRSs relies on the crystal structures of approximately 80 aaRS:tRNA complexes. They correspond to 10 class I and 8 class II identities, with only the Asn and Lys complexes missing. Noteworthy, many originate from *Bacteria* and were solved with unmodified tRNA transcripts. The two seminal structures (with modified tRNA) from class I *Escherichia coli* Glu (Rould *et al.*, 1989) and class II yeast (*Saccharomyces cerevisiae*) Asp (Ruff *et al.*, 1991) systems show interaction schemes that are characteristic for tRNA recognition by class I and class II aaRSs (Figure 2). The case of dimeric  $\alpha_2$  class I TyrRS and tetrameric  $(\alpha\beta)_2$  class II PheRS is peculiar with tRNA binding across the synthetase subunits.

## Characterising identity elements in tRNA for aminoacylation

### ***Determinants or positive identity elements***

Three main methods were used to find identity determinants. In its simplest version, the first one consists in searching consensus sequences, for example, of tRNAs charged by a same synthetase (see Giegé *et al.*, 1993). In more elaborated studies, first applied to *E. coli* tRNAs and developed after 2005 by Ardell and coworkers (Collins-Hed and Ardell, 2019), the search was done by computational methods. Presently it is possible to screen very large data sets of tRNA sequences (e.g. Galili *et al.*, 2016; Branciamore *et al.*, 2018; Zamudio *et al.*, 2020).

The second method, widely used, is *in vitro* transcription of artificial tRNA genes that can readily produce any type of tRNA transcript for enzymatic or physical studies. It takes advantage of advances in DNA synthesis, allowing cloning of tRNA genes with any number of nucleotide changes. *In vitro* transcription of tRNA genes using T7 RNA polymerase has allowed preparation of milligram quantities of specific tRNAs. However, the method does not evaluate the role of post-transcriptional modifications in identity. Fortunately, most wild-type transcripts lacking base modifications are substrates for aaRSs because only in few cases have base modifications been shown to influence the identity of a tRNA. *In vitro* assays can evaluate the effects of mutations on the catalysis of tRNA charging (i.e. the strength of determinants). Atomic mutagenesis (removal or replacement of chemical groups in tRNA) would be the method of choice to discover the chemical signals in tRNA important for identity (Musier-Forsyth and Schimmel, 1999). But due to practical difficulties, the approach is restricted to those systems where aaRSs can aminoacylate small RNA substrates (Martinis and Schimmel, 1995).

The third method that yielded important information on tRNA identity has resulted from *in vivo* assays using nonsense suppression (e.g. McClain and Foss, 1988). Here the anticodon of a tRNA must be changed to one that is complementary to a nonsense codon. A nonsense test codon is introduced in a reporter protein, and suppression of the codon leads to a readily analysable active protein. The fact that changes in an anticodon often affect tRNA recognition by an aaRS is a limitation of the method. Results can also be affected by the level of synthesis of the mutated tRNAs, the type and extent of base modifications, and the interaction with ribosomes and translation factors. However, *in vivo* results take into account cellular parameters like competition between the different tRNAs and different aaRSs as well as possible correction mechanisms. Recently a new *in vivo* approach (McShane *et al.* 2016) was used to refine the yeast Arg identity set obtained by the *in vitro* method.

These methods were essentially employed to detect identity determinants in *E. coli* and *S. cerevisiae* tRNAs (for 20 and 18 identities) and only scarce data concern tRNAs from other organisms. Explicit proof of the identity nature of a tRNA residue (or of a set of residues) comes from transplantation experiments, when the suspected residue(s) transferred in another tRNA confer the new identity to this tRNA. In practice, tRNAs with transplanted identity sets seldom recover the catalytic aminoacylation efficiency of the donor tRNA and completely lose their initial identity, meaning that some identity signals are missing.

Table 1 summarises the data confirmed by transplantation experiments. They indicate a clustering of identity determinants in the distal extremities of the tRNA molecule and in a few cases the presence of determinants in the core region of tRNA. Only a limited number of determinants (2–11 nts) specify a given identity. All 20 tRNA families show determinants in the acceptor stem and 18 in the anticodon loop, except the tRNA<sup>Ala</sup> and tRNA<sup>Ser</sup> families. However, only 11 identities utilise the complete anticodon triplet as a determinant (Cys, Gln, Ile, Met, Trp

and Asn, Asp, His, Lys, Phe, Thr) (Figure 1). Noteworthy, identity sets can overlap (i.e. a given nt at a same tRNA position can code for several identities in a given organism) such as A<sub>73</sub> coding for 13 identities or C<sub>35</sub> coding for 4 identities. At first sight, these conclusions seem to apply for tRNAs from all over the Tree of Life, since identity nts (predicted by sequence inspection) are well conserved, suggesting that universal rules govern identity. However, phyla-dependent idiosyncrasies are observed such as variable number of determinants for a given identity set (e.g. for Phe identity), the strength of determinants (e.g. Asp or Met determinants much stronger in *E. coli* than in yeast) or even the nonconservation of determinants in evolution (e.g. for Tyr identity with G<sub>1</sub>•C<sub>72</sub> in *Bacteria* and *Archaea* and C<sub>1</sub>•G<sub>72</sub> in *Eukaryota*). Altogether, 3 identities (Gly, His, Tyr) are not fully conserved in evolution.

### **Antideterminants or negative identity elements**

In addition to positive elements (determinants), negative elements (antideterminants) contribute to identity by blocking false recognitions between synthetases and noncognate tRNAs (Table 2). Two modified bases in the anticodon loop of tRNA are the first antideterminants that were discovered. They are lysidine (a hypermodified C abbreviated k<sup>2</sup>C) at the first anticodon position in a minor *E. coli* tRNA<sup>Ile</sup> that is responsible for rejection of this tRNA by MetRS (Muramatsu *et al.*, 1988) and a methylated G at position 37 in yeast tRNA<sup>Asp</sup> that is an antideterminant against ArgRS (Perret *et al.*, 1990). Interestingly, k<sup>2</sup>C<sub>34</sub> is also a positive element for aminoacylation by IleRS. Likewise, the identity determinants in *E. coli* tRNA<sup>Gln</sup> (see Table 1) are antideterminants in *E. coli* tRNA<sup>Glu</sup> and vice versa for the determinants in tRNA<sup>Glu</sup> that are antideterminants in tRNA<sup>Gln</sup> (Hadd and Perona, 2014).

Amino acids that act as negative signals have been found, first in bacterial MetRS and TyrRS. Thus, *E. coli* MetRS mutants able to charge ochre and amber tRNA<sup>Met</sup> possess acidic residues (e.g. Asp449 and Asp456) involved in rejection of tRNAs carrying nonmethionine anticodons. Their substitution renders possible recognition of nonsense suppressors, without affecting recognition of native tRNA<sup>Met</sup> (Schmitt *et al.*, 1993). Note the presence in aaRSs of amino acids with dual positive and negative functions. Evidence came from studies on yeast AspRS (Ador *et al.*, 1999) and *E. coli* GlnRS (Bullock *et al.*, 2003) that characterised amino acids rejecting noncognate tRNAs and interacting positively with the cognate tRNA. Thus, if one considers that chemical groups ensuring positive and negative interactions have to be accessible, clustering both functions on a single residue may be a way to optimise discrimination by minimising the number of residues required for specificity, thereby reducing the structural complexity of the enzyme.

### **Role of nucleotide modification**

In general, tRNA modifications do not participate in tRNA identity. However, there are cases where they are crucial, either as positive determinants or negative antideterminants. To date, 9 modified nts (Figure 1) were explicitly characterised as positive determinants in 12 tRNAs (see Giegé and Lapointe, 2009). These residues play an active role in aminoacylation catalysed by *E. coli* IleRS, GluRS, GlnRS, LysRS, TyrRS, yeast IleRS, TyrRS, PheRS and archaeal SepRS (an atypical aaRS that charges phosphoserine on tRNA<sup>Cys</sup>, see below). They are all located in anticodon loops and more precisely within anticodons and at neighbouring N<sub>37</sub>. Most of them are hypermodified and have a dual function in aminoacylation and codon reading. A typical example is k<sup>2</sup>C<sub>34</sub> in *E. coli* tRNA<sup>Ile</sup><sub>2</sub> anticodon. Although the gene of this tRNA codes for a Met CAU anticodon, replacement of C<sub>34</sub> by k<sup>2</sup>C<sub>34</sub> allows incorporation of Ile on AUA instead AUG codons and specifies aminoacylation by IleRS instead MetRS. Likewise, hypermodified derivatives of U<sub>34</sub> of *E. coli* tRNA<sup>Glu</sup> are strong Glu identity element and influence codon reading. This limited panel does not

mean that only a few aaRSs use post-transcriptional modifications of tRNA as identity signals but more likely reflects the fact that their search was limited to only few investigations. Other modified nucleotides outside the anticodon loop, participate collectively in tuning the competent configuration of individual tRNAs for optimal interaction with their cognate aaRSs. The effect is especially striking for *E. coli* tRNA<sup>Phe</sup> where the absence of its 10 modified nts strongly impairs phenylalanylation by *E. coli* PheRS (Tinkle-Peterson and Uhlenbeck, 1992). In contrast, modified residues are without effect on expression of Ala identity in *E. coli* or Asp and Tyr identities in yeast.

## Functional role of tRNA domains

### ***Acceptor stem and 'discriminator' base at position 73***

The amino acid acceptor region was postulated to contain the primordial recognition elements that appeared in the course of tRNA evolution. One reason for that is the proximity of the tRNA 3'-terminus with the synthetase active site. Other reasons rely on the activity of nonsense suppressors derived from glutamine-, leucine-, serine- and tyrosine-specific tRNAs that have their anticodon replaced by nonsense anticodons. These molecules can be charged but on mutagenesis in the most terminal part of their acceptor stems they can lose activity or acquire new identities. Finally, the ability of aaRSs to specifically aminoacylate minimalist RNA structures restricted to the whole or part of the amino acid accepting branch of tRNA (the so-called mini- or microhelices) (Table 3), although with reduced efficiency, brought the decisive proof showing that binding of the tRNA anticodon branch on aaRSs is not compulsory for tRNA charging. This ability, together with considerations on evolution, led to the concept of 'RNA operational code' (Schimmel *et al.*, 1993). Figure 1 shows that the first five base pairs, N<sub>-1</sub>, and the fourth base from the 3'-end (the unpaired discriminator N<sub>73</sub>) are used as identity determinants. Noticeable is the wide usage of the discriminator base, of the first three base pairs, and in the case of tRNA<sup>Ala</sup> the role of the G<sub>3</sub>-U<sub>70</sub> pair that is the major signal required for Ala identity (Hou and Schimmel, 1988; McClain and Foss, 1988), except in mitochondria (Zeng *et al.*, 2020). The His system is unique, since all tRNA<sup>His</sup> species have an additional residue at their 5'-end (mainly G<sub>-1</sub>, sometimes U<sub>-1</sub>). This residue forms a standard base pair with the discriminator base N<sub>73</sub>. The large decrease of the charging activity of tRNA mutants lacking G<sub>-1</sub> (in *E. coli* tRNA<sup>His</sup>) and thus lacking the extra G<sub>-1</sub>•C<sub>73</sub> pair, indicates the key importance of this peculiarity for recognition by HisRS. The exact role of this additional pair was probed by atomic mutagenesis (Rosen and Musier-Forsyth, 2004). Results indicate that G<sub>-1</sub> serves to position the 5'-monophosphate, which is critical for aminoacylation, and additionally that C<sub>73</sub> and G<sub>-1</sub> contain exocyclic atomic groups located in the major groove of the accepting RNA helix that contribute to HisRS recognition. Note a prevalence of identities coded by class II aaRSs, in line with their possible ancestry (see below).

The N<sub>3</sub>•N<sub>70</sub> pair deserves particular attention since it is part of 10 identity sets recognised by class I (Cys, Gln, Leu, Met, Trp, Val) and class II (Ala, Gly, Ser, Thr) aaRSs. Except for the strong universally conserved Ala identity determinant (G<sub>3</sub>-U<sub>70</sub>), it is always a weak nonconserved N<sub>3</sub>•N<sub>70</sub> pair. Crystal structure of *Archaeoglobus fulgidus* AlaRS provided the basis for the G<sub>3</sub>-U<sub>70</sub> recognition that allows the correct positioning of the 3'-CCA accepting end in the catalytic site (Naganuma *et al.*, 2014).

Altogether, the data show that the presence of an unpaired base at the discriminator position (nt 73) brings chemical groups able to specifically interact with amino acids of the synthetases

close to the CCA acceptor end in the active site. Therefore, the discriminator base is highly critical to provide specificity during tRNA recognition and convert the resulting binding in catalytic efficiency. Compared to the other positive determinants found in the acceptor stem (Table 1) the discriminator 73 is almost universally found as determinant in all tRNA isoacceptors, further highlighting its critical function.

### ***Anticodon and anticodon branch***

Many speculations and experiments focused on the anticodon branch, the second structural domain of tRNA (Figure 1). The anticodon itself, at a distance of approximately 75 Å from the accepting end, would be the most logical site to specify the amino acid to be coupled to a tRNA. Inactivation of several tRNAs after chemical or enzymatic modification of their anticodons was a strong argument in favour of this possibility. Subsequent studies on mutagenized tRNAs have explicitly shown that changes in anticodons have dramatic effects on the activity of many tRNAs (e.g. tRNA<sup>fMet</sup>, tRNA<sup>Arg</sup>, tRNA<sup>Val</sup>). These results were a prelude to numerous others that definitely demonstrated that anticodon bases are implicated in recognition in most isoaccepting tRNA families. Now it is established that the anticodons of 18 *E. coli* tRNAs contain at least one important recognition element for cognate aaRSs (Figure 1). The tRNAs that do not contain identity elements in the anticodon (tRNA<sup>Ser</sup> and tRNA<sup>Ala</sup>) include isoacceptors differing at the anticodon level so that they can recognise the multiplicity of codons (up to six) they have to decode. The high degeneracy of anticodons in these tRNAs probably accounts for the peculiar recognition requirements of the corresponding aaRSs. Beside the anticodon triplets, positions 37 and 38 of the anticodon loop are also involved in the identity of several tRNAs (Table 1). Likewise, the A<sub>31</sub>•U<sub>39</sub> pair from the anticodon stem of *E. coli* tRNA<sup>Tyr</sup> affects identity, since its replacement by a G<sub>31</sub>•C<sub>39</sub> pair affects its mischarging capacity by GlnRS and LysRS, respectively increased and decreased [see (Giegé and Springer, 2006) for Yokogawa data].

In summary, the critical elements involved in tRNA recognition are mainly located at the two distal extremities of the tRNA. Near the acceptor end the discriminator base is a major determinant. On the opposite end the nucleotides at positions 34-36 of the anticodon are equally critical, the 3 together or in various combinations. Both constitute the two essential recognition poles of the tRNA molecule by aaRSs.

### ***Other recognition sites***

Beside the determinants within the two distal domains of tRNA, there are structural features located in the tRNA core that contribute to identity. This core consists of a network of largely conserved tertiary interactions that connect the D- and T-loops to the variable region. The network maintains all tRNAs in a similar three-dimensional fold (Figure 1). However, faint idiosyncratic architectural differences exist and characterise each tRNA.

*E. coli* tRNA<sup>Cys</sup> is unique in that one of the universally conserved tertiary interaction, the so-called 15–48 Levitt pair (G–C or A–U pairs in reverse WC conformation), is replaced by an atypical G<sub>15</sub>–G<sub>48</sub> interaction. As could be anticipated, this unusual tertiary pair plays a role in Cys identity. Further, the interdigitation of the D- and T-loops reveals a cluster of nonconserved residues called ‘variable pocket’. These form a specific three-dimensional arrangement that can be recognised by aaRSs. In this pocket, residue A<sub>20</sub> of *E. coli* tRNA<sup>Arg</sup> specifies Arg identity in a major fashion in anchoring tRNA<sup>Arg</sup> on ArgRS, as demonstrated by crystallography (Stephen *et al.*, 2018). Note that the degenerated N<sub>20</sub> position (U/C) in the yeast tRNA<sup>Arg</sup> isoacceptors intervenes in the identity of the isoacceptor with ICG anticodon (McShane *et al.*, 2016). Similarly, residue G<sub>20</sub> is an identity element in yeast tRNA<sup>Phe</sup> for its specific charging by yeast



PheRS. Finally, in the serine and leucine systems aaRSs recognise the long variable arms of tRNA<sup>Ser</sup> and tRNA<sup>Leu</sup>. These arms are pivotal for the discrimination between cognate and noncognate tRNAs.

### **Computational predictions versus experimental findings**

Given the rather large panel of experimentally defined tRNA identity determinants (see Table 1) and the new computational and statistical analyses of big sequence data sets it is legitimate to compare predictions with experimental findings. Thus, global analyses of tRNA sequences, e.g. by Bayesian Network modeling (Branciamore *et al.*, 2018), revealed networks of positions in tRNA sequences associated with intrinsic structural properties. However, to find positions associated with focused structural or functional properties, sequence data sets should be hierarchised as advocated by Ardell and coworkers already 15 years ago. This was done following various strategies (e.g. phylogenies, isoaccepting tRNA families, RNA operational code) and allowed prediction of tRNA determinant and antideterminant positions.

Thus, analyses of the tRNA sequences from the main domains of life subdivided in families representing canonical amino acids, predicted for each family 3 groups of determinant positions with decreasing statistical likelihood of Relative Entropy values (Branciamore *et al.*, 2018). Many such positions, but not all, correspond to strong aminoacylation identity determinants biochemically characterised (see Giegé *et al.*, 1998); but the exact meaning of most positions remains elusive, indicating the informative complexity of the tRNA molecule (see Schimmel, 2018). In another approach, the most informative positions in archaeal tRNAs were searched, such as e.g. C<sub>73</sub> for His, G<sub>20</sub> for Phe and A<sub>20</sub> for Arg identities (Galili *et al.*, 2016); interestingly, these positions were found as positive aminoacylation determinants in *Aeropyrum pernix* tRNA<sup>His</sup> and tRNA<sup>Phe</sup> and in bacterial tRNA<sup>Arg</sup> (Table 1). In agreement, the crystal structure of *E. coli* ArgRS in complex with tRNA<sup>Arg</sup> shows contacts of A<sub>20</sub> with 6 amino acids of ArgRS, notably Asn84 which replacement by Ser leads to an important loss of catalytic efficiency (Stephen *et al.*, 2018). Finally, a computational approach unveiled identity clusters in tRNA isoacceptors that contain positions experimentally validated as tRNA aminoacylation determinants. These clusters are partly conserved in evolution and in particular show high diversity in *Archaea* (Zamudio *et al.*, 2019). In conclusion, computational sequence analysis yields a holistic view of tRNA properties and provide a robust background for future experimental work.

## **B. Identity in Atypical tRNA Aminoacylation Systems**

Among the first documented atypical tRNA aminoacylation systems one finds systems encompassing viral tRNA-like structures and mitochondrial tRNAs, both of unusual architecture. Another atypical system present in most *Archaea* and in a few *Bacteria* consists in canonical tRNA<sup>Lys</sup> and noncanonical class I LysRS. However, many eukaryal and archaeal organisms lack one or more members of the aaRS family. In these organisms the proper aminoacyl-tRNAs are formed by the pretranslational conversion of a noncognate amino acid attached to tRNA by enzymes unrelated to aaRSs. This implies deliberate tRNA misacylation by noncognate aaRSs referred as non-discriminating (ND). Misacylation is also used to synthesise selenocysteinyl-tRNA<sup>Sec</sup> in methanogenic organisms. Finally, pyrrolysine incorporation in a few archaeal proteins uses a specialised aaRS named PylRS.

### **Identity of viral tRNA mimics**

tRNA-like domains are found at the 3'-end of genomic RNAs of several genera of plant viral RNAs (see Dreher, 2010). Although these structures are aminoacylated by aaRSs, they do not

participate in protein synthesis but act in amplification of viral genomes. Three groups of tRNA-like domains were found on the basis of their aminoacylation by ValRS, HisRS and TyrRS. These domains deviate from the canonical tRNA cloverleaf and show closest sequence similarity with tRNA for the valine accepting structures from tymo- and furoviruses, including *turnip yellow mosaic virus* (TYMV). All viral tRNA mimics present pseudoknotted acceptor stems, which dictate their intricate three-dimensional architectures and a CCA-end, which is the unifying feature needed for aminoacylation.

Identity of tRNA-like structures follows the universal identity rules of tRNAs and include mimics from the canonical Val, Tyr and His identity sets, notably for:

1. Val identity of TYMV and related tRNA-like structures: discriminator base  $A_{73}$  and three nts from the tRNA<sup>Val</sup> anticodon loop, that is,  $A_{35}$  and  $C_{36}$  from the Val anticodon and  $C_{38}$ .
2. Tyr identity of bromoviral tRNA-like structures: discriminator base  $A_{73}$ , the  $C_1 \bullet G_{72}$  pair in the acceptor helix, as in *Eukaryota*, and  $U_{35}$  and  $A_{36}$  from the Tyr anticodon.
3. His identity of tobamoviral tRNA-like structures: atypical  $N_{-1}-N_{73}$  pair and  $G_{34}$  and  $U_{35}$  from the His anticodon.

Interestingly, the pseudoknotted acceptor stem of all tRNA-like structures contains a mimic of  $N_{-1}$  characteristic of tRNA<sup>His</sup>, so that an atypical  $N_{-1}-N_{73}$  pair can form, meaning that tRNA-like structures with Val and Tyr identity possess also His identity and are substrates of HisRS. Furthermore, as feasible with canonical tRNAs, the identity of tRNA-like structures can be engineered. Thus, replacement of the Val by Met determinants confers Met identity to the tRNA-like structure from TYMV (Dreher *et al.*, 1996).

### **Identity of mitochondrial tRNAs**

Mitochondrial tRNAs (mt-tRNA) show remarkable structural diversity with bias in nt-composition in favour of A, U and C and frequent occurrence of mismatched base pairs or other defects. So-called RNA editing, by base deamination or base replacement, restores the functional sequences; interestingly such editing events are often associated with base methylations (Dixit *et al.*, 2019). Moreover, most mt-tRNAs show atypical and even resected cloverleaf folds. mt-tRNA<sup>Aia</sup> species from *Acaria* are extreme examples of structural simplification. They are deprived of both D- and T-arms and have only conserved a canonical 7 nt anticodon loop and a 3'- $A_{73}$ CCA terminus (see Salinas-Giegé *et al.*, 2015). However, most mt-aaRSs are of bacterial type and in many cases aminoacylate both mitochondrial and bacterial tRNAs, but surprisingly, bacterial aaRSs do not or only weakly aminoacylate mt-tRNAs. Experiments on the Asp system support the hypothesis that non-aminoacylation of mt-tRNAs by bacterial aaRSs is linked to the large sequence and structural relaxation of the organelle encoded tRNAs, itself a consequence of the high rate of mt-genome divergence due to rapid evolution in mitochondria (Fender *et al.*, 2012). These peculiarities raise questions about the structure of mt-tRNAs required in ribosome-dependent protein synthesis, the peculiarities of mt-aaRSs and the identity of mt-tRNAs. For functional necessity, mt-tRNAs must be L-shaped with an adequate distance between anticodon and terminal CCA. Because anticodon loops and discriminator  $N_{73}$  are conserved in mt-tRNAs, it can be anticipated that elements from these tRNA extremities belong to mt-identity sets. This is explicitly verified for the Asp and Tyr identities of mammalian mt-tRNA<sup>Asp</sup> (Fender *et al.*, 2012) and mt-tRNA<sup>Tyr</sup> (Bonfond *et al.*, 2005a). However, because of the atypical and often simplified frameworks sustaining the L-shape of mt-tRNAs it can also be anticipated that mt-identity sets present idiosyncrasies and are simplified. Four examples illustrate the point:

1. Bovine mt-SerRS recognises two aberrant mt-tRNA<sup>Ser</sup> with GCU and UGA anticodons, namely, a truncated isoacceptor with GCU anticodon lacking the entire D-arm and, importantly, both are deprived of the long variable arm that is the major Ser identity determinant in cytosolic tRNA<sup>Ser</sup>. Combining crystallographic with mutagenesis and docking data, it could be shown that mt-SerRS recognises the distinct shape of each tRNA<sup>Ser</sup> isoacceptor by indirect readout involving recognition of the backbone of the acceptor helix and alternative interactions with the tRNA core (Chimnaronk *et al.*, 2005).
2. To prevent mistranslation of the AGA codon as arginine in metazoan organelles (because of a reassigned genetic code), mt-tRNA<sup>Arg</sup> species require a set of identity elements distinct from those of cytoplasmic tRNA<sup>Arg</sup> species. Using an insect mt-ArgRS, it was shown that the anticodon bases G<sub>36</sub> and C<sub>35</sub> are essential for arginylation, but that the major A<sub>20</sub> cytoplasmic identity determinant located in the D-loop has been lost (Igloi and Leisinger, 2014).
3. Recent functional studies compared the His identity in *Eukaryota* and in mitochondria. Thus, in humans, efficient histidylolation of mt-tRNA<sup>His</sup> primarily relies on a C<sub>73</sub> discriminator base and on the N-terminal WHEP sequence [present in 3 aaRSs, i.e. TprRS, HisRS and EPRS (fused GluRS–ProRS)] of mt-HisRS (Lee *et al.*, 2019).
4. Similarly, it was shown that human mt-AlaRS does not recognise the quasi-universal G<sub>3</sub>–U<sub>70</sub> wobble pair of tRNA<sup>Ala</sup> but a translocated G<sub>5</sub>–U<sub>68</sub> pair (Zeng *et al.*, 2019).

#### **Identity of canonical tRNAs recognised by atypical aaRSs**

In most extant *Archaea*, *Bacteria* and in many organelles, AsnRS and GlnRS are absent. Thus, Asn-tRNA<sup>Asn</sup> and Gln-tRNA<sup>Gln</sup> are formed in a two-step process consisting first to aminoacylate tRNA<sup>Asn</sup> or tRNA<sup>Gln</sup> with ND-AspRS or ND-GluRS (or paralogous GluRS) that are mischarging aaRSs with relaxed tRNA specificity. The Glu or Asp moieties bound on tRNA<sup>Asn</sup> or tRNA<sup>Gln</sup> are then amidated by amidotransferases (Asp-AdT or Glu-AdT) in the presence of ATP and an amide donor to form the properly aminoacylated Gln-tRNA<sup>Gln</sup> or Asn-tRNA<sup>Asn</sup> species. The whole catalytic process takes place on ribonucleoprotein complexes that assemble ND-AspRS/tRNA<sup>Asn</sup>/Asp-AdT or ND-GluRS/tRNA<sup>Gln</sup>/Glu-AdT in structures called Asp or Glu transamidosomes (Blaise *et al.*, 2010; Ito and Yokoyama, 2010). Misacylation occurs because of the similarity of Asp and Asn identities and Glu and Gln identities, with both couples taking advantages of determinants within anticodon and discriminator N<sub>73</sub> positions (Figure 1). However, for strict mischarging GluRSs, anticodon and discriminator residues of tRNA<sup>Gln</sup> are not identity determinants; instead, the major determinants for Gln acceptance are the two first base pairs localised in the acceptor helix of tRNA<sup>Gln</sup> [see (Giegé and Springer, 2016) for details and Hendrickson data]. Altogether, structural idiosyncrasies within ND-aaRSs favour misacylation and, importantly, because mischarging occurs within transamidosome protects cells against false use of the mischarged tRNA on the ribosome. Recently, peculiar AlaRSs were added to the constitutively mischarging aaRSs since they alanylolate tRNA<sup>Thr</sup> species with a shifted G–U Ala determinant at position 4–69 (Sun *et al.*, 2016) (see below).

Atypical class I LysRSs are evolutionary and structurally unrelated with class II LysRSs although they recognise tRNA via similar interactions. However, differences exist, as revealed by the peculiar Lys identity set in *Borrelia burgdorferi* tRNA<sup>Lys</sup> that specifies lysylation by *B. burgdorferi* LysRS (Ambrogelly *et al.*, 2005a). Thus, anticodon bases U<sub>35</sub> and U<sub>36</sub> are determinants for both class I and class II LysRSs, but the strength of U<sub>36</sub> is more important for class I LysRS. In contrast, the discriminator base A<sub>73</sub> plays only a marginal role, but the closely located G<sub>2</sub>–U<sub>71</sub> pair in the acceptor stem is essential for lysylation by class I LysRS. Remarkably, this pair is an antideterminant that affects lysylation by *E. coli* class II LysRS (Table 2). Finally, the structural context of the acceptor stem is crucial, since a shift of the wild-type identity pair G<sub>2</sub>–U<sub>71</sub> to

another position in the acceptor stem has dramatic effects on *B. burgdorferi* LysRS lysylation efficiency. All these features reflect differences in the mode of interaction of the tRNA acceptor helix by the two LysRSs.

A canonical class I CysRS is either absent or dispensable in *Euryarchaeota*, so that cysteinyl-tRNA<sup>Cys</sup> formation occurs via an indirect pathway in these organisms. This pathway couples protein synthesis with cysteine production and consists of a two-step amino acid transformation on tRNA<sup>Cys</sup>. The tRNA is first aminoacylated with O-phosphoserine (Sep) by O-phosphoseryl-tRNA synthetase (SepRS) and the phosphoserine moiety is then transformed to tRNA-bound cysteine by Sep-tRNA:Cys-tRNA synthase (SepCysS). Crystal structures show a close relationship between SepRS and the  $\alpha$  subunit of class II PheRS and a similarity of SepCysS with the CsdB protein, which is a member of the pyridoxal 5'-phosphate-dependent cysteine desulfurases. SepRS differs from CysRS by recruiting the m<sup>1</sup>G<sub>37</sub> modification as an identity determinant for aminoacylation (Zhang *et al.*, 2008). Both SepRS and SepCysS bind the reaction intermediate Sep-tRNA<sup>Cys</sup>, thus promoting its conversion to Cys-tRNA<sup>Cys</sup> and preventing binding to elongation factor or infiltration into the ribosome. Formation of Cys-tRNA<sup>Cys</sup> proceeds in a multienzyme complex without release of the intermediate phosphoseryl-tRNA<sup>Cys</sup> (Liu *et al.*, 2014).

The *E. coli* YadB protein is a minimalist GluRS lacking the anticodon binding domain. This atypical aaRS is in fact a tRNA modification enzyme that glutamylates the Q<sub>34</sub> residue at the wobble position of the QUC anticodon of *E. coli* tRNA<sup>Asp</sup>. YadB proteins are conserved in *Bacteria* and were renamed glutamyl-Q-tRNA<sup>Asp</sup> synthetases. Likely, they recognise in the anticodon branch of tRNA<sup>Asp</sup> sequence features reminiscent to those in the acceptor stem of tRNA<sup>Glu</sup> (Blaise *et al.*, 2004).

#### **Identity of atypical tRNA<sup>Sec</sup> and tRNA<sup>Pyl</sup>**

Selenocysteine (Sec) and pyrrolysine (Pyl) are generally regarded as the 21st and 22nd genetically encoded amino acids. Although selenoproteins are widespread in nature but not ubiquitous, Pyl containing proteins are rare and only found in a restricted number of archaeal taxa, such as *Methanosarcina barkeri*. Although the cloverleaf folds of tRNA<sup>Sec</sup> and tRNA<sup>Pyl</sup> is atypical in terms of size of helical stems and loop sequences, these tRNAs adopt an L-shaped conformation compatible with functional binding on ribosomes. Briefly, selenocysteine is delivered to the ribosome by dedicated elongation factor SelB and inserted in selenoproteins in response to UGA stop codons located in specific mRNAs sequences deciphered by different mechanisms in *Bacteria* and *Eukaryota*. As to pyrrolysine, it is coded by UAG stop codons under the control of PYLIS elements by analogy to SECIS elements controlling translational expression of selenoproteins.

tRNA<sup>Sec</sup> species are misacylated by SerRSs to Ser-tRNA<sup>Sec</sup>, and then converted to Sec-tRNA<sup>Sec</sup> by a selenocysteine synthase (SelA). Discriminator base G<sub>73</sub> is essential for serylation of tRNA<sup>Sec</sup> by SerRS. In addition, the long extra arm of tRNA<sup>Sec</sup> contributes to identity, due to its specific orientation and length, as well as its other atypical secondary structure elements (9 base pair acceptor stem, 6 base pair D-stem and 4 base pair T-stem), but to a lesser extent. These remarkable features are observed in the crystal structure of human SerRS in complex with tRNA<sup>Sec</sup>. The long extra arm of tRNA<sup>Sec</sup> and the conformation of the base triple U<sub>20</sub>:G<sub>19</sub>•C<sub>56</sub> confer to tRNA<sup>Sec</sup> a distinct structure compared to the related tRNA<sup>Ser</sup> (Wang *et al.*, 2015).

In contrast to Sec, Pyl is directly esterified to tRNA<sup>Pyl</sup> by pyrrolysyl-tRNA synthetase (PylRS). The structure of PylRS is a close relative of the  $\beta$  subunit of class II PheRS. The secondary structures of bacterial and archaeal tRNA<sup>Pyl</sup> are well known and contain distinct sequence features, especially in the variable loop and anticodon stem. Discriminator base G<sub>73</sub> and the first base pair

(G<sub>1</sub>•C<sub>72</sub>) in the acceptor stem from tRNA<sup>Pyl</sup> are the major identity elements for PylRS, whereas the CUA anticodon of tRNA<sup>Pyl</sup> is not involved (see Tharp *et al.*, 2018).

### ***Beyond canonical tRNA recognition by aaRSs: Interactions with other RNA species***

In addition to their established role in tRNA charging, aaRSs are involved in a variety of cellular processes distinct from tRNA aminoacylation. Some do not concern RNA binding (not described here), others involve interactions with RNA species other than tRNA. The non-tRNA binding ability of aaRSs is quite well known [see (Giegé and Springer, 2016) for initial data from Grunberg-Manago/Springer/Romby and Lambowitz laboratories]. Early studies showed binding of several aaRSs with mRNAs or viral rRNAs. Recent RNA interactome analyses revealed that almost all aaRSs bind mRNAs suggesting that their role may be more general than initially expected (Levi and Arava, 2019). The association of aaRSs with mRNA may impact different stages of its expression, including transcription antitermination, RNA splicing, translation regulation and viral RNA replication. In the present context it is important to emphasise that aaRS:RNA associations were often found under the control of tRNA identity rules. Several examples illustrate the point:

1. Transcription attenuation is an abundant regulation mechanism in *Bacteria*. Following transcription initiation, RNA polymerases synthesize regulatory elements that fold into structures that terminate transcription before synthesis of the downstream coding region. An emblematic example is expression of the *E. coli* MetRS gene that is controlled by a leader RNA with a structure mimicking tRNA<sup>Met</sup>, its natural substrate. This tRNA-like structure when binding to MetRS enables a terminator stem to fold and thereby represses its expression. In this process, an excess of free MetRS triggers premature termination of its own transcription. A similar antitermination mechanism was proposed for *Bacillus subtilis* ThrRS [see (Giegé and Springer, 2016) for Grunberg-Manago data].
2. In the eukaryote *S. cerevisiae*, GlyRS is involved in transcription termination of mRNAs that share a tRNA-like structure in their 3' untranslated region (UTR). Upon binding to the 3'-UTR, GlyRS plays a direct role in promoting efficient termination of transcription (Johanson *et al.*, 2003). The process involves the tRNA binding site of GlyRS as shown by mutant P552F which simultaneously impacts the affinity for its cognate tRNA and mRNA 3'-UTR.
3. Translation can be regulated through binding of an aaRS to RNA elements that mimic the cognate tRNA anticodon region. The best-studied example is that of *E. coli* ThrRS. Here, the protein binds to a structure that resembles the tRNA<sup>Thr</sup> anticodon region located in the 5'-UTR of its own mRNA, and negatively regulates ThrRS translation by competing with ribosome binding. Crystal structures demonstrated that ThrRS binds its own mRNA in a similar manner as the tRNA anticodon region. Furthermore, recognition by ThrRS follows tRNA<sup>Thr</sup> identity rules and can be exchanged to MetRS by changing the anticodon-like triplet from Thr to Met. This study clearly established the concept of mRNA recognition through tRNA mimics and the interrelation between RNA binding and translation regulation [see (Giegé and Springer, 2016) for Romby/Springer data]. Recently, RNA immunoprecipitation experiments coupled to deep sequencing revealed that several *S. cerevisiae* aaRSs (MetRS, GluRS, ValRS, GlnRS and HisRS) bind their own mRNA indicating that the phenomenon of autogenous regulation that exploits binding mimicry to control mRNA translation is more common than previously supposed (Levi and Arava, 2019).
4. Two aaRSs are involved in the splicing of group I introns (*Neurospora crassa* TyrRS and *S. cerevisiae* LeuRS). Thus, *N. crassa* TyrRS is directly involved in the splicing of group I introns within the 25S ribosomal RNA through its binding to the intron. Although the *in vitro* binding of

tRNA<sup>Tyr</sup> to TyrRS inhibited its splicing activity, it was not clear whether the intron motif that binds TyrRS would resemble tRNA. In fact, the co-crystal structure of *N. crassa* TyrRS with a group I intron identified the binding site on one face of the intron, with little resemblance to tRNA-specific features and almost exclusive interactions with the backbone. In addition, the structure shows that the group I intron binds across the two subunits of the homodimeric protein with a phylum-specific and recently evolved RNA-binding surface distinct from that which binds tRNA<sup>Tyr</sup> [see (Giegé and Springer, 2016) for Lambowitz data]. On the other side, the splicing mechanism by yeast LeuRS shows remarkable differences with that of TyrRS. In particular, LeuRSs from diverse origins are capable of splicing, even though these species lack group I introns. This results from the fact that LeuRS binds introns primarily through its conserved editing domain that has functionally diverged to confer a robust splicing activity (Sarkar *et al.*, 2012). Thus, LeuRS differs from TyrRS which uses a newly evolved domain not involved in tRNA aminoacylation. These two examples show that mRNA intron recognition by aaRS may be achieved by repurposing tRNA-binding domains or through new domains added to the aaRS later in evolution.

5. Beside the nontranslational role of aaRSs that interact with the tRNA-like domains present at the 3-termini of several plant viral RNAs (see above), additional nontranslational functions of aaRSs include interactions with viruses during cell infection via IRES (Internal Ribosome Entry Site) elements. Thus, the poliovirus IRES uses tRNA<sup>Gly</sup> anticodon stem-loop mimicry to recruit GlyRS to the apical part of its RNA domain V (Andreev *et al.*, 2012). The binding of GlyRS stabilizes the IRES in a conformation that promotes its accommodation in the mRNA binding site of the ribosome, thereby promoting the 48S initiation complex formation and then the synthesis of viral proteins. In terms of specificity, the synthetase binding requires a Gly anticodon in the tRNA mimic showing that the process obeys the tRNA identity rules. Additional evidences showed that the regulation of translation initiation via GlyRS is a universal mechanism for this family of viruses.

### C. Overview on the Mechanism of tRNA Aminoacylation

Synthetases are classified into two groups of 10 enzymes each that differ by their catalytic site and aminoacylation on the 2' or 3' OH of the terminal ribose of tRNA. The two classes are also characterised by sequence homologies of short peptides, called class-defining motifs, located in crucial parts of the catalytic domains. The catalytic domain of class I aaRSs comprises a Rossmann fold and that of class II enzymes is built around an antiparallel  $\beta$  sheet. In class I aaRSs, the conserved peptidic motifs HIGH and KMSKS interact with ATP and assist catalysis, whereas in class II enzymes, residues of the so-called motifs 2 and 3 play this role. The third class II-defining motif, motif 1, is responsible for subunit associations of  $\alpha_2$  dimers. The class II catalytic domain with a large antiparallel central sheet (approximately 250 residues) is larger than that of the Rossmann fold (approximately 150 residues) common to class I aaRSs. This difference is explained by the presence in the catalytic domain of class II aaRSs of the binding site of the CCA-end of tRNA, a role performed by a distinct domain in class I enzymes. For ATP binding, two class-specific binding modes are observed (Figure 2). Despite these differences, all aaRSs catalyse the same global type of reaction, using an activated aminoacyl-adenylate intermediate from which the amino acid is transferred to the tRNA, forming the 'charged' tRNA and releasing AMP. The result is that the high energy of the anhydride-phosphate bond of ATP (7 kcal) is transferred to the ester bond between the tRNA and the amino acid, via the acyl-phosphate bond of the aminoacyl-adenylate (high energy because 2' and 3' OH groups are in

*cis*). The crucial features in aaRSs that are important for tRNA aminoacylation are summarised in Table 3.

### **Common mechanism but two ways to approach ATP and tRNA**

A mechanism accounted for by crystallographic data on GlnRSs and AspRSs, the two best-known class I and class II aaRSs, is common to all aaRSs. In the first step of the tRNA aminoacylation reaction, that is, amino acid activation, the amino acid and ATP enter into their respective binding pockets, with ATP in its reactive conformation characteristic of each aaRS class (Figure 2). Then, the carboxylate group of the amino acid substrate attacks the  $\alpha$ -phosphate of ATP, leading to a penta-coordinated intermediate stabilised by conserved residues in critical peptidic motifs of aaRSs. Hydrolysis of the phosphate bond followed by release of the pyrophosphate group is activated by  $Mg^{2+}$  ion(s) which withdraw electrons with the assistance of chemical groups on the enzyme. In the second step, the 2' or 3' OH (for class I or class II systems, respectively) of the terminal tRNA ribose attacks the carbon of the mixed anhydride intermediate of the adenylate, forming a tetrahedral second transition state. Then one of the two free oxygen atoms of the  $\beta$ -phosphate attracts the proton from the attacking-OH of the terminal ribose, forming a cyclic intermediate that converts into the ester linkage between the amino acid and the ribose.

Crystal structures reveal aaRS class-dependent tRNA recognition modes (Figure 2). Thus, the acceptor stem of tRNA is recognised by class I aaRSs by its minor groove and by its major groove in the case of class II enzymes. As a consequence of this symmetric binding, the variable region of tRNA is exposed to the solvent in complexes with class I aaRSs, while it faces the protein in class II enzymes. Another consequence is a preferential distribution of identity elements on tRNAs. For instance, the major determinant in tRNA<sup>Arg</sup> (i.e. A<sub>20</sub>) recognised by class I ArgRS is found in the D-loop, whereas the long variable arm of tRNA<sup>Ser</sup> recognised by class II SerRS is on the opposite side of tRNA.

Taken together, the juxtaposition of the reactive groups of the amino acid, ATP and tRNA substrates trigger the reaction and the structure of the intermediates promotes by itself its forward progression. The key functional difference between the two aaRS classes is the mode of entry of the helical acceptor stem of tRNA into the active site that dictates the positioning of the acceptor-OH of terminal ribose. The conservation of the 2' or 3' specificity through evolution remains obscure since after release from the enzyme a rapid isomerisation of the amino acid occurs between the 2' and 3' OHs. This isomerisation step is required to produce the 3' species that are used during protein synthesis and is facilitated when elongation factor interacts with charged tRNA. In summary, as usual for catalysts, the main function of the aaRSs is to correctly orient the substrates and to stabilise the intermediate transition states.

### ***Kinetic aspects of identity expression and the role of tRNA in aminoacylation***

Within identity sets, all determinants do not play the same role, as reflected by their strength exerted on tRNA aminoacylation. Thus, losses of aminoacylation efficiency ranging from less than 10 to more than 1000 have been found. For instance, mutating anticodon residue U<sub>35</sub> to A in yeast tRNA<sup>Asp</sup> impairs aminoacylation efficiency 530-fold, while changing the wobble pair G<sub>10</sub>-U<sub>25</sub> to A<sub>10</sub>•U<sub>25</sub> leads only to a 33-fold loss (Pütz *et al.*, 1991). In case of multiple mutations effects can be additive, but often they are cooperative (cumulative effect greater than the sum of effects brought by individual mutations) and sometimes anticooperative (cumulative effect smaller than the sum of individual effects) (Pütz *et al.*, 1993).

A variety of structural features within tRNA play important roles in tRNA aminoacylation (Table 4). Examination of crystal structures shows that the strongest identity elements at distal ends of the tRNA are in direct contact with amino acids from the synthetases. These contacts, in particular with anticodon residues, are facilitated by the huge conformational change tRNA often undergoes when it recognises aaRSs. Such contacts may be lost upon mutation of the determinants. Other determinants, however, are not in direct contact with the enzyme and act indirectly. This is the case of the G<sub>10</sub>-U<sub>25</sub> pair in yeast tRNA<sup>Asp</sup> that participates in a tertiary interaction with G<sub>45</sub>. Interestingly, the loss of activity of the A<sub>10</sub> mutant is mainly due to  $K_m$  while that of mutants impairing contacts with the catalytic domain or the anticodon-binding domain affect predominantly  $k_{cat}$  (Pütz *et al.*, 1991). Likewise, reorganising the D-loop and variable region in *E. coli* tRNA<sup>Leu</sup> and changing its G<sub>15</sub>-C<sub>48</sub> Levitt pair to A<sub>15</sub>-U<sub>48</sub> impairs Leu identity (Tocchini-Valentini *et al.*, 2000). Such functional effects triggered by structure alterations in tRNA, as observed in the Asp and Leu systems, are of general occurrence in aminoacylation systems.

Concluding, identity expression of tRNAs are intricate molecular processes that go far away from the simple view of a static recognition of determinants by synthetases (see Li *et al.*, 2015). Kinetic effects are essential as well as indirect readout (Perona and Hou, 2007) and other indirect structural effects of determinants that tune the conformation of tRNA, as can be the case of modified nucleosides or base pairs in anticodon and acceptor stems.

## D. Allosteric Phenomena in tRNA-Synthetase Systems

Efficiency of tRNA aminoacylation is dependent on tRNA shape and full sets of identity determinants, the relative contribution of each being system-dependent (even taxa-dependent in case of weak determinants). In addition, structural plasticity of aaRSs (Crnkovic *et al.*, 2019) and their tRNA ligands (Chan *et al.*, 2020) is a critical necessity to ensure specificity because functional aaRS complexes must process the chemical information brought by the interaction of the small and macromolecular substrates. Contacts with substrates occur necessarily in the catalytic site and also in distant domains, located at approximately 50 to 70 Å apart from the catalytic site. On the other hand, the fact that aaRSs are multi-domain proteins implies existence of communication between domains and of coupled domain motions. Together, induced fit/allosteric phenomena are anticipated and indeed were found.

### Early highlights on asymmetry and using allostery to enforce specificity

First evidence of allostery in aaRSs came from kinetic analyses that reported an inability of the active site of some dimeric aaRSs to perform the amino acid activation step simultaneously. This was later supported by structural data that pinpointed differences in the occupancy of active sites by the small substrates. This functional asymmetry was first demonstrated in *Bacillus stearothermophilus* TyrRS where the dimer exhibits half-of-the-sites reactivity, binds one tyrosine and forms one tyrosyl-adenylate intermediate (Fersht *et al.*, 1975). Likewise, the stoichiometry of tRNA binding can also show asymmetry, as found for example with dimeric class II ProRS from *Methanocaldococcus jannashii* (Ambrogelly *et al.*, 2005b). Therefore, asymmetry, and hence allostery, was considered to be a general feature of aminoacylation reactions catalysed by dimeric aaRSs from both classes.

More generally, allostery allows aaRSs to reach specificity with only moderate affinity for their cognate amino acid ATP and tRNA substrates (tight binding would make product release more difficult and decrease turnover) (see Li *et al.*, 2015). In fact, the critical goal for aaRSs is to



provide the highest affinity to the adenylate intermediate which has to remain bound to the enzyme since its premature release would lead to its hydrolysis in solution and prevent subsequent transfer to the acceptor end of tRNA. Given such functional necessity, allostery occurs through conformational changes or stabilisation of previously flexible regions of the substrate binding sites. Structural studies have revealed a broad spectrum of substrate binding modes, ranging from the rigid lock and key recognition to various types of induced-fit recognition. Examples of induced fit have been described in the active site of AspRS where a 'flipping loop', disordered in absence of amino acid, becomes visible and interacts with aspartic acid in the presence of tRNA. Similarly, proline- and ATP-binding on ProRS cause conformational changes in the proline and ATP-binding loops. Subsequent prolyl-adenylate formation is also a prerequisite for the final conformational ordering of a loop of 10 residues essential for functional tRNA binding. More widely, local induced-fit changes coupled to global inter-domain and inter-subunits movements, have important effects on the aminoacylation catalysis and generate the functional asymmetry. Overall, the induced-fit mode of substrate binding is probably sufficiently accurate for some aaRSs to avoid the need of editing activity against misactivated amino acids.

### **Allostery for communication between synthetase domains and cooperativity between identity determinants**

Optimal aminoacylation efficiency is dependent both on tRNA shape and full sets of identity determinants in tRNA, the relative contribution of each being system dependent. Cooperative, additive or anticooperative effects between determinants have been observed in the Asp system, with pairs of determinants located far apart in the three-dimensional structure of tRNA<sup>Asp</sup> acting cooperatively, whereas those clustered in the anticodon triplet act additively or anticooperatively (Pütz *et al.*, 1993). Similar effects occur in other systems. A structural interpretation is not straightforward and implies transfer of chemical information from remote aaRS and/or tRNA regions to the catalytic site of aaRSs. Such intra- or intermolecular communications are accompanied by induced fit/allosteric phenomena and by coupled domain motions in aaRSs. In other words, functional aaRS:tRNA complexes can be considered as 'signal transduction' systems in which specific conformational changes occur, which can be subtle or dramatic.

Presently, based on crystallography, advanced enzymology and molecular dynamic simulations (see Li *et al.*, 2015), allosteric phenomena and communication paths (Table 3) could be documented for 14 aaRS families. Subtle conformational changes in aaRSs essential for tRNA aminoacylation are well documented in the Asp and Gln systems. Binding of tRNA<sup>Asp</sup> to the anticodon-binding domain of AspRS stabilizes the flipping loop that controls the proper positioning of aspartate in the active site. Interestingly, all regions of AspRS that contact tRNA<sup>Asp</sup> show local flexibility in the apo-enzyme, which suggests that analysing the flexibility of apo-aaRSs may inform about the communication pathways of these enzymes (see Giegé and Springer, 2016). In the case of the *E. coli* GlnRS:tRNA<sup>Gln</sup> complex, pre-steady-state kinetics on GlnRS mutants were employed to discover allosteric signaling pathways in the aaRS body that would regulate glutamine binding and Gln-tRNA formation. Long-range signal propagation from the tRNA anticodon to the catalytic site reveals protein contacts that weaken glutamine binding affinity across distances up to 40 Å (Rodriguez-Hernandez and Perona, 2011). In TrpRS, ATP acts as an allosteric effector and domain motion during induced-fit closes and twists the anticodon-binding and catalytic domains. Moreover, the anticodon-binding domain moves as a rigid body

with both catalytic signature sequences to deliver ATP to the tryptophan carboxyl group in the active site within the Rossmann fold (Laowanapiban *et al.*, 2009).

## E. Evolution of tRNA-Synthetase Systems

Positive identity determinants specifying tRNA aminoacylation (see Table 1) are of two types: i.e. strong determinants that are the major identity signals and weak determinants that fine-tune identity. Given that strong determinants are located quasi-exclusively at the two distal extremities of tRNA, it was tempting to engineer artificial tRNAs restricted to these extremities connected together by linkers. Provided the strong determinants contact recognition sites on aaRSs, it was conjectured that such RNAs are functional and this was truly verified with artificial tRNA<sup>Asp</sup> molecules that can be aminoacylated by yeast AspRS (Wolfson *et al.*, 1999). This suggested a functional plasticity of identity rules in modern systems and agrees with the present knowledge of the evolution of canonical and atypical tRNA-synthetase systems.

### Origin, ancestry, doubtful coevolutions, and idiosyncrasies

Understanding the complexity of tRNA aminoacylation systems is a central question in evolution and goes far beyond tRNA aminoacylation and protein synthesis (Schimmel, 2018). This complexity emerged in the early history of life when nature established the basic mechanisms of replication, translation and metabolisms. It is often believed that this occurred in the RNA World, but this remains open. When primordial ribosome-dependent protein synthesis emerged, ancestral aminoacylation systems were simplified with a reduced number of proto-tRNAs and aaRSs of minimalist size. The operational RNA code for amino acids keeps memory of this evolution (Schimmel *et al.*, 1993). Yet, the origin and evolution of the two aaRS classes and the associated recognition of the tRNA acceptor stem in its minor- or major-groove by class I or class II aaRSs remain puzzling. The recent finding of a hierarchical groove discrimination by aaRSs and of thermodynamic signatures in acceptor stems seen by modern aaRSs shed new light in the field (Carter and Wills, 2019).

The relationship between tRNA aminoacylation and genetic code remains a conundrum, although recent data uncover new perspectives (e.g. Saint-Léger *et al.*, 2016; Kim *et al.*, 2019). With the increasing amount of phylogenetic data on tRNAs and aaRSs, the thought that tRNA aminoacylation and genetic code coevolved by similar pathways is gradually broken. Data supporting this view, for long were exceptions, today they are becoming the consensus view. The first example of an atypical tRNA identity rule was found in the His system by automated statistical methods (see Table 1). Another example is the recognition of canonical tRNAs by atypical archaeal or bacterial (including mitochondrial) ND-AspRSs or ND-GluRSs (see above). In this context it was shown that extant eukaryal GlnRSs derive from an ancient eukaryal GlnRS that evolved by gene duplication from a ND-GluRS and that eukaryal GluRSs and GlnRSs acquired additional domains when tRNA<sup>Glu</sup> and tRNA<sup>Gln</sup> differentiated their identity determinants (Hadd and Perona, 2014). Likewise, in *Bacteria* and *Archaea*, the Asp and Asn systems evolved peculiar strategies for Asp-tRNA<sup>Asp</sup> and Asn-tRNA<sup>Asn</sup> production (Chuawong *et al.*, 2020). More extreme, in the Tyr system evolution differentiated identity rules in the three domains of life. Thus, the N<sub>1</sub>•N<sub>72</sub> identity pair has lost the capacity to discriminate between the G<sub>1</sub>•C<sub>72</sub> pair, typical of bacterial and mitochondrial tRNA<sup>Tyr</sup>, and the reverse C<sub>1</sub>•G<sub>72</sub> pair, present in archaeal and eukaryal tRNA<sup>Tyr</sup>, so that identity primarily relies on the discriminator base A<sub>73</sub> and on phylum-specific binding modes with TyrRS (see Bonnefond *et al.*, 2005b). Moreover, recent phylogenetic data on the distribution of the Trp and Tyr systems in the Tree of Life uncovered in certain lineages of *Archaea*, *Bacteria*, *Eukaryota* and giant viruses, unprecedented relationships

between the class Ic TrpRSs and TyrRSs distinct from those of their cognate tRNAs (Mukai *et al.* 2017).

The mitochondrial and apicoplasmic tRNA aminoacylation systems deserve particular attention. For instance, in mammals, the mt-Asp identity evolved to a reduced identity set with the sole U<sub>35</sub> and C<sub>36</sub> anticodon determinants, a fact explained by structural and biophysical properties of the human mitochondrial tRNA<sup>Asp</sup>-AspRS system. Even though the three-dimensional structure of mt-AspRS is close to that of its *E. coli* homologue, it differentiates by an enlarged catalytic groove, a more electropositive surface and a reduced thermal stability. Moreover, isothermal titration calorimetry shows an affinity of mt-AspRS for cognate tRNA one order of magnitude higher than for noncognate tRNAs, but with different enthalpic and entropic contributions. Altogether, this illustrates an evolutionary mechanism for adaptation of nuclear-encoded aaRSs to degenerate mt-encoded tRNAs (Neuenfeldt *et al.*, 2013). In apicoplasts, identity rules are likely also simplified as found in the Tyr system. Here, charging of api-tRNA<sup>Tyr</sup> by cognate *Plasmodium falciparum* api-TyrRS is promoted by only three weak positive identity elements in the tRNA and likely relies on negative recognition determinants and on the idiosyncratic sequence insertions in api-aaRSs, notably in TyrRS (Cela *et al.*, 2018).

Other idiosyncrasies in tRNA-synthetase systems correlated with identity are linked to tax-specific sequence variabilities in both canonical and atypical systems. Peculiarities essentially rely on the distribution of weak identity determinants in the core of the tRNA structure (see Table 1) and on divergent aaRS structures. Thus, despite mt-tRNA<sup>His</sup> species are lacking the universal G<sub>-1</sub> His identity determinant, they are histidylated because of the presence of distinct HisRS isoforms differing only by their N-termini code for His identity in *Eukaryota*. This suggests a new balance of minor His determinants overcoming the absence of G<sub>-1</sub> for HisRS recognition, with e.g. yeast HisRS recognizing both A<sub>73</sub> and C<sub>73</sub> minor determinants while mt-HisRS prefers A<sub>73</sub> (Lee *et al.*, 2019).

Taken together, these data prove a functional plasticity of tRNA identity rules in modern systems that, in extreme cases, is correlated with a huge structural plasticity of tRNA molecules. These conclusions are best illustrated by the unexpected property of human AlaRS to mischarge tRNAs (i.e. tRNA<sup>Thr(AGU)</sup> and tRNA<sup>Cys(GCA)</sup> isoacceptors, well represented in vertebrates) with the strong G<sub>3</sub>-U<sub>70</sub> Ala identity pair shifted in the acceptor stem to position 4-69 (Sun *et al.*, 2016) and by the recent discovery of a miniaturized mt-tRNA, i.e. a nematode tRNA<sup>Arg</sup> that keeps a boomerang-like L-shape despite a length of only 45 nts (Jühling *et al.*, 2018). Note that in all mt-systems so far examined, the canonical structures of the tRNA 3'-end and anticodon loop is conserved, thereby allowing correct reading of the genetic code.

## Errors and error prevention

Inaccuracy of amino acid and tRNA selection by aaRSs explain that tRNA-synthetase systems are not fully specific and that errors are tolerated up to thresholds that are system-dependent. It can be conjectured that false tRNA charging, by increasing the diversity of proteomes, is a motor of evolution. Because too high error levels could threaten cellular life, nature has evolved prevention and correction mechanisms. However, under certain circumstances high error levels lead to dysfunctions (see below).

Accurate tRNA aminoacylation depends on the successful discrimination between cognate and noncognate or nonstandard amino acid and tRNA substrates. Inaccuracy in amino acid selection ( $10^{-4}$  to  $10^{-5}$ ) by synthetases is more frequent than tRNA selection ( $10^{-6}$ ) due to the larger surface area of the tRNA molecules and greater structural diversity that serve as discriminating factors. It has been postulated that errors in protein synthesis should not exceed 1/3000 *in vivo*.

About half of the aaRSs misactivate noncognate amino acids (proteinogenic or nonproteinogenic) that are similar to their cognate substrates. Likewise, aaRSs can recognise and aminoacylate noncognate tRNAs. To prevent detrimental effects of mischarging, evolution has worked out editing/proofreading mechanisms to clear mistakes before wrong amino acids would be misincorporated into proteins. Kinetic proofreading based on high levels of ATP hydrolysis, as proposed in the seventies by Hopfield to clear misactivated amino acids, has been refreshed and, among others, it was shown that energy dissipation accounts for tRNA selection in protein synthesis (Banerjee *et al.*, 2020). Editing is mediated by 9 aaRSs covering both aaRS classes (Table 3). It is ubiquitous in all domains of life, but not in mitochondria. Interestingly, an editing activity of human mt-AlaRS clears mischarged mt-tRNA<sup>Ala</sup> and is likely a biological necessity since its loss causes embryo lethality in mice (Hilander *et al.*, 2018). Editing mechanisms are globally well understood, although some aspects remain elusive. Editing was first formalized by Fersht in the seventies and presently is supported by a huge panel of enzymology, sequence and crystallographic data. Thus, pretransfer editing (first sieve) clears amino acid misactivated in the catalytic site, posttransfer editing (second sieve) edits mischarged tRNAs with the assistance of distinct editing domains, and if needed *trans*-editing by freestanding proteins complete editing (third sieve) (see Yadavalli and Ibba, 2012). Editing domains distinct from the synthetic aminoacylation domains are conserved in 5 cytosolic aaRSs (Ile, Val, Leu, Phe and AlaRS), notably the connective peptide CP1 (Ile, Val, LeuRS), the B3/B4 domain in the  $\beta$  subunit of PheRS and the C-terminal domain of AlaRS. In 2 aaRSs lacking distinct editing domains (MetRS and LysRS), editing occurs in the aminoacylation synthetic site. The situation is contrasted for AlaRS, ProRS and ThrRS that edit their mischarged tRNAs by serendipitous and interconnected routes. These aaRSs have distinct editing domains not strictly conserved in evolution, and in addition they often use *trans*-editing factors (AlaXps, ProXps, YbaKps, ThrRS-ed and/or d-aminoacyl-tRNA-deacylases abbreviated DTD). Note that in the 20 canonical tRNA–aaRS systems, mistranslations are minimized because of the generally very low yield of mischarged tRNA production due to a balance between deacylation (enzymatic and nonenzymatic) and acylation rates favouring correct aminoacylation (see Giegé *et al.*, 1993). In case of constitutively ND-aaRSs (Ala, Asp, Glu) with relaxed aminoacylation specificity, potential toxicity of mischarged tRNAs is threatened by dedicated mechanisms (see Blaise *et al.*, 2010; Ito and Yokoyama, 2010; Kuncha *et al.*, 2019).

A few highlights and recent findings deserve attention:

1. The controversial mechanistic of editing by Ile-, Val- and LeuRS, i.e. tRNA-independent or tRNA-dependent pretransfer editing, has been clarified. In the case of *E. coli* IleRS, tRNA-dependent pretransfer editing accounts for one-third of the total proofreading and employs a conserved tyrosine determinant within the synthetic site for both editing and aminoacylation. This dual process is kinetically controlled, and in *E. coli* LeuRS it relies almost entirely on posttransfer tRNA-dependent editing in which the A<sub>76</sub> 3'-OH group of tRNA<sup>Leu</sup> plays a crucial role in posttransfer editing (see Dulic *et al.*, 2018).
2. Posttransfer editing cleaves mischarged tRNAs in domains that are distinct from the synthetic aminoacylation domains. This necessitates conformational changes in aaRSs and tRNAs (i.e. a significant rotation of the CP1 domain inserted about 35 Å from the synthetic aminoacylation site and thus a translocation of the 3'-end of the mischarged tRNA to this site, as seen e.g. in LeuRSs). Further, it implies partially distinct tRNA interactions in editing and synthetic sites, as reflected by a clear segregation of tRNA determinants for editing (in the corner of the L-shaped tRNA) and aminoacylation (in the anticodon) by IleRS. Similarly, the anticodon arm of tRNA<sup>Leu</sup> is essential for editing by LeuRS but dispensable for aminoacylation. In an insightful study, the functionality of CP1 domains was examined by interspecies domain

exchange. Thus, in human cytoplasmic LeuRS, lack of the CP1 domain leads to a complete loss of tRNA binding and catalytic activities (amino acid activation and transfer) and only the CP1 domain from yeast can partially rescue the LeuRS functions. This demonstrates the importance of context-dependent and structurally coordinated cross talks between CP1 and synthetic domains (Huang *et al.*, 2014).

3. Editing of AlaRS, ProRS and ThrRS follows serendipitous and interconnected routes. These aaRSs have distinct editing domains not strictly conserved in evolution, and in addition they often use *trans*-editing factors (AlaXps, ProXps, YbaKps, ThrRS-eds, *d*-aminoacyl-tRNA-deacylases (DTD). Thus, ThrRS-eds are ThrRSs with a resected catalytic domain that are just present in a few *Archaea*. Also, in many archaeal ThrRSs the N-terminal editing domain is replaced by a module homologous to the DTD fold. Interestingly, ProXp factors with relaxed specificity recognizing multiple mischarged tRNAs, even with nonproteinogenic amino acids, were recently discovered (Bacuso *et al.*, 2018).

4. The DTD proteins are present under different types all over the Tree of Life and are major actors in proofreading and quality control processes. DTDs clear tRNAs mischarged with *d* amino acids and control hydrolysis of tRNA bound achiral glycine (i.e. charged on tRNA<sup>Gly</sup> and mischarged on tRNA<sup>Ala</sup>). One type of DTD is likely the progenitor of ThrRS in *Archaea*. Remarkably, the Ala identity pair G<sub>3</sub>–U<sub>70</sub> was found to be a universal determinant for DTD, thus explaining its functional relationship with AlaRS. In particular, bacterial DTD clears efficiently noncognate Gly-tRNA<sup>Ala</sup> but much less cognate Gly-tRNA<sup>Gly</sup> due to their discriminator base U<sub>73</sub> that acts as an antideterminant. In addition, elongation factor Tu further protects Gly-tRNA<sup>Gly</sup> from deacylation. Moreover, in higher *Eukaryota* (i.e. *Animalia*), a paralog of DTD, named ADT, exhibits a relaxed specificity and proofreads tRNA<sup>Thr</sup> isoacceptors with a shifted Ala identity pair U<sub>4</sub>–G<sub>69</sub> that were alanylated by ND-AlaRSs. In fact, ADT turned out to be a glycine deacylase that hydrolyses mischarged Gly-tRNA<sup>Ala</sup> in *Bacteria* and *Eukaryota* (Kuncha *et al.*, 2019).

5. The origin of the prevalence of *l* amino acid homochirality in extant protein synthesis remains elusive. In this context, a novel function of the AlaRS editing domain is insightful. Thus, *Thermus thermophilus* AlaRS forms at high level *d* Ala-tRNA<sup>Ala</sup> that, surprisingly, is not edited by DTD but progressively deacylated via posttransfer-editing in the synthetic site of AlaRS. Thereby this indicates an active role of the editing domain of AlaRS in chirality control (Ryback *et al.*, 2019).

## Quality control and error biology in tRNA–synthetase systems

The erroneous genetic code interpretation induced by misacylated tRNAs may be a real advantage under stress conditions. For instance, in fungi of the *Candida* clade, ambiguous decoding of the Leu codon is increasing phenotype diversity (Bezerra *et al.*, 2013). Also, nutrient starvation and viral infection reveal that, up to a certain threshold, tRNA mischarging may favour the formation of a statistic proteome offering real advantages for the cell survival and adaptation. Statistical synthesis may provide phenotypic or functional plasticity that can be subsequently refined by codon reassignment and genetic code evolution.

### **Expanding proteomes through tRNA mischarging**

For a long time, high level of accuracy during protein synthesis was considered essential for life since tRNA mischarging would result in genetic code ambiguity and yield statistical proteins likely harmful for cellular functions. Indeed, severe mistranslation leads to growth inhibition, mitochondrial dysfunction and apoptosis (Hilander *et al.*, 2018; Schimmel, 2018). However, different types of cells tolerate various levels of mistranslation, and mistranslation

appears to be advantageous under certain physiological conditions, leading to a diversification of proteomes (Schwartz and Pan, 2017). Thus, some aaRSs have altered or lost editing domains and produce misacylated tRNAs *in vivo*. This occurs in *Mycoplasma* parasites, where several aaRSs with naturally perturbed editing function (LeuRS, PheRS and ThrRS) catalyse high levels of tRNA mischarging resulting in a statistical proteome (Li *et al.*, 2011). The loss of editing capacity may benefit *Mycoplasma* to escape host immune responses by increasing the antigen diversity. Similarly, *E. coli* cells not only tolerate the presence of misacylated tRNAs but can even require it for growth under selective pressure. Quite surprising, the cells survive with up to 10% of mismatched protein, suggesting that the editing function of aaRSs is not essential for survival under certain circumstances (Ruan *et al.*, 2006).

### **Biology of dysfunctions**

Dysfunctions of tRNA–synthetase systems are frequent in human diseases and soon attracted interest. Several examples illustrate the point:

1. Idiopathic inflammatory myopathy (IIM) is an uncommon type of human autoimmune disease studied for several decades. Patients with IMM exhibit a spectrum of pulmonary, arthritic and/ or inflammatory muscle signs. A subset of IIM is referred to as the “anti-synthetase syndrome.” In this case, autoantibodies against AlaRS, AsnRS, GlyRS, HisRS, IleRS, ThrRS have been found, with highly neutralizing and aaRS-specific properties that inhibit aminoacylation and critically reduce protein synthesis (see Mahler *et al.*, 2014).
2. A mutation in the editing domain of human AlaRS was shown to cause cerebellar Purkinje cell loss and ataxia because of a compromised proofreading activity leading to tRNA mischarging and accumulation of misfolded proteins in neurons (Lee *et al.*, 2006).
3. More recently, the precise mechanism underlying the toxicity of the plant non-proteinogenic amino acid azetidine, a mimic of proline and alanine, was elucidated. This amino acid, sometimes present in the food chain, is activated by human ProRS and AlaRS and misincorporated into proteins in place of proline only, what has pathological consequences (Song *et al.*, 2017).

Many other studies report aaRSs involved in cancer cell survival and tumor progression. Low expression of TrpRS in colorectal cancer correlates with increased risk of metastasis and poor prognosis, LysRS is involved in the development of melanoma, and GlnRS blocks the proapoptotic pathway of an enzyme that controls tumorigenesis and stress responses and inhibits apoptosis. The bifunctional GluProRS acts on Vascular Endothelial Growth Factor A (VEGF-A) and promotes tumor growth. MetRS was found to be overexpressed in non-small cell lung cancer. Although these aaRSs have clear roles in tumorigenesis, the molecular mechanism underlying their role in cancer remains unclear (see Kwon *et al.*, 2019).

Today the number of diseases linked to dysfunction of cytosolic and mitochondrial tRNA–synthetase systems is increasing. They cover a large panel of pathologies, often neuropathies and cardiopathies with a wealth of clinical manifestations, and their patho-mechanisms are worldwide investigated (see Tahmasebi *et al.*, 2018). Understanding such dysfunctions is a prerequisite for the discovery of disease-focused therapeutics and, in a more basic perspective, gives clues for engineering tRNA–aminoacylation systems, in view either to expand the genetic code or to prepare alloproteins containing unnatural amino acids.

## Engineering tRNA aminoacylation

Engineering of tRNA and synthetase became recently routine practice, both for structural and functional studies. Efficient transplantation of an identity set implies that the architectural framework of the donor and acceptor tRNAs are sufficiently similar to allow correct positioning of the identity bases for optimal recognition by amino acids on the aaRSs. If matching is not optimal, engineering of the architectural core of tRNA often improves activity. Interestingly, atypical RNA frameworks can fulfil the requirements for recognition by aaRSs and are even found in nature (see above). Remarkably, since tRNA identity is defined by a small number of nts, it is possible to design artificial tRNAs with multiple specificities when these nts do not overlap. This has been done by transplanting the Ala and Phe identity sets in the framework of tRNA<sup>Asp</sup>; the resulting chimeric tRNA had a triple aminoacylation specificity (Frugier *et al.*, 1993). tRNA aminoacylation systems can also be engineered at the aaRS level. Thus, a single amino acid swap (Arg/Trp) between homologous positions in IleRS and MetRS is sufficient to switch identities (Auld and Schimmel, 1995).

An expanding area concerns design of orthogonal tRNA/aaRS pairs specifically producing aminoacylated tRNAs with unnatural amino acids in view to extend the genetic code. To ensure the fidelity of noncanonical protein synthesis, engineered tRNA/aaRS pairs must be orthogonal to the other tRNA/aaRSs pairs present in translation systems. To date, various methods were used to site-specifically incorporate more than 200 unnatural amino acids into proteins in response to nonsense or frameshift codons (see Xiao and Schultz, 2016; Katoh *et al.*, 2018; Cervettini *et al.*, 2020). Efficient orthogonality relies on subtle differences in identity sets and/or in catalytic sites of aaRSs of same canonical specificity within *Archaea*, *Bacteria* and *Eukaryota*. Noticeable, mutagenesis in editing domains of aaRSs are useful to improve orthogonality (Yadavalli and Ibba, 2012).

Typical examples are orthogonal tRNA/aaRS pairs based on the noncanonical identity of archaeal tRNA<sup>Pyl</sup> (e.g. Tharp *et al.*, 2018) or an engineered TyrRS from *Geobacillus stearothermophilus* that is orthogonal in yeast and mammalian cells (Qin *et al.*, 2020). The genetic code can also be reprogrammed using the flexizyme system in which an aaRS is replaced by a *de novo* tailored ribozyme (i.e. flexizyme) that is capable of charging unnatural amino acids on tRNAs that are subsequently used by a ribosome-dependent translation machinery (Katoh *et al.*, 2018). These methodologies, by generating proteins with novel properties, open a wealth of perspectives in basic research, molecular medicine and biotechnologies.

## F. A Refined View of tRNA Identity

For long, tRNA identity exclusively referred to identity for aminoacylation. At present, the concept of identity also refers to the many other functions of tRNA. In this context, only few experimental data concern this aspect, although it was already shown that such determinants can be distinct (e.g. G<sub>16</sub>, D<sub>20</sub> and D<sub>21</sub> in tRNA<sup>Ile</sup> for editing only) or overlapping (e.g. D/T loop elbow in tRNA<sup>Val</sup> and tRNA<sup>Leu</sup> or G<sub>3</sub>-U<sub>70</sub> in tRNA<sup>Ala</sup> for both aminoacylation and editing) (see Yadavalli *et al.*, 2012). On the other side, the increasing amount of data on tRNA sequences increased dramatically in the last decade and their analysis by novel bioinformatic and experimental tools (e.g. Branciamore *et al.*, 2018; Carter and Wills, 2019; Zamudio *et al.*, 2020) brought new insight in the structural, functional, dynamic and evolutionary understanding of tRNA. To date it becomes clear that tRNA identity recapitulates the evolutionary history of protein biosynthesis and even of life. This implies a shift of paradigm towards code biology (Barbieri, 2018) and the necessity to enlarge and refine the concept of identity (e.g. Galili *et al.*, 2016; Collins-Hed and Ardell, 2019).

For that, it becomes crucial to deconvolute identity determinants in terms of aminoacylation and associated functions. To achieve this, the field is actively seeking new approaches to find tRNA interacting partners or explore the tRNA aminoacylation *in vivo*. The MIST (Microarray Identification of Shifted tRNAs) approach validated for yeast ArgRS (Eriani *et al.*, 2015) but applicable to other tRNA binding proteins explores the ability of a given protein to discriminate between cognate and noncognate cellular tRNAs based on their concentrations and affinities (Figure 3A). At another level, the ISAP (tRNA Isoacceptor Specific Aminoacylation Profiling) approach allows to identify and quantify the amino acids attached to a tRNA species *in vivo* (Mohler *et al.*, 2017). Coupled with the highly sensitive mass spectrometry analysis, ISAP allows detection of subpopulations of noncognate or nonproteogenic amino acids in addition to cognate amino acid (Figure 3B) (Mohler *et al.*, 2017).

Altogether, much remains to be discovered on tRNA identities, notably specifying recognition by other tRNA binders such those involved during the modification processes. There are several dozens of tRNA modification enzymes catalysing the biogenesis of the hundred known tRNA modifications (Grosjean, 2009) in addition to the enzymes catalyzing tRNA splicing (see Schimmel, 2018).

Undoubtedly, the tRNA identity concept applies far beyond tRNA recognition by aaRSs. With the tremendous complexity of cellular isoacceptors and isodecoders tRNA species and multiplicity of tRNA interacting partners, it clearly appears that the field has still long-term prospects and a promising future ahead of him.

## References

- Ador L, Camasses A, Erbs P *et al.* (1999) Active site mapping of yeast aspartyl-tRNA synthetase by *in vivo* selection of enzyme mutations lethal for cell growth. *Journal of Molecular Biology* **288**: 231–242.
- Aldinger CA, Leisinger AK and Igloi GL (2012) The influence of identity elements on the aminoacylation of tRNA<sup>Arg</sup> by plant and *E. coli* arginyl-tRNA synthetases. *FEBS Journal* **279**: 3622–3638.
- Ambrogelly A, Frugier M, Ibba M *et al.* (2005a) Transfer RNA recognition by class I lysyl-tRNA synthetase from the Lyme disease pathogen *Borrelia burgdorferi*. *FEBS Letters* **579**: 2629–2634.
- Ambrogelly A, Kamtekar S, Stathopoulos C *et al.* (2005b) Asymmetric behavior of archaeal prolyl-tRNA synthetase. *FEBS Letters* **579**: 6017–6022.
- Andreev DE, Hirnet J, Terenin IM *et al.* (2012) Glycyl-tRNA synthetase specifically binds to the poliovirus IRES to activate translation initiation. *Nucleic Acids Research* **40**: 5602–5614.
- Auld DS and Schimmel P (1995) Switching recognition of two tRNA synthetases with an amino acid swap in a designed peptide. *Science* **267**: 1994–1996.
- Bacusmo JM, Kuzmishin AB, Cantara WA *et al.* (2018) Quality control by trans-editing factor prevents global mistranslation of non-protein amino acid  $\beta$ -aminobutyrate. *RNA Biology* **15**: 576–585.
- Banerjee K, Das B, and Gaugopadhyay G (2020) The guiding role of dissipation in kinetic proofreading networks: Implications for protein synthesis. *Journal of Chemical Physics* **152**: 111102.
- Barbieri M (2018) What is code biology? *BioSystems* **164**: 1–10.
- Bezerra AR, Simoes J, Lee W *et al.* (2013) Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. *Proceedings of the National Academy of Sciences of the USA* **110**: 11079–11084.



- Blaise M, Bailly M, Fréchin M *et al.* (2010) Crystal structure of a transfer-ribonucleoprotein particle that promotes asparagine formation. *EMBO Journal* **29**: 3118–3129.
- Blaise M, Becker HD, Keith G *et al.* (2004) A minimalist glutamyl-tRNA synthetase dedicated to aminoacylation of the tRNA<sup>Asp</sup> QUC anticodon. *Nucleic Acids Research* **32**: 2768–2775.
- Bonnefond L, Frugier M, Giegé R *et al.* (2005a) Human mitochondrial TyrRS disobeys the tyrosine identity rules. *RNA* **11**: 558–562.
- Bonnefond L, Giegé R and Rudinger-Thirion J (2005b) Evolution of the tRNA<sup>Tyr</sup>/TyrRS aminoacylation systems. *Biochimie* **87**: 873–883.
- Branciamore S, Gogoshin G, Di Giulio M *et al.* (2018) Intrinsic properties of tRNA molecules as deciphered via Bayesian network and distribution divergence analysis. *Life (Basel)* **8**: E5.
- Bullock TL, Uter N, Nissan TA *et al.* (2003) Amino acid discrimination by a class I aminoacyl-tRNA synthetase specified by negative determinants. *Journal of Molecular Biology* **328**: 395–408.
- Carter CW Jr and Wills PR (2019) Class I and II aminoacyl-tRNA synthetase tRNA groove discrimination created the first synthetase-tRNA cognate pairs and was therefore essential to the origin of genetic coding. *International Union of Biochemistry and Molecular Biology – Life* **71**: 1088–1098.
- Cela M, Paulus C, Santos MAS *et al.* (2018) *Plasmodium* apicoplast tyrosyl-tRNA synthetase recognizes an unusual, simplified identity set in cognate tRNA<sup>Tyr</sup>. *PLoS ONE* **13**: e0209805.
- Cervettini D, Tang S, Fried SD *et al.* (2020) Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase-tRNA pairs. *Nature Biotechnology* **in press** doi:10.1038/s41587-020-0479-2.
- Chan CW, Badong D, Rajan R *et al.* (2020) Crystal structures of an unmodified bacterial tRNA reveal intrinsic structural flexibility and plasticity as general properties of unbound tRNAs. *RNA* **26**: 278–289.
- Chimnaronk S, Gravers Jeppesen M, Susuki T *et al.* (2005) Dual-mode recognition of noncanonical tRNAs<sup>Ser</sup> by seryl-tRNA synthetase in mammalian mitochondria. *EMBO Journal* **24**: 3369–3379.
- Chuawong P, Likittrakulwong W, Suebka S *et al.* (2020) Anticodon-binding domain swapping in a nondiscriminating aspartyl-tRNA synthetase reveals contributions to tRNA specificity and catalytic activity. *Proteins* **in press** doi:10.1002/prot.25881.
- Collins-Head AI and Ardell AH (2019) Match fitness landscapes for macromolecular interaction networks: Selection for translational accuracy and rate can displace tRNA-binding interfaces of non-cognate aminoacyl-tRNA synthetases. *Theoretical Population Biology* **129**: 68–80.
- Crnkovic A, Vargas-Rodriguez O and Söll D (2019) Plasticity and constraints of tRNA aminoacylation define directed evolution of aminoacyl-tRNA synthetases. *International Journal of Molecular Sciences* **20**: 2294.
- Cusack S, Berthet-Colominas C, Härtlein M *et al.* (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.
- Dixit S, Henderson JC and Alfonzo JD (2019) Multi-substrate specificity and the evolutionary basis for interdependence in tRNA editing and methylation enzymes. *Frontiers in Genetics* **10**: 104.
- Dreher TW (2010) Viral tRNAs and tRNA-like structures. *Wiley Interdisciplinary Reviews: RNA* **1**: 402–414.
- Dreher TW, Tsai CH and Skuzeski JM (1996) Aminoacylation identity switch of turnip yellow mosaic virus RNA from valine to methionine results in an infectious virus. *Proceedings of the National Academy of Sciences of the USA* **93**: 12212–12216.

- Dulic M, Cvestesic N, Zivkovic I *et al.* (2018) Kinetic origin of substrate specificity in post-transfer editing by leucyl-tRNA synthetase. *Journal of Molecular Biology* **430**: 1–16.
- Ebel J-P, Giegé R, Bonnet J *et al.* (1973) Factors determining the specificity of the tRNA aminoacylation reaction. *Biochimie* **55**: 547–557.
- Eriani G, Delarue M, Poch O *et al.* (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**: 203–206.
- Eriani, G, Karam J, Jacinto J *et al.* (2015) MIST, a novel approach to reveal hidden substrate specificity in aminoacyl-tRNA synthetases. *PLoS One* **10**: e0130042.
- Fender A, Gaudry A, Jühling F *et al.* (2012) Adaptation of aminoacylation identity rules to mammalian mitochondria. *Biochimie* **94**: 1090–1097.
- Fersht AR, Ashford JS, Bruton CJ *et al.* (1975) Active site titration and aminoacyl adenylate binding stoichiometry of aminoacyl-tRNA synthetases. *Biochemistry* **14**: 1–4.
- Francin M and Mirande M (2006) Identity elements for specific aminoacylation of a tRNA by mammalian lysyl-tRNA synthetase bearing a nonspecific tRNA-interacting factor. *Biochemistry* **45**: 10153–10160.
- Frugier M, Florentz C, Schimmel P *et al.* (1993) Triple aminoacylation specificity of a chimerized transfer RNA. *Biochemistry* **32**: 14053–14061.
- Galili T, Gingold H, Shaul S *et al.* (2016) Identifying the ligated amino acid of archaeal tRNAs based on positions outside the anticodon. *RNA* **22**: 1477–1491.
- Giegé R and Lapointe J (2009) Transfer RNA aminoacylation and modified nucleosides. In: Grosjean H (ed.) *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*, pp. 475–492. Georgetown, TX: Landes Bioscience.
- Giegé R, Puglisi JD and Florentz C (1993) tRNA structure and aminoacylation efficiency. *Progress in Nucleic Acid Research and Molecular Biology* **45**: 129–206.
- Giegé R, Sissler M and Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Research* **26**: 5017–5035.
- Giegé R and Springer M (2016) Aminoacyl-tRNA synthetases in the bacterial world (updated version). In: *EcoSal Plus, Domain 4: Synthesis and Processing of Macromolecules* (Lovett ST, ed) *EcoSal – Escherichia coli and Salmonella: Cellular and Molecular Biology*. Washington, DC: ASM Press. <http://www.ecosal.org>
- Grosjean H (ed.) (2009) *DNA and RNA Modification Enzymes: Comparative Structure, Mechanism, Functions, Cellular Interactions and Evolution*. Austin, TX: Landes Bioscience.
- Hadd A and Perona JJ (2015) Coevolution of specificity determinants in eukaryotic glutamyl- and glutaminyl-tRNA synthetases. *Journal of Molecular Biology* **426**: 3619–3633.
- Hilander T, Zhou X-L, Konovalova S *et al.* (2018) Editing activity for eliminating mischarged tRNAs is essential in mammalian mitochondria. *Nucleic Acids Research* **46**: 849–860.
- Hou Y-M and Schimmel P (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* **333**: 140–145.
- Huang Q, Zhou XL, Hu QH *et al.* (2014) A bridge between the aminoacylation and editing domains of leucyl-tRNA synthetase is crucial for its synthetic activity. *RNA* **20**: 1440–1450.
- Igloi GL and Leisinger AK (2014) Identity elements for the aminoacylation of metazoan mitochondrial tRNA<sup>Arg</sup> have been widely conserved throughout evolution and ensure the fidelity of the AGR codon reassignment. *RNA Biology* **11**: 1313–1323.
- Ito T and Yokoyama S (2010) Two enzymes bound to one tRNA assume alternative conformations for consecutive reactions. *Nature* **467**: 612–616.
- Johanson K, Hoang T, Sheth M *et al.* (2003) GRS1, a yeast tRNA synthetase with a role in mRNA 3' end formation. *Journal of Biological Chemistry* **278**: 35923–35930.

- Jühling T, Duchardt-Ferner E, Bonin S *et al.* (2018) Small but large enough: Structural properties of armless mt-tRNAs from the nematode *Romanomermis culicivora*. *Nucleic Acids Research* **46**: 9170–9180.
- Katoh T, Iwane Y and Suga H (2018) tRNA engineering for manipulating genetic code. *RNA Biology* **15**: 453–460.
- Kim Y, Opron K and Burton ZF (2019) A tRNA- and anticodon-centric view of the evolution of aminoacyl-tRNA synthetases, tRNAomes, and the genetic code. *Life (Basel)* **9**: E37.
- Kuncha SK, Kruparani SP and Sankaranarayanan R (2019) Chiral checkpoints during protein biosynthesis. *Journal of Biological Chemistry* **294**: 16535–16548.
- Kwon NH, Fox PL and Kim S (2019) Aminoacyl-tRNA synthetases as therapeutic targets. *Nature Reviews Drug Discovery* **18**: 629–650.
- Laowanapiban P, Kapustina M, Vornrhein C *et al.* (2009) Independent saturation of three TrpRS subsites generates a partially assembled state similar to those observed in molecular simulations. *Proceedings of the National Academy of Sciences of the USA* **106**: 1790–1795.
- Lee JW, Beebe K, Nangle LA *et al.* (2006) Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* **443**: 50–55.
- Lee YH, Yo LT, Chang CP *et al.* (2019) Naturally occurring dual recognition of tRNA<sup>His</sup> substrates with and without a universal identity element. *RNA Biology* **16**: 1275–1285.
- Levi O and Arava Y (2019) mRNA association by aminoacyl tRNA synthetase occurs at a putative anticodon mimic and autoregulates translation in response to tRNA levels. *PLoS Biology* **17**: e3000274.
- Li L, Boniecki MT, Jaffe JD *et al.* (2011) Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in *Mycoplasma* parasites. *Proceedings of the National Academy of Sciences of the USA* **108**: 9378–9383.
- Li R, Macnamara LM, Leuchter JD *et al.* (2015) MD simulations of tRNA and aminoacyl-tRNA synthetases: Dynamics, folding, binding, and allostery. *International Journal of Molecular Sciences* **16**: 15872–15902.
- Liu Y, Nakamura A, Nakazawa Y *et al.* (2014) Ancient translation factor is essential for tRNA-dependent cysteine biosynthesis in methanogenic archaea. *Proceedings of the National Academy of Sciences of the USA* **111**: 10520–10525.
- Mahler M, Miller FW, Fritzler MJ (2014) Idiopathic inflammatory myopathies and the anti-synthetase syndrome: a comprehensive review. *Autoimmunity Reviews* **13**: 367–371.
- McClain WH and Foss K (1988) Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science* **240**: 793–796.
- McShane A, Hok E, Tomberlin J *et al.* (2016) The enzymatic paradox of yeast arginyl-tRNA synthetase: Exclusive arginine transfer controlled by a flexible mechanism of tRNA recognition. *PLoS One* **11**: e0148460.
- Martinis SA and Schimmel P (1995) Small RNA oligonucleotide substrates for specific aminoacylations. In: Söll D and RajBhandary UL (eds) *tRNA: Structure, Biosynthesis, and Function*, pp. 349–370. Washington, DC: American Society for Microbiology Press.
- Mohler K, Mann R and Ibba M (2017) Isoacceptor specific characterization of tRNA aminoacylation and misacylation *in vivo*. *Methods* **113**: 127–131.
- Mukai T, Reynolds NM, Crnkovic A *et al.* (2017) Bioinformatic analysis reveals archaeal tRNA<sup>Tyr</sup> and tRNA<sup>Trp</sup> identities in *Bacteria*. *Life (Basel)* **7**: E8.
- Muramatsu T, Nishikawa K, Nemoto F *et al.* (1988) Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature* **336**: 179–181.

- Musier-Forsyth K and Schimmel P (1999) Atomic determinants for aminoacylation of RNA minihelices and relationship to genetic code. *Accounts of Chemical Research* **32**: 368–375.
- Naganuma M, Sekine S, Chong YE *et al.* (2014) The selective tRNA aminoacylation mechanism based on a single G-U pair. *Nature* **510**: 507–511.
- Neuenfeldt A, Lorber B, Ennifar E *et al.* (2013) Thermodynamic properties distinguish human mitochondrial aspartyl-tRNA synthetase from bacterial homolog with same 3D architecture. *Nucleic Acids Research* **41**: 2698–2708.
- Perona JJ and Hou Y-M (2007) Indirect readout of tRNA for aminoacylation. *Biochemistry* **46**: 10419–10432.
- Perret V, Garcia A, Grosjean H *et al.* (1990) Relaxation of transfer RNA specificity by removal of modified nucleotides. *Nature* **344**: 787–789.
- Pütz J, Puglisi JD, Florentz C *et al.* (1991) Identity elements for specific aminoacylation of yeast tRNA<sup>Asp</sup> by cognate aspartyl-tRNA synthetase. *Science* **252**: 1696–1699.
- Pütz J, Puglisi JD, Florentz C *et al.* (1993) Additive, cooperative and anti-cooperative effects between identity nucleotides of a tRNA. *EMBO Journal* **12**: 2949–2957.
- Qin X, Tang H, Cao W *et al.* (2020) An orthogonal tyrosyl-tRNA synthetase/tRNA pair from a thermophilic bacterium for an expanded eukaryotic genetic code. *Biochemistry* **59**: 90–99.
- Rao BS and Jackman JE (2015) Life without post-transcriptional addition of G-1: Two alternatives for tRNA<sup>His</sup> identity in Eukarya. *RNA* **21**: 243–253.
- Rodriguez-Hernandez A and Perona JJ (2011) Heat maps for intramolecular communication in an RNP enzyme encoding glutamine. *Structure* **19**: 386–396.
- Rosen AE and Musier-Forsyth K (2004) Recognition of G<sub>-1</sub>:C<sub>73</sub> atomic groups by *Escherichia coli* histidyl-tRNA synthetase. *Journal of the American Chemical Society* **126**: 64–65.
- Rould MA, Perona JJ, Söll D *et al.* (1989) Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA<sup>Gln</sup> and ATP at 2.8 Å resolution. *Science* **246**: 1135–1142.
- Ruan B, Palioura S, Sabina J *et al.* (2008) Quality control despite mistranslation caused by an ambiguous genetic code. *Proceedings of the National Academy of Sciences of the USA* **105**: 16502–16507.
- Ruff M, Krishnaswamy S, Boeglin M *et al.* (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA<sup>Asp</sup>. *Science* **252**: 1682–1689.
- Rybak MY, Rayevsky AV, Gudzera OI *et al.* (2019) Stereospecificity control in aminoacyl-tRNA-synthetases: New evidence of d-amino acids activation and editing. *Nucleic Acids Research* **47**: 9777–9788.
- Saint-Léger A, Bello C, Dans PD *et al.* (2016) Saturation of recognition elements blocks evolution of new tRNA identities. *Science Advances* **2**: e1501860.
- Salinas-Giegé T, Giegé R and Giegé P (2015) tRNA biology in mitochondria. *International Journal of Molecular Sciences* **16**: 4518–4559.
- Sarkar J, Poruri K, Boniecki M *et al.* (2012) Yeast mitochondrial leucyl-tRNA synthetase CP1 domain has functionally diverged to accommodate RNA splicing at expense of hydrolytic editing. *Journal of Biological Chemistry* **287**: 14772–14781.
- Schimmel P (2018) The emerging complexity of the tRNA world: Mammalian tRNAs beyond protein synthesis. *Nature Review in Molecular and Cellular Biology* **19**: 45–58.
- Schimmel P, Giegé R, Moras D *et al.* (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proceedings of the National Academy of Sciences of the USA* **90**: 8763–8768.

- Schmitt E, Meinnel T, Panvert M *et al.* (1993) Two acidic residues of *Escherichia coli* methionyl-tRNA synthetase act as negative discriminants towards the binding of non-cognate tRNA anticodons. *Journal of Molecular Biology* **233**: 615–628.
- Schwartz MH and Pan T (2017) Function and origin of mistranslation in distinct cellular contexts. *Critical Reviews in Biochemistry and Molecular Biology* **52**: 205–219.
- Song Y, Zhou H, Vo M-N *et al.* (2017) Double mimicry evades tRNA synthetase editing by toxic vegetable-source non-proteinogenic amino acid. *Nature Communications* **8**: 2281.
- Stephen P, Ye S, Zhou M *et al.* (2018) Structure of *Escherichia coli* arginyl-tRNA synthetase in complex with tRNA<sup>Arg</sup>: Pivotal role of the D-loop. *Journal of Molecular Biology* **430**: 1590–1606.
- Sun LT, Gomes AC, He WW *et al.* (2016) Evolutionary gain of alanine mischarging to noncognate tRNAs with a G4:U69 base pair. *Journal of the American Chemical Society* **138**: 12948–12955.
- Tahmasebi S, Khoutorsky A, Mathews MB *et al.* (2018) Translation deregulation in human disease. *Nature Review in Molecular and Cellular Biology* **19**: 791–807.
- Tharp JM, Ehnbohm A and Liu WR (2018) tRNA<sup>Pyl</sup>: Structure, function, and applications. *RNA Biology* **15**: 441–452.
- Tinkle-Peterson E and Uhlenbeck OC (1992) Determination of recognition nucleotides for *Escherichia coli* phenylalanyl-tRNA synthetase. *Biochemistry* **31**: 10380–10389.
- Tocchini-Valentini G, Saks M and Abelson J (2000) tRNA leucine identity and recognition sets. *Journal of Molecular Biology* **298**: 779–793.
- Tsuchiya W and Hasegawa T (2009) Molecular recognition of tryptophan tRNA by tryptophanyl-tRNA synthetase from *Aeropyrum pernix* K1. *Journal of Biochemistry (Tokyo)* **145**: 635–641.
- Wang C, Guo Y, Tian Q *et al.* (2015) SerRS-tRNA<sup>Sec</sup> complex structures reveal mechanism of the first step in selenocysteine biosynthesis. *Nucleic Acids Research* **43**: 10534–10545.
- Wolfson AD, Khvorova AM, Sauter C *et al.* (1999) Mimics of yeast tRNA<sup>Asp</sup> and their recognition by aspartyl-tRNA synthetase. *Biochemistry* **38**: 11926–11932.
- Xiao H and Schultz PG (2016) At the interface of chemical and biological synthesis: An expanded genetic code. *Cold Spring Harbor Perspectives in Biology* **8**: a023945.
- Yadavalli SS and Ibba M (2012) Quality control in aminoacyl-tRNA synthesis its role in translational fidelity. *Advanced Protein Chemistry and Structural Biology* **86**: 1–43.
- Zamudio GS, Palacios-Perez M and José MV (2020) Information theory unveils the evolution of tRNA identity elements in the three domains of life. *Theory in Biosciences* **139**: 77–85.
- Zeng QY, Peng GX, Li G *et al.* (2019) The G3-U70-independent tRNA recognition by human mitochondrial alanyl-tRNA synthetase. *Nucleic Acids Research* **47**: 3072–3085.
- Zhang CM, Liu C, Slater S *et al.* (2008) Aminoacylation of tRNA with phosphoserine for synthesis of cysteinyl-tRNA<sup>Cys</sup>. *Nature Structural and Molecular Biology* **15**: 507–514.

---

## Further Reading

- Beuning PJ and Musier-Forsyth K (1999) Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers* **52**: 1–28.
- Francklyn CS, First EA, Perona JJ and Hou Y-M (2008) Methods for kinetic and thermodynamic analysis of aminoacyl-tRNA synthetases. *Methods* **44**: 100–118.
- Köhler C and RajBhandary UL (eds) (2009) *Protein Engineering*. Berlin: Springer-Verlag. In vol **22** of 'Nucleic Acids and Molecular Biology', Gross HJ (series ed).
- Kuo ME and Antonellis A (2020) Ubiquitously expressed proteins and restricted phenotypes: Exploring cell-specific sensitivities to impaired tRNA charging. *Trends in Genetics* **36**: 105–117.

- Levi O, Garin S, Arava Y. (2020) RNA mimicry in post-transcriptional regulation by aminoacyl tRNA synthetases. *Wiley Interdiscip Review RNA* **11**: e1564.
- McClain WH (1993) Rules that govern tRNA identity in protein synthesis. *Journal of Molecular Biology* **234**: 257–280.
- Mans MW, Pleij CWA and Bosch L (1991) tRNA-like structures. Structure, function and evolutionary significance. *European Journal of Biochemistry* **201**: 303–324.
- Melnikov SV and Söll D. (2019) Aminoacyl-tRNA Synthetases and tRNAs for an expanded genetic code: what makes them orthogonal? *International Journal Molecular Sciences* **20**: 1929.
- Mohler K and Ibba M (2017) Translational fidelity and mistranslation in the cellular response to stress. *Nature Microbiology* **2**: 17117.
- Watanabe K (2010) Unique features of animal mitochondrial translation systems – The non-universal genetic code, unusual features of the translational apparatus and their relevance to human mitochondrial diseases. *Proceedings of the Japan Academy Series B, Physical and Biological Sciences* **86**: 11–39.

## Glossary

*Alloproteins* = Proteins containing unnatural amino acids, such as analogues of natural amino acids or spectroscopically active amino acid derivatives.

*Editing/Proofreading* = A panel of mechanisms important to ensure fidelity of tRNA aminoacylation consisting in efficient hydrolysis of misactivated amino acids or improperly charged/mischarged tRNAs.

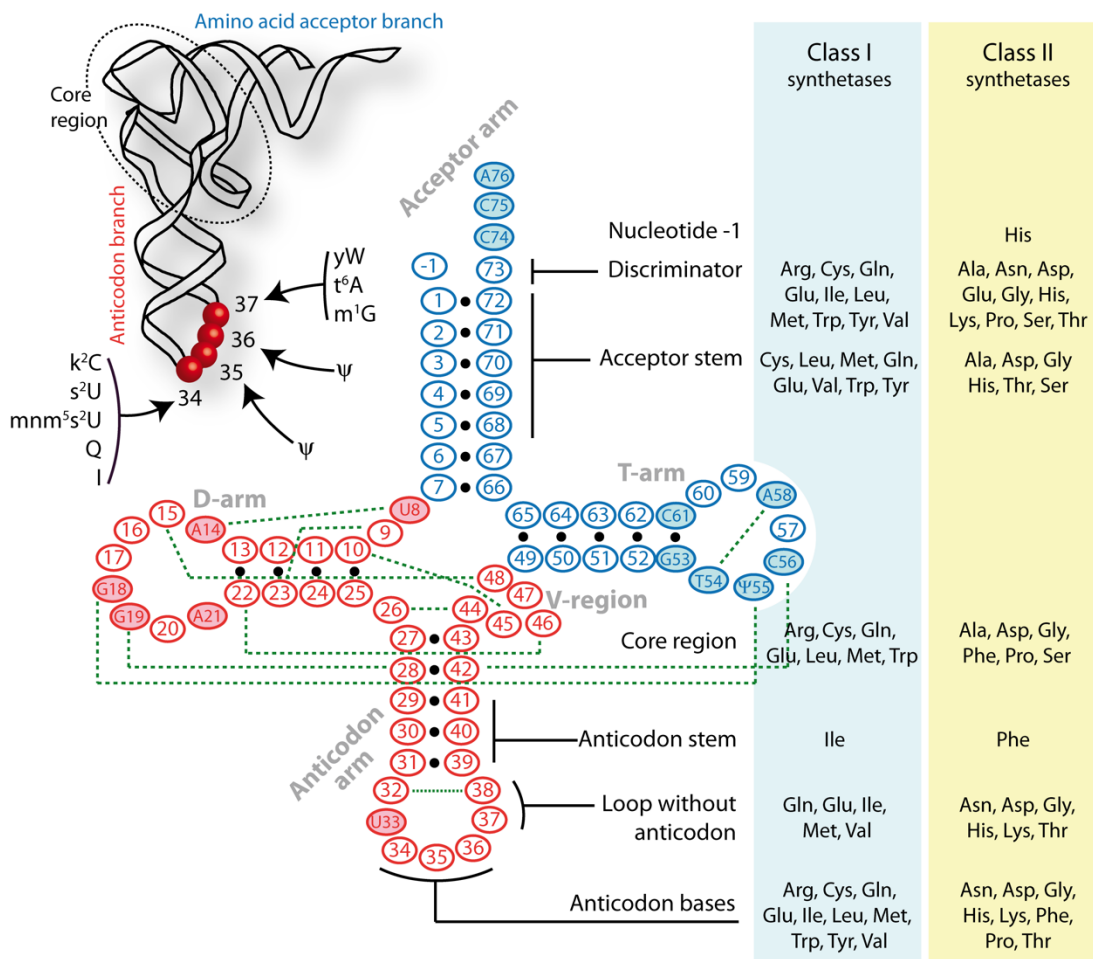
*Identity rules* = These rules account for the specificity of tRNA aminoacylation by their cognate aminoacyl-tRNA synthetases and rely on several determinants generally interacting with aaRSs and on less well-characterised antideterminants that prevent tRNA interactions with noncognate aaRSs.

*Identity sets* = Such sets are constituted by the diverse determinants (mainly nucleotidic bases but also structural features) defining the aminoacylation specificity of a tRNA. Identity sets are composed of major and minor determinants (having strong and weak effects on tRNA aminoacylation), with major determinants usually conserved during evolution.

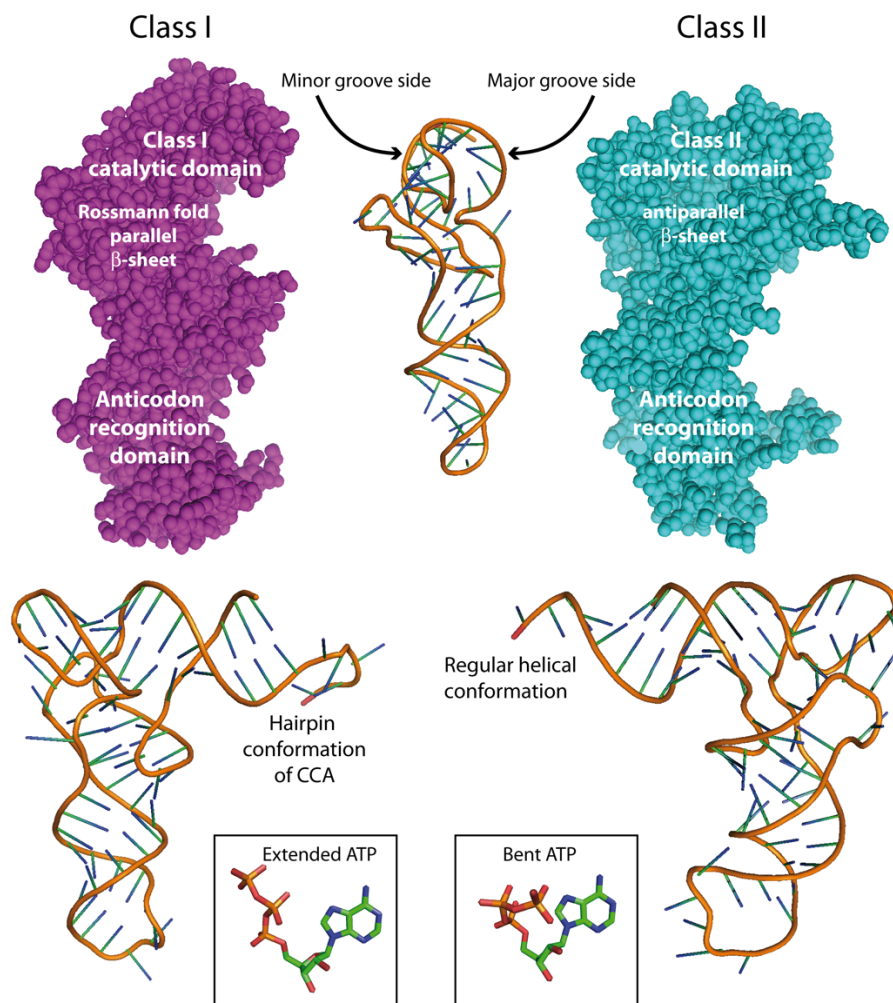
*Kinetic/catalytic specificity/efficiency* = In the present context it is defined by the  $k_{\text{cat}}/K_m$  ratio of tRNA aminoacylation reactions, with  $k_{\text{cat}}$  the catalytic rate and  $K_m$  the Michaelis constant representing an approximation of the inverse of tRNA affinity for a synthetase.

*Orthogonal aaRS:tRNA pairs* = They consist in variant aaRS and tRNA molecules engineered in such a way that an orthogonal aaRS becomes specific of an unnatural amino acid. This requires (1) that an orthogonal aaRS solely charges the unnatural amino acid on an orthogonal tRNA partner and (2) that this orthogonal tRNA is not charged by a natural amino acid by any other aaRS present in the protein synthesis system for which the orthogonal aaRS:tRNA pair was engineered.

*Strength of an identity determinant* = Is defined *in vitro* by the effect on the  $k_{\text{cat}}/K_M$  ratio produced by its mutation; this leads to a loss of catalytic efficiency (defined as  $L = (k_{\text{cat}}/K_m)_{\text{wild type}} / (k_{\text{cat}}/K_m)_{\text{mutant}}$ ). *In vivo*, strength of a determinant, as a result of its mutation, is defined by the suppression efficiency of a stop codon in a reporter gene, measured by the frequency of amino acid incorporation at the suppressed position (often at position 10 of *E. coli* dihydrofolate reductase). Major determinants are conserved in evolution and are aided by minor determinants that show phylogenetic idiosyncrasies.

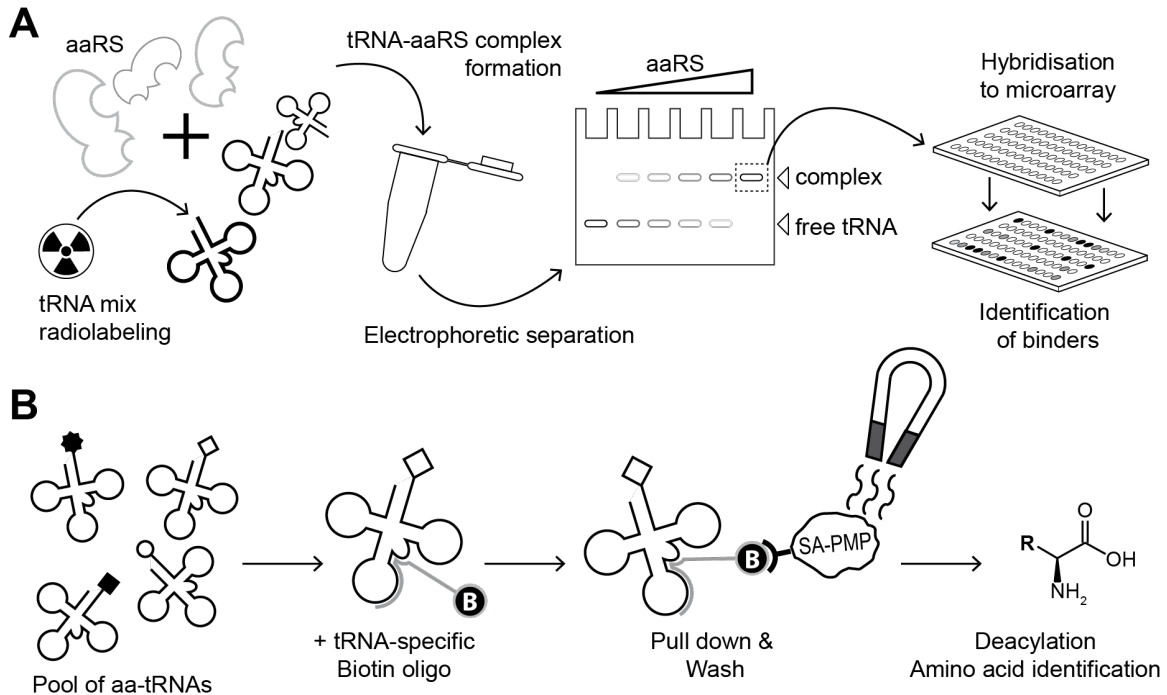


**Figure 1** Cloverleaf folding of tRNA with location of known identity determinants and its three-dimensional L-shaped organisation. The standard cloverleaf structure of cytosolic tRNAs and conventional numbering system are used. Constant nucleotides (nts) are explicitly indicated (T and  $\Psi$  are modified residues: ribothymidine and pseudouridine). Note the presence within the variable region (nts 44 to 48) of the long extra arm of tRNA<sup>Leu</sup>, tRNA<sup>Ser</sup> and tRNA<sup>Tyr</sup>. The • symbol indicates Watson–Crick base pairings (including G–U pairs); dotted green lines indicate other pairings, mostly between constant and/or semi-constant residues important for tRNA L-shaped architecture. The inset represents the L-shaped structure of tRNA and highlights its different domains. Location of identity elements in the tRNA molecule is shown with a distinction between identities of tRNAs recognised by class I and class II synthetases. Characterised individual modified residues that act as identity determinants in tRNA anticodon loops are shown in the inset ( $k^2C$ , lysidine;  $s^2U$ , 2-thiouridine;  $mnm^5s^2U$ , 5-methylaminomethyl-2-thiouridine; Q, queuosine; I, inosine;  $\Psi$ , pseudouridine; m1G, 1-methylguanosine; t6A, N6-threonylcarbamoyladenine; yW, wybutosine).



**Figure 2** Synthetase class-dependent binding modes of tRNA and ATP molecules as revealed in GlnRS:tRNA<sup>Gln</sup>:ATP and AspRS:tRNA<sup>Asp</sup>:ATP complexes. The upper part of the figure shows the different binding approaches of the catalytic domain of the aaRS molecules relative to the acceptor arm of tRNA. The structures of monomeric *E. coli* GlnRS and dimeric *S. cerevisiae* AspRS (only one monomer represented) are chosen as class-representative aaRSs and are displayed so that to emphasise the two binding modes of tRNA via the minor or major groove side of its amino acid acceptor helix. The lower part of the figure shows the conformation of tRNA and ATP in the complexes. Binding of tRNA implies that CCA folds back in class I complexes and remains in regular helical conformation in class II complexes. The class-specific architecture of the aaRS catalytic domain implies further that ATP exhibits an extended conformation in class I aaRSs (reminiscent of that found in other enzymes containing a Rossmann fold) and a bent conformation in class II aaRSs (with the  $\gamma$ -phosphate folded back over the adenine base). The tRNA and ATP conformations are those found in the GlnRS:tRNA<sup>Gln</sup>:ATP (left side) and AspRS:tRNA<sup>Asp</sup>:ATP (right side) complexes. Notice that tRNA binding on GlnRS and AspRS implies an unfolding of the anticodon loops with the unstacking of the anticodon bases favouring specific contacts with amino acids from the anticodon-binding domains.





**Figure 3** Outline of two recent methods for investigation on tRNA recognition and aminoacylation by aaRSs. **(A)** The MIST method was designed to study the formation of tRNA/aaRS complexes independently from the aminoacylation reaction (Eriani *et al.*, 2015). A crude tRNA preparation is radiolabeled in 5' with <sup>32</sup>P, purified and incubated with varying amounts of purified aaRS. Free tRNAs are separated from complexed species on non-denaturing gel. Complexed tRNAs corresponding to high molecular weight bands are eluted and hybridised on tRNA microarrays where they are identified. **(B)** The ISAP method allows to identify and quantify the amino acids attached to a tRNA species *in vivo* (Mohler *et al.*, 2017). ISAP allows compilation of aminoacylation profiles for specific isoacceptors tRNAs. A given tRNA is isolated and hybridised to a biotinylated oligonucleotide probe. The probe is captured on Streptavidin conjugated paramagnetic particles (SA-PMP) and the amino acid moiety is isolated and analysed. Subsequent analysis using the highly sensitive mass spectrometry approach allows detection of the amino acids characterising the small populations of noncognate or nonproteogenic amino acids in addition to the cognate amino acid.

**Table 1 Positive identity determinants for aminoacylation experimentally characterised in cytosolic tRNAs**

tRNA domain	Position	A	U	G	C	Sub domain	
Accepto r arm	-1			<u>His<sup>a</sup></u>			
	<b>73</b>	Ile, <b>Leu, Met</b> , <i>Trp, Tyr, Val</i>	<b>Cys</b>	Arg, Gln, <b>Trp</b>	<u>Arg</u>		
		<u>Ala, Gly, His</u> , <u>Lys, <b>Phe</b>, Pro</u>	<u>Gly, Thr</u>	<b>Asp, Asn, Ser</b>	<u>His</u>		
	72			<b>Pro</b>	<i>Ser</i>		
	1•72 <sup>b</sup>	<i>Trp</i>	<i>Gln</i>	<i>Glu, Trp, Tyr</i>	<u>Tyr</u>		
				<b>Gly, Thr</b>			
	2•71	<i>Glu</i>	<i>Cys, Gln, Met, Trp</i>	<i>Gln<sup>c</sup></i>			
			<b>Ala, Asp, Ser</b>		<b>Gly<sup>b</sup>, Thr</b>		
	3•70 <sup>d</sup>			<i>Leu, Gln<sup>c</sup>, Trp, Val</i>	<i>Gln, Cys, Leu, Met</i>		
		<u>Ser</u>	<u>Thr</u>	<b>Ala<sup>d</sup>, Gly, Ser</b>			
	4•69	<i>Leu</i>	<i>Val, Met, Trp</i>		<i>Ile</i>		
		<u>Ser</u>		<b>Ala, Ser</b>			
	5•68	<i>Met</i>		<i>Leu, Trp</i>			
	Core region	9	<i>Trp</i>				
		10			<i>Gln<sup>c</sup></i>		
				<b>Asp</b>			
10•25				<b>Asp<sup>d,e</sup>, Gly</b>			
11•24			<i>Glu</i>				
12•24			<i>Ile<sup>f</sup></i>				
20 <sup>g</sup>		<i>Arg</i>	<i>Arg</i>		<i>Arg, Leu</i>		
			<b>Phe</b>	<b>Ala, Phe</b>			
44				<b>Phe</b>			
45			<b>Phe</b>				
59			<b>Phe</b>				
60			<b>Phe</b>				
VR <sup>h</sup>						<b>Ser</b>	
D-arm					<b>Met</b>		
Tertiary interacti ons					<b>Cys<sup>i</sup>, Glu<sup>j</sup>, Ile<sup>k</sup>, Leu<sup>l</sup></b>		
					<b>Pro<sup>k</sup></b>		
Anticodo n arm	27•43			<b>Phe</b>			
	28•42			<b>Phe</b>			
	29•41				<i>Ile</i>		
	30•40				<b>Phe</b>		
	31•39	<b>Phe</b>					
	32		<i>Gln</i>				
	33		<i>Gln, Glu, Met</i>				

34	Ile <sup>m</sup>	Gln, Glu <sup>m</sup> , Tyr	Cys, Gln, Ile, Tyr	Gln, Met, Trp
	<i>Phe</i>	Lys <sup>m</sup>	<i>Asn, Asp, His, Phe, Thr</i>	<i>His</i>
35	Ile, Leu, Met, Val	Gln, Glu, Tyr <sup>m</sup>		Arg, Cys, Trp, Tyr
	<i>Phe</i>	<i>Asn, Asp, His, Lys</i>	<i>His, Pro, Thr</i>	<i>Phe<sup>m</sup></i>
36	Cys, Trp, Tyr	<b>Arg, Ile, Met</b>	<b>Arg, Gln</b>	
	<i>Phe<sup>m</sup></i>	<i>Asn, His, Lys, Thr</i>	<i>His, Pro</i>	<b>Asp</b>
37	Gln, Glu, Ile <sup>m</sup> , Met		Leu	
	<i>Phe<sup>m</sup></i>			
38	Arg, Gln, Ile, Met	Gln, Ile		
				<b>Asp</b>

The three-letter code for amino acids is used to distinguish the determinants specifying the 20 canonical tRNA identities, with determinants recognised by class I (black) and class II (blue print and background) aaRSs explicitly shown: if in normal script, identity determinants were characterised by the *in vitro* method, sometimes completed by *in vivo* data; if in italics, they were only searched by the *in vivo* method; if in bold, they were characterised in both *Bacteria* and *Eukaryota* (mostly in the 2 model organisms *E. coli* and *S. cerevisiae*; less data come from *T. thermophilus* and human and a few from 12 other organisms). Nonconservation of identity determinants in evolution is indicated by underlined amino acids. Data cover cytosolic tRNAs and do not discriminate between the strength of determinants (see text for details). Watson–Crick (WC) N•N pairs as determinants in helical stems are numbered as in **Figure 1** (for a given identity, these N•N pairs are generally not conserved in phylogenies). References for early data (Giegé *et al.*, 1998) and for newer data covering plant Arg (Aldinger *et al.*, 2012), archaeal *Aeropyrum pernix* Phe, Thr, Trp and Tyr and *Haloferax volcanii* Leu, Thr and Tyr (see Tsuchiya and Hasegawa, 2009), bacterial/eukaryal Tyr (Bonfond *et al.*, 2005a) and mammalian Lys (Francin and Mirande, 2006) identities. The most informative tRNA positions coding for ancient discriminator determinants N<sub>73</sub> and anticodon triplets N<sub>34</sub>N<sub>35</sub>N<sub>36</sub> are highlighted (red print).

<sup>a</sup> Except in a few eukaryal taxa, e.g. *Trypanosoma brucei* (predicted by Ardell and biochemically validated by Rao and Jackman, 2015).

<sup>b</sup> In case of WC identity pairs referred in the A, U, G or C columns, they are A•U, U•A, G•C or C•G pairs, respectively; the same for all N•N pairs.

<sup>c</sup> Importance of the 2-amino group of G, as demonstrated by reduced activity of *E. coli* tRNA<sup>Gln</sup> mutants with G<sub>2</sub>, G<sub>3</sub> and G<sub>10</sub> replaced by Inosine.

<sup>d</sup> Is G<sub>3</sub>–U<sub>70</sub> in all tRNA<sup>Ala</sup> except mitochondrial tRNA<sup>Ala</sup> and G<sub>10</sub>–U<sub>25</sub> in *S. cerevisiae* tRNA<sup>Asp</sup>.

<sup>e</sup> Alteration of the G<sub>45</sub>:G<sub>10</sub>–U<sub>25</sub> triple.

<sup>f</sup> Alteration of the G<sub>46</sub>:U<sub>12</sub>•A<sub>23</sub> triple pairing in *E. coli* tRNA<sup>Ile</sup>.

<sup>g</sup> Location at N<sub>20a</sub>; for *S. cerevisiae* tRNA<sup>Arg</sup> (ICG isoacceptor) there is an ambiguity in the D-loop position (McShane *et al.*, 2016).

<sup>h</sup> VR for variable region (from N<sub>44</sub> to N<sub>48</sub>, with insertions of 4 to 24 nts between N<sub>47</sub> and N<sub>48</sub> and deletion of N<sub>47</sub>).

<sup>i</sup> Levitt pair G<sub>15</sub>–C<sub>48</sub> (trans WC pair) in *E. coli* tRNA<sup>Cys</sup>.

<sup>j</sup> Lack of nt 47 in *E. coli* tRNA<sup>Glu</sup> alters the stability of the U<sub>13</sub>–G<sub>22</sub>:A<sub>46</sub> triple pairing.

<sup>k</sup> Alteration of the C<sub>13</sub>•G<sub>22</sub>:G<sub>46</sub> triple pairing and of reverse Hoogsteen pair U<sub>8</sub>–A<sub>14</sub> in *E. coli* tRNA<sup>Ile</sup>.

<sup>l</sup> Alteration of A<sub>14</sub> in the reverse Hoogsteen pair U<sub>8</sub>–A<sub>14</sub> in *E. coli* tRNA<sup>Leu</sup>.

<sup>m</sup> Post-transcriptional modifications involved in identity.

**Table 2 Examples of antideterminants identified in tRNAs and synthetases from the same organism**

Antideterminant	Organism	In tRNA/aaRS	Against aaRS/tRNA
Nucleoside			
k <sup>2</sup> C <sub>34</sub>	<i>Escherichia coli</i>	tRNA <sup>Ile</sup>	MetRS
m <sup>1</sup> G <sub>37</sub>	yeast	tRNA <sup>Asp</sup>	ArgRS
A <sub>36</sub>	<i>E. coli</i>	tRNA <sup>Arg</sup>	TrpRS
G <sub>37</sub>	yeast	tRNA <sup>Ser</sup>	LeuRS
A <sub>73</sub>	human	tRNA <sup>Leu</sup>	SerRS
G <sub>2</sub> -U <sub>71</sub>	Spirochetes (bacterial taxa) <sup>a</sup>	tRNA <sup>Lys</sup>	LysRS <sup>b</sup>
G <sub>3</sub> -U <sub>70</sub>	yeast	tRNA <sup>Ala</sup>	ThrRS
U <sub>30</sub> -G <sub>40</sub>	yeast	tRNA <sup>Ile</sup>	GlnRS, LysRS
Amino acid			
Asp449, Asp456	<i>E. coli</i>	MetRS	Noncognate anticodons
Glu152	<i>Bacillus stearothermophilus</i>	TyrRS	Noncognate acceptor stems (discriminator N <sub>73</sub> )

Abbreviations and numbering of nucleosides as in Figure 1. <sup>a</sup>See below for identity of canonical tRNA<sup>Lys</sup> species by atypical archaeal class I LysRSs; <sup>b</sup>*E. coli* class II LysRS.

**Table 3 Structural and functional features in synthetases critical for tRNA aminoacylation**

	Class I aaRSs	Class II aaRSs
aaRS		
Amino acid specificity	Arg, Cys, Leu, Ile, Gln, Glu, Met, Trp, Tyr, Val, (Lys) <sup>a</sup>	Ala, Asn, Asp, Gly, His, Lys, Phe, Pro, Ser, Thr, (Pyl, Sep) <sup>b</sup>
Oligomeric structure	Mainly $\alpha$ ( $\alpha_2$ for TyrRS & TrpRS)	Mainly $\alpha_2$ ( $\alpha_2\beta_2$ for most GlyRSs & PheRSs and $\alpha_4$ for AlaRSs)
Ability to charge minimalist RNA structures	IleRS, GlnRS, LeuRS, MetRS, ValRS	AlaRS, AspRS, GlyRS, HisRS, ProRS, SerRS
Catalytic domains		
Architectural organisation	Parallel $\beta$ sheet of 5 strands or Rossmann fold	Antiparallel $\beta$ sheet of 6 strands
Consensus peptidic motifs <sup>c,d</sup>	HiGh	g $\Phi$ xx $\Phi$ xxP (motif 1)
	kmSKs	fRxe (motif 2)
		gxgxfgd/eR (motif 3)
ATP-binding conformation	Extended	Bent
Amino acid attachment site on the last ribose of the tRNA	2' hydroxyl	3' hydroxyl
Anticodon-binding domains	3 subclasses	3 subclasses
Subclasses Ia or Iia	ArgRS, CysRS, IleRS, LeuRS, MetRS, ValRS	GlyRS, HisRS, ProRS, SerRS, ThrRS
Subclasses Ib or Iib	GlnRS, GluRS, ( <i>LysRS1</i> )	AsnRS, AspRS, LysRS
Subclasses Ic or Iic	TrpRS, TyrRS	AlaRS, PheRS, ( <i>PylRS</i> , <i>SepRS</i> )
tRNA approach		
Acceptor arm	Minor groove side	Major groove side
Variable region	Exposed to solvent	Faces the protein
D-loop side	Faces the protein	Exposed to solvent
Editing/Proofreading aaRSs	IleRS, LeuRS, ValRS, MetRS	ProRS, ThrRS, LysRS, AlaRS, PheRS
Allosteric effects revealed		
by Enzymology <sup>e</sup>	GlnRS, TyrRS	AspRS, ProRS
by Structural data <sup>f</sup>	MetRS, TrpRS, TyrRS, GlnRS, TyrRS, LeuRS	AspRS, ProRS, PylRS

<sup>a</sup>A second LysRS, found in a few *Bacteria* and mainly in *Archaea*, belongs to class I

<sup>b</sup>Atypical PylRS and SepRS found in *Archaea* belong to class II aaRSs; PylRS charges pyrrolysine (Pyl) on tRNA<sup>Pyl</sup> and SepRS charges cysteine on tRNA<sup>Cys</sup>.

<sup>c</sup>Single-letter code for amino acids.

<sup>d</sup>Lower case letters for the less conserved residues, with  $\Phi$  for hydrophobic residues and x for any residue.

<sup>e</sup>Including thermodynamic analyses.

<sup>f</sup>Including data from X-ray crystallography and molecular dynamic simulations.

**Table 4 Structural features within tRNA critical for specific aminoacylation by synthetases**

Structural features	Effect <sup>a</sup> on		Comments
	Interaction	Specificity	
<i>Shape of tRNA recognised by an aaRS</i>			
Canonical L-shape	+++	±	*Under special conditions, certain aaRSs recognise all tRNAs, but tRNA mischarging efficiency is low
Atypical shapes	+++	++	*tRNAs with large variable region (e.g. tRNA <sup>Leu</sup> ) or with atypical tertiary interactions (e.g. tRNA <sup>Cys</sup> ) *tRNA mimics in mRNA, e.g. <i>E. coli</i> mRNA <sup>Thr</sup> or viral genomic RNA – Aminoacylation of mimics can be efficient
<i>Determinants in tRNA for specific tRNA aminoacylation ...</i>			
<i>... directly read by an aaRS</i>			
Bases (2–11 residues in identity sets; mainly located in single-stranded regions, sometimes in WC pairs)	+++	+ /++++	*System-specific contacts, for example, in <i>E. coli</i> Asp and Gln systems *Atomic determinants in <i>E. coli</i> tRNA <sup>Ala</sup>
Ribose (from identity bases)	+	+	*Few examples of direct contacts of O2' with aaRSs, for example, in <i>E. coli</i> Asp system *Few examples of ribose as identity determinant (e.g. determined by atomic mutagenesis)
Wobble G–U base pairs	±/+	± /++++	*Wobble pairs are frequent in tRNAs and except for Ala identity are not involved in identity (perhaps marginally by structural effects) *Mutations within the strong G3–U70 Ala identity pair affect mainly <i>k</i> <sub>cat</sub> of alanylation
Modified residues	+	++	*Few documented system-specific examples
<i>... indirectly read by an aaRS</i>			
Sequence-dependent areas (individual or collective)	+ /++	+++	*Recognition of idiosyncratic conformations in cognate tRNAs *Water mediated recognition of individual determinants, e.g. in <i>E. coli</i> Asp system

<i>Antideterminants in tRNA for noncognate tRNA rejection by aaRSs</i>			
Individual nucleotides (canonical or modified)	–	++++	*Not systematically searched: only few cases validated by experiments
Base pairs	–	+++	*Only few examples (see Table 2); can prevent cross-reactions between species as G <sub>2</sub> •U <sub>71</sub> in tRNA <sup>Lys</sup> from <i>Borellia</i> against <i>E. coli</i> LysRS
<i>Nucleotide constituents for stability of a given aaRS:tRNA complex and for tuning specificity</i>			
Bases	+ / +++	± / ++	*System-specific contacts contributing to overall binding affinity, e.g. in <i>E. coli</i> Gln system
Ribophosphate backbones	+ / +++	± / ++	*System-specific contacts as above

<sup>a</sup> Effects: +++, very important; ++, important; +, medium; +weak; ±, low. WC, Watson–Crick; n.d., not determined. See text for details and main references; for other references, see Giegé *et al.* (1998), Giegé and Springer (2016) and Naganuma *et al.*, (2014). Note that the structural features in tRNA are mirrored by proteic elements in the aaRSs (less well defined and characterised than their counterparts on tRNA), notably amino acids that contact identity nts (e.g. in *E. coli* AspRS, Arg76, Gln46 and Glu93 make H-bonds with identity determinants Queuosine<sub>34</sub> and U<sub>35</sub> from the anticodon of cognate tRNA<sup>Asp</sup>). Two of these amino acids (Gln46 and Glu93) are conserved through evolution in AspRSs and therefore can be considered as Asp identity amino acids.