Evolutionary Adjustment of tRNA Identity Rules in Bacillariophyta for Recognition by an Aminoacyl-tRNA Synthetase Adds a Facet to the Origin of Diatoms

Gabor L. Igloi

Received: 27 January 2022 / Accepted: 7 March 2022 / Published online: 24 March 2022
© The Author(s) 2022

Abstract
Error-free protein synthesis relies on the precise recognition by the aminoacyl-tRNA synthetases of their cognate tRNAs in order to attach the corresponding amino acid. A concept of universal tRNA identity elements requires the aminoacyl-tRNA synthetases provided by the genome of an organism to match the identity elements found in the cognate tRNAs in an evolution-independent manner. Identity elements tend to cluster in the tRNA anticodon and acceptor stem regions. However, in the arginine system, in addition to the anticodon, the importance of nucleotide A20 in the tRNA D-loop for cognate enzyme recognition has been a sustained feature for arginyl-tRNA synthetase in archaea, bacteria and in the nuclear-encoded cytosolic form in mammals and plants. However, nuclear-encoded mitochondrial arginyl-tRNA synthetase, which can be distinguished from its cytosolic form by the presence or absence of signature motifs, dispenses with the A20 requirement. An examination of several hundred non-metazoan organisms and their corresponding tRNAArg substrates has confirmed this general concept to a large extent and over numerous phyla. However, some Stramenopiles, and in particular, Diatoms (Bacillariophyta) present a notable exception. Unusually for non-fungal organisms, the nuclear genome encodes tRNAArg isoacceptors with C or U at position 20. In this case one of two nuclear-encoded cytosolic arginyl-tRNA synthetases has evolved to become insensitive to the nature of the D-loop identity element. The other, with a binding pocket that is compatible with tRNAArg-A20 recognition, is targeted to organelles that encode solely such tRNAs.

Keywords Aminoacyl-tRNA synthetase · tRNA · Identity element · Bacillariophyta · Diatoms · Non-metazoan eukaryotic evolution

Introduction
Aminoacylation is the enzymatic process by which amino acids are attached to their specific tRNAs during the first step of protein biosynthesis. It requires a virtually error-free recognition at two levels; the association of one of 20 aminoacyl-tRNA synthetases with its specific amino acid and the simultaneous identification of the cognate tRNA. Heterologous aminoacylation between macromolecules from different species foretold the existence of a common sequence within a given set of tRNAs isoacceptors that is responsible for their specific enzymatic association almost 60 years ago (Loftfield et al. 1968). However, such tRNA identity elements were experimentally verified only some 20 years later (Normanly et al. 1986; McClain and Foss 1988; Sampson et al. 1989; Schulman and Pelka 1989; Normanly and Abelsson 1989). Extensive detailed investigation over many years of bacterial and yeast systems have provided good evidence that, despite some exceptions, for this rather limited selection of organisms a general set of universal rules governing the identity of each synthetase/tRNA pair exists (Giegé et al. 1998; Giegé and Eriani 2021). Further interspecies differences have provided some examples of a more idiosyncratic nature for tRNA recognition (Xue et al. 1993; Nameki et al. 1995; Stehlin et al. 1998). An evolutionary complication to the generalization had, however, arisen upon the emergence of eukaryotes since the synthetase/tRNA pair of the engulfed endosymbiont—the subsequent mitochondrion—possess distinct identity characteristics from the
host (Kumazawa et al. 1989, 1991). Notably, the truncation of metazoan mitochondrial tRNAs and elimination of at least part of the anticipated identity elements provided an adequate minimal set of recognition sites (Ueda and Watanabe 1993; Fender et al. 2012) and opened a whole range of exceptions to the convenient universal identity element concept (Lovato et al. 2001; Yuan et al. 2011; Fender et al. 2012; Igloi and Leisinger 2014; Zeng et al. 2019) leading in some instances even to natural codon reassignments (Ling et al. 2014). In metazoans, the appearance of mitochondria and their encoded tRNAs, therefore, required the retention of both nuclear-encoded cytosolic aminocyl-tRNA synthetase and nuclear-encoded mitochondrial aminocyl-tRNA synthetase with distinct recognition potentials.

Identity elements tend to cluster in the tRNA anticodon and acceptor stem regions (Carter and Wolfenden 2016). However, in the arginine system, the importance of nucleotide A20 in the tRNA D-loop for cognate enzyme recognition was first established both in vivo (McClain and Foss 1988) and in vitro for E.coli (Tamura et al. 1992) and has been a sustained feature also for arginyl-tRNA synthetase for archaea, (Mallick et al. 2005) and the cytosolic form in mammals (Guigou and Mirande 2005) and plants (Aldinger et al. 2012). Its role in binding to the enzyme has been examined in crystallographic detail for Eubacteria (Escherichia coli; Stephen et al. 2018), Archaea (Pyrococcus horikoshii; Konno et al. 2009) and a docking model for Thermus thermophilus has been presented (Shimada et al. 2001). The variable pocket encompassing the tertiary structure of the D and T loops of tRNAARG constitutes an essential feature in its recognition by arginyl-tRNA synthetase. In these three microorganisms, aminocacylation is strictly dependent on the presence of adenosine at position 20 (A20) in tRNAARG. In contrast, Saccharomyces cerevisiae, often taken as a good model for eukaryotic systems, possesses a cytosolic arginyl-tRNA synthetase that is conspicuous in not requiring this A20 for recognition (Sissler et al. 1996) and, indeed, is indifferent to the nature of the base at that position.

Phylogenetic analysis has revealed that the arginyl-tRNA synthetase of S. cerevisiae is the product of an ancestral mitochondrial gene that, after migration to the nucleus and following duplication, has replaced the gene for the host cytosolic form (Karberg et al. 2000; Brindfalk et al. 2007). Therefore, the identity elements documented for the S. cerevisiae arginyl-tRNA synthetase, and in particular, the indifference to the base at position 20 in tRNA, correspond to those found in mitochondrial tRNAs having canonical cloverleaf structures, as in Fungi. An alignment of derived amino acid sequences of arginyl-tRNA synthetases over a wide phylogenetic range has uncovered sequence domains specific to the nuclear-encoded cytosolic and to the nuclear-encoded mitochondrial arginyl-tRNA synthetases, respectively (Igloi 2020a) which one can use to classify the ancestral source of the corresponding gene (Table 1). Using this classification and a knowledge of the corresponding identity elements (Aldinger et al. 2012) it is possible to follow the co-evolutionary development of the macromolecules.

A concept of universal identity elements would require that the enzyme(s) provided by the genome of an organism match the identity elements found in the cognate tRNAs in an evolution-independent manner. In its simplest form it would be expected that cytosolic-type arginyl-tRNA synthetase has a requirement for tRNAARG-A20, irrespective of the cellular compartment encoding the tRNA, as, for example, in plants. In contrast, the mitochondrial-type enzyme could recognize a tRNAARG with any nucleotide at position 20, again in a subcellular-independent manner, as, for example, in Fungi. The validity of such universal identity elements has previously been confirmed (Igloi 2021) for the relatively restricted sample size of amitochondrial organisms (Makiuchi and Nozaki 2014). Whether it also applies to a much more complex system of multi-organelle species, possibly with multiple arginyl-tRNA synthetase genes, has been examined here.

Table 1 Signature sequences used to classify arginyl-tRNA synthetase to their ancestral cytosolic or mitochondrial origin and their associated tRNA identity elements

| Arginyl-tRNA synthetase type | Classification based on: | tRNA Identity Element at D-loop Position 20 |
|-------------------------------|------------------------|-------------------------------------------|
|                               | N-term D-loop binding domain | KMSK catalytic domain                     |
| Cytosolic “GDYQ”               | “FKTR”                  | A                                         |
| Mitochondrial Poorly discernible “GDYQ” | “5∆MSTR” or “MSSR” | U, C, or A                               |

Recognition rules for non-metazoans are poorly understood but have been examined in an Apicomplexan tyrosine system from Plasmodium (Cela et al. 2018). In that organism a minimalistic set of elements compared with the evolutionarily conserved positions was detected. Taking this observation as an indication that non-metazoans might be a source of unconventional identity rules, a screening of arginyl-tRNA synthetases within the phylogenetic groups of non-metazoan/non-fungal eukaryotes was undertaken.
Results

A set of 404 derived protein sequences corresponding to one or more distinct arginyl-tRNA synthetase genes from 264 non-metazoan, non-fungal eukaryote species and covering 32 taxonomic groups in the National Center for Biotechnology Information (NCBI) taxonomic classification were assembled (Igloi 2022). Species possessing a mitosomal organelle have been discussed previously (Igloi 2021) and were not included. Sequences from individual taxonomic groups were visually inspected in order to classify them into the cytosolic or mitochondrial category (Online Resource 1) using the sequence features described previously (Igloi 2020a) (Table 1). For non-metazoans, the elements distinguishing the forms were not as clear-cut as in the case of Eumetazoans. In particular the five amino acid-deletion in the characteristic mitochondrial 5ΔMSTR motif (Igloi 2020a) was not always present. Instead, the cytosolic KFKTR-like motif (corresponding to the conserved Class I aminoacyl-tRNA synthetase KMSK region (Sekine et al. 2001)) was frequently replaced by MSSR, or similar. More reliably, the cytosolic N-terminal GDYQ-motif becomes barely detectable in the mitochondrial form. However, in some cytosolic instances, the GDYQ-sequence was also considerably degraded (Online Resource 2). For questionable placement, an alignment with a clustering of distinct forms of the sequences could provide clarification.

No tRNA sequence data (either cytosolic or organelle) for 61 of these species could be extracted from the databases and could not be used for identity analysis but the protein sequences were nevertheless included in the alignments to facilitate classification. In order to detect robust and systematic evidence for a molecular mechanism of tRNA recognition by alternative or unconventional co-evolution, 931 tRNAArg isoacceptors; (43 isoacceptors from 15 species) have retained the algal nucleotide A20 in all sub-cellular compartments. They may be destined for recognizing all cellular tRNAs (as in higher plants (Duchêne et al. 2005)) or in the case of pairs of different genes in some taxa (e.g. Phaeophyceae), have distinct subcellular targets.

The polyphyletic group of photosynthetic organisms (including Dinophyceae, Cryptophyta, Euglenophyta, Haptophyta, Rhodophyta, Eustigmatophyceae, Pelagophyceae, Phaeophyceae, Xanthophyceae and Chlorophyta) contribute 103 algal species and 136 arginyl-tRNA synthetase sequences. All enzymes are of the cytosolic-type and are associated with tRNAs having A20 in all sub-cellular compartments. They may be destined for recognizing all cellular tRNAs (as in higher plants (Duchêne et al. 2005)) or in the case of pairs of different genes in some taxa (e.g. Phaeophyceae), have distinct subcellular targets.

Of these photosynthetic phyla, Pelagophyceae, Phaeophyceae and Xanthophyceae belong to the clade of Stramenopiles. However, within the Stramenopiles, the phylum Bacillariophyta provides a remarkable and consistent exception to the presumed recognition of universal tRNA identity elements. Bacillariophyta (Diatoms) are represented in the databases by 51 arginyl-tRNA synthetase sequences from 26 species (Table 2) and correspondingly 97 tRNAArg isoacceptors (Igloi 2022), encoded by the nucleus, mitochondrion or plastid, from 18 species. They provide a convincing example of a novel evolutionary divergence from the previously held concept of how arginyl-tRNA synthetase deals with distinct identity elements in its cognate tRNAs.

In all but three species, two genes for the enzyme have been revealed by BLAST searches [the missing genes are likely to be due to gaps in the respective genome assemblies or transcriptome (TSA) databases] (Table 2). Both gene products are clearly of the cytosolic-type, possessing conserved GDYQ and KFKTR motifs (Fig. 1, Panels A and C). The plastids are said to originate from Rhodophyta (Falkowski et al. 2004) and accordingly all encoded tRNAArg isoacceptors (43 isoacceptors from 15 species) have retained the algal nucleotide A20. The encoded mitochondrial tRNAArg, without exception (22 isoacceptors from 11 species) also carry the A20 (Fig. 2A) identity element for recognition by the cytosolic form of the enzyme. However, the available nuclear-encoded tRNAs from 15 species possess C20 and/or U20 (Fig. 2B) without exception (Table 2). This results in the paradox of one form of the cytosolic enzyme needing, atypically, to be indifferent to the nature of the base at position 20 and requires a more detailed examination.

Sequence alignment of the proteins using species which provided data from two genes, resulted in two well-defined phylogenetic clusters each containing one of the gene products which were then designated arbitrarily as Sequence1 or Sequence2 depending on their grouping in the clusters (Fig. 3). In order to determine which cluster was responsible for recognizing A20-containing tRNAArg isoacceptors, attention was focussed on the amino acids that have been defined crystallographically as being involved in the binding of the tRNA variable pocket (including position 20) in yeast (Delagoutte et al. 2000), bacteria (Shimada et al.
Numerous amino acids make up the D-loop binding pocket and these can vary between organisms (Fig. 4). Furthermore, the N-terminal domain which comprises this pocket adopts a different orientation in bacteria compared to yeast (Bi et al. 2014). In E.coli A20 in the D-loop is locked into the enzyme by a mesh of H-bonds involving F36, Q40, A78 and N84 and stacking with F82 (Stephen et al. 2018) (Fig. 4). A32 in E.coli corresponds to P29 in T.thermophilus which forms a van der Waals interaction with A20 (Shimada et al. 2001) and is conserved as T in all Protein2 members (Fig. 1 Panels A and B). Compared to E.coli A78, the analogous N106 of yeast contacts tRNA D20, but mutation N106A in yeast is viable and the kinetic constants of this mutation, as well as F109A and Q111A in yeast arginyl-tRNA synthetase are the same as those on the wild-type (Geslain et al. 2003). This shows that the interaction between the N-terminal domain of S. cerevisiae arginyl-tRNA synthetase and the D-loop of tRNAArg are not important for the specific interaction. Indeed, the arginine-accepting activity was not decreased by the replacement of tRNA C20 by A in an S. cerevisiae tRNAArg^UCU species (Guigou and Mirande 2005). Thus, structural changes in the protein from the strictly A20-binding E.coli-like domain to the yeast-like relaxed binding network can result in indifference to position 20 of the tRNA.

In Bacillariophyta, the Sequence2 representatives were, in general, more E.coli-like than Sequence1 proteins (Fig. 4). Within the conserved FGDYQ motif in cytosolic enzymes, the F corresponding to F36 in E.coli motif, which is in contact with A20, is conserved throughout the Sequence2 cluster and is replaced frequently, but not uniformly by H in Sequence1. On the other hand, E.coli Q^40 of GDYQ, close to A20 (Stephen et al. 2018) is present in all Gene1 as well as in Gene2 products (Fig. 1, Panel A). As noted, position 106 of S. cerevisiae arginyl-tRNA synthetase and the D-loop of tRNAArg^UCU species (Guigou and Mirande 2005). Thus, structural changes in the protein from the strictly A20-binding E.coli-like domain to the yeast-like relaxed binding network can result in indifference to position 20 of the tRNA.

In Bacillariophyta, the Sequence2 representatives were, in general, more E.coli-like than Sequence1 proteins (Fig. 4). Within the conserved FGDYQ motif in cytosolic enzymes, the F corresponding to F36 in E.coli motif, which is in contact with A20, is conserved throughout the Sequence2 cluster and is replaced frequently, but not uniformly by H in Sequence1. On the other hand, E.coli Q^40 of GDYQ, close to A20 (Stephen et al. 2018) is present in all Gene1 as well as in Gene2 products (Fig. 1, Panel A). As noted, position 106 of S. cerevisiae arginyl-tRNA synthetase and the D-loop of tRNAArg^UCU species (Guigou and Mirande 2005). Thus, structural changes in the protein from the strictly A20-binding E.coli-like domain to the yeast-like relaxed binding network can result in indifference to position 20 of the tRNA.

### Table 2: Bacillariophyta species for which arginyl-tRNA synthetase gene products were compiled

| Species                          | Nature of base at position 20 in tRNA |
|----------------------------------|--------------------------------------|
|                                  | Cyto   | Mito | Plastid |
| Asterionella formosa 1, Asterionella formosa 2 | A, C, U | A | A |
| Asterionellopsis glacialis1, Asterionellopsis glacialis2 | C, U | A |
| Chaetoceras neograecalis1, Chaetoceras neograecalis2 | A, U | A |
| Conticribra weissflogii1, Conticribra weissflogii2 | C | A |
| Corethron pennatum1, Corethron pennatum2 | | |
| Coscinodiscus walesii1, Coscinodiscus walesii2 | | |
| Cylindrotheca closterium1, Cylindrotheca closterium2 | | |
| Dactyliosolen fragilissimums1, Dactyliosolen fragilissimums2 | | |
| Ditylum brightwellii1, Ditylum brightwellii2 | C, U | A | A |
| Fistulifera solaris1, Fistulifera solaris2 | C, U | A | A |
| Fragilariopsis cylindrus 1, Fragilariopsis cylindrus 2 | C, U | A | A |
| Guinardia flaccida1, Guinardia flaccida2 | | |
| Leptocylindrus danicus1, Leptocylindrus danicus2 | | |
| Minidiscus sp | | |
| Minutocellus polymorphus1, Minutocellus polymorphus2 | | |
| Navicula sp1, Navicula sp2 | C | A |
| Nitzschia sp1, Nitzschia sp2 | U | A | A |
| Odontella aurita | | |
| Phaeodactylum tricornutum 1, Phaeodactylum tricornutum 2 | U | A | A |
| Pseudo-nitzschia multisiria, Pseudo-nitzschia fraudulenta | U | A | A |
| Skeletonema marinoi1, Skeletonema marinoi2 | U | A | A |
| Stephanopyxis turris1, Stephanopyxis turris2 | | |
| Synedra sp1, Synedra sp2 | | |
| Synedropsis cf. recta | | |
| Thalassiosira pseudonana 1, Thalassiosira pseudonana 2 | C, U | A | A |
| Tryblionella compressa1, Tryblionella compressa2 | | |

Those species, which on the basis of available tRNA data, require a relaxed recognition property, are highlighted in bold. For these, where available, the nature of the base at position 20 in tRNA^Arg is given. Unbolded correspond to species with unavailable tRNA records.

Cyto cytosolic, Mito mitochondrial.
Fig. 1 Alignment of Bacillariophyta arginyl-tRNA synthetase tRNA D-loop binding regions (Panels A and B) and of the conserved “KMSK” catalytic region (Panel C). For orientation, segments from the \textit{E. coli} and \textit{S. cerevisiae} enzymes and the numbering of their critical residues is given. The species order is that given by the CLUSTAL Ω alignment and results in the two distinct clusters. Amino acid identity is shown as Red > 95%; Green 80–95%, Blue 60–80% (Color figure online)

in yeast (\textit{E. coli} 78) is replaced by a small residue (here, A) to accommodate the adenosine ring (Geslain et al. 2003) in Gene2 products, whereas N/Q is retained in Gene1 products. Of the other residues in this domain, \textit{E. coli} F82 which is stacked on A20 and interacts with D20 in yeast is conserved throughout Bacillariophyta. In \textit{E. coli}, N84 would appear to play a central role in A20 recognition. The corresponding amino acid N79 in \textit{T. thermophilus} has been replaced by a selection of different amino acids (Sekine et al. 2001) and revealed that N78D could aminoacylate both tRNA^{ARS}A20 and tRNA^{ARG}G20 (although G20 occurs very rarely, if at all, in natural tRNA^{Arg} isoscipients). Interestingly, the N79Q mutation, which would correspond to the Q111 position in yeast proved to be inactive with all position 20 variants, indicating that the importance of N79 relies on its associated architecture. This crucial D-loop recognition which is guided by N84 (Stephen et al. 2018) is conserved throughout Gene2 products of Bacillariophyta but is highly variable in the other cluster (Fig. 1, Panel B).

Although the domain KMSK that is typical of Class I aminoacyl-tRNA synthetases is not involved tRNA binding, it is vital in catalysis (Sekine et al. 2001) and its alignment (Fig. 1, Panel C) demonstrates its conserved nature. It is worth noting, however, that even within this highly conserved sequence segment, conserved differences between Sequence1 and 2 exist. Examples are V/I at position \textit{E. coli} 367, T/L at 371, R/K at 379 and K/A at 381. Such conserved divergence is found at several sites throughout the protein as in cluster-specific deletions (Panel A). The compilation also shows the nature of the 5ΔMSTR motif found in yeast (position 408) and which is typical of nuclear-encoded mitochondrial arginyl-tRNA synthetases (Igloi 2020a). It is evident that none of the Bacillariophyta enzymes correspond to this motif but, as indicated additionally by the N-terminal GDYQ
feature both Sequences 1 and 2 all fulfill the requirements to be of the cytosolic form. Nuclear-encoded tRNA<sub>Arg</sub> in Bacillariophyta are, without exception within these samples, characterized by having U20 or C20 isoacceptors (Table 2). Conventional identity rules would require these to be recognized by yeast-like, mitochondrial enzymes. However in the absence of this enzyme type, one of these two cytosolic enzymes must, atypically, be insensitive to the nature of the nucleotide at position 20.

From the examination of the D-loop binding domain of bacterial arginyl-tRNA synthetases (above) Protein2 clearly fulfills the requirements for A20 binding, whereas Protein1 lacks some of the essential interacting amino acids (F36, A78, N84 in <i>E. coli</i>). In order to assess the likelihood of proteins of group 2 being involved in protein synthesis involving organelle-encoded tRNA-A20 recognition, their targeting potential was examined with five different prediction algorithms, including DEEPLOC 1.0, which has proved to be robust for mitochondrial proteins in diatoms (Cainzos et al. 2021), (Table 3). Despite the relative uncertainty of predictions and the evident N-terminal extension in the Protein1 group (Igloi 2022), the general tendency is for the Protein2 group to become transported in the organelles (Table 3). Some
exceptions are apparent but these concern sequences whose N termini are either not complete or are ill-defined from genomic assemblies. One can, therefore, with a degree of reliability maintain that the enzymes whose D-loop binding characteristics are in accordance with tRNA-A20 recognition, are targeted to the cellular location harbouring such tRNAs. In contrast group 1 enzymes are retained in the cytosol and have evolved to abstain from using position 20 as a recognition element.

Discussion

Rules governing the specific recognition of tRNAs by their aminoacyl-tRNA synthetases have accumulated over the past decades for all aminoacyl-tRNA synthetase/tRNA pairs. Although some exceptions have been recognized and by no means all taxonomic groups have been investigated, in view of their conservation, it is generally accepted the some key identity elements embedded in the structure of tRNA have been maintained throughout evolution and encompass the Tree of Life (Giegé and Eriani 2021). Nevertheless, exceptions, in particular involving metazoan mitochondrial systems and in the case of some much less-studied taxonomic groups such Apicomplexans have been noted (Giegé and Eriani 2021).

A frequently cited apparent exception that has been known for decades (Benzer and Weisblum 1961; Giegé et al. 1998) is the distinct recognition mechanism by arginyl-tRNA synthetase from *E. coli* and from *S. cerevisiae*. Both enzymes rely on the presence of C35-U/G36 in the anticodon but the former is, in addition, strictly dependent on the major identity element adenosine at position 20 in the tRNA D-loop. The yeast enzyme is characteristically indifferent to the base in the D-loop. However, a frequently overlooked aspect (McShane et al. 2016) is that the yeast cytosolic enzyme [or the enzyme from Fungi, in general (Igloi 2020a)] despite being propagated in the literature, is not comparable to the bacterial or indeed to cytosolic eukaryotic enzymes that were derived from bacteria after the endosymbiotic evolution of mitochondria.

### Table 1

| Amino acid involved in binding of identity element A20 in tRNA^Arg^ in *E. coli* | Corresponding position in *P. horikoshii* / *T. thermophilus* | Corresponding amino acid in aligned Cluster1 of Bacillariophyta | Corresponding amino acid in aligned Cluster2 of Bacillariophyta | Corresponding amino acid in Yeast |
|---|---|---|---|---|
| A32 | P/P | D/N/S/T/Q | T | N62 |
| F36 | I/P | H/F/D/L | F | R66 |
| Q40 | G/G | Q | Q | L70 |
| A78 | V/V | A/N/Q/T | A | N106 |
| F82 | Y/Y | F | F | F109 |
| N84 | N/N | Q/M/L/T/C/I | N | Q111 |

![Diagram](image.png)
Table 3  Organelle targeting prediction for Bacillariophyta arginyl-tRNA synthetase Gene1 and Gene2 products

| Organism                        | Algorithm | Busca | Tagetp | Deeploc | Mulocdeep | Hectar | Number of algorithms giving organelle location |
|---------------------------------|-----------|-------|--------|---------|-----------|--------|-----------------------------------------------|
| Asterionella formosa1           |           | c     | c      | c       | c         | c      | 1                                             |
| Asterionella formosa2           |           | c     | O      | c       | c         | c      | 2                                             |
| Asterionellopsis glacialis1     |           | c     | c      | c       | c         | c      | 1                                             |
| Asterionellopsis glacialis2     |           | O     | O      | O       | O         | O      | 5                                             |
| Chaetoceros neogracilis1        |           | c     | c      | c       | c         | c      | 1                                             |
| Chaetoceros neogracilis2        |           | O     | O      | O       | O         | O      | 5                                             |
| Centricrissa weissflogii1       |           | c     | c      | c       | c         | c      | 1                                             |
| Centricrissa weissflogii2       |           | O     | O      | O       | O         | O      | 5                                             |
| Corethron pinnatum1             |           | c     | c      | c       | c         | c      | 1                                             |
| Corethron pinnatum2             |           | O     | c      | c       | c         | c      | 2                                             |
| Coscinodiscus wailesti1         |           | c     | O      | c       | c         | c      | 1                                             |
| Coscinodiscus wailesti2         |           | O     | O      | O       | O         | O      | 5                                             |
| Cylindrotheca closteriun1       |           | c     | c      | c       | c         | c      | 1                                             |
| Cylindrotheca closteriun2       |           | O     | O      | O       | O         | O      | 5                                             |
| Dactyliosolen fragilissimus1    |           | c     | c      | c       | c         | c      | 1                                             |
| Dactyliosolen fragilissimus2    |           | c     | c      | c       | c         | c      | 0                                             |
| Ditylum brightwellii1           |           | c     | O      | O       | c         | c      | 1                                             |
| Ditylum brightwellii2           |           | c     | O      | O       | O         | O      | 4                                             |
| Fistulifera solariis1           |           | c     | c      | c       | c         | c      | 1                                             |
| Fistulifera solariis2           |           | O     | O      | O       | O         | O      | 5                                             |
| Fragilariopsis cylindrus1       |           | c     | c      | c       | c         | c      | 1                                             |
| Fragilariopsis cylindrus2       |           | O     | O      | O       | O         | O      | 4                                             |
| Guinardia flaccida1             |           | c     | c      | c       | c         | c      | 1                                             |
| Guinardia flaccida2             |           | O     | O      | O       | O         | O      | 5                                             |
| Leptocylindrus danicus1         |           | c     | c      | c       | c         | c      | 1                                             |
| Leptocylindrus danicus2         |           | O     | c      | O       | O         | O      | 4                                             |
| Minatocellus polymorphus1       |           | c     | c      | c       | c         | c      | 1                                             |
| Minatocellus polymorphus2       |           | O     | O      | O       | O         | O      | 5                                             |
| Navicula sp1                    |           | c     | c      | c       | c         | c      | 1                                             |
| Navicula sp2                    |           | O     | O      | O       | O         | O      | 5                                             |
| Nitzschia sp1                   |           | c     | c      | c       | c         | c      | 1                                             |
| Nitzschia sp2                   |           | O     | O      | O       | O         | O      | 5                                             |
| Phaeodactylum tricornutum1      |           | c     | c      | c       | c         | c      | 1                                             |
| Phaeodactylum tricornutum2      |           | O     | O      | O       | O         | O      | 5                                             |
| Pseudo-nitzschia multistriata1 |           | c     | c      | O       | c         | c      | 1                                             |
| Pseudo-nitzschia fraudulenta2   |           | O     | c      | c       | O         | O      | 3                                             |
| Skeletonema marinoi1            |           | c     | c      | O       | O         | c      | 1                                             |
| Skeletonema marinoi2            |           | c     | c      | c       | c         | c      | 0                                             |
| Stephanopysis turris1           |           | c     | c      | c       | c         | c      | 1                                             |
| Stephanopysis turris2           |           | c     | c      | c       | c         | c      | 0                                             |
| Synedra sp1                     |           | c     | c      | c       | c         | c      | 1                                             |
| Synedra sp2                     |           | O     | O      | O       | O         | O      | 5                                             |
| Thalassiosira oceanica1         |           | c     | c      | c       | c         | c      | 1                                             |
| Thalassiosira oceanica2         |           | O     | c      | c       | c         | c      | 1                                             |
| Thalassiosira pseudonana1       |           | c     | c      | c       | c         | c      | 1                                             |
| Thalassiosira pseudonana2       |           | c     | c      | c       | c         | c      | 0                                             |
| Tryblionella compressa1         |           | c     | c      | c       | c         | c      | 1                                             |
| Tryblionella compressa2         |           | O     | O      | O       | O         | O      | 5                                             |

Predictions were performed with the algorithms indicated c cytosol, O organelle

Numbers in brackets refer to different species from the same genus

aDerived from genomic sequence; N terminus not confirmed

bTSA encodes a protein which is N-terminally incomplete
An alignment of several hundred arginyl-tRNA synthetase sequences from numerous taxonomic sources (Igloi 2020b) revealed that one can distinguish between the cytosolic and the mitochondrial enzyme type by characteristic sequence motifs (Igloi 2020a). Using these markers the arginyl-tRNA synthetase of any organism can be matched to the identity elements presented by its cognate tRNA. In the simplest application, an examination of amitochondrial species revealed the single arginyl-tRNA synthetase encoded by the genome always corresponded to the identity elements found in the cytosolic tRNAArg irrespective of whether the enzyme was of the cytosolic or mitochondrial class. It was proposed that in amitochondrial organisms the choice between loss and retention of one type of arginyl-tRNA synthetase depended on the nature of the nucleotide at position 20 of the cognate tRNA (Igloi 2021).

An extension of this analysis to include organisms possessing subcellular genetic compartments in order to determine whether this principle is valid in more complex systems, has now been performed. Attention has been focussed on less well-studied non-metazoans/non-fungi whose subcellular tRNAs are of a canonical structure and would be expected to follow conventional identity rules.

Although the supposition that nucleotides that are conserved at given sites can be equated with identity elements needs to be validated experimentally in each case, the overall trend appeared to substantiate the concept of cytosolic-type arginyl-tRNA synthetases requiring tRNA-A20, whereas mitochondrial-type enzymes were needed to recognize tRNA-N20. Of the 264 species providing sequence data, 168 from 26 phyla were found to match the macrogeneral identity rules. A20 identity element required by the cytosolic arginyl-tRNA synthetase (Igloi and Leisinger 2014). Consistent with the notion that such an altered enzyme was destined for cytosolic protein synthesis, was the finding that the second alignment cluster was predicted to be targeted to the organelles. The proteins in this group have all the structural properties of a tRNA-A20-binding arginyl-tRNA synthetase which is required for the mitochondrial- and plastid-encoded tRNAArg isoacceptors.

An alignment of several hundred arginyl-tRNA synthetase sequences could be recovered by BLAST searches, 21 proved to have two distinct gene products which were all ascertained to originate from ancestral nuclear genes, on the basis of characteristic signature sequences (Igloi 2020a). The two gene products gave rise to two well-defined alignment clusters. This, in itself, would not be a cause for concern since dual targeting has been reported in Diatoms (Gile et al. 2015) as well as in plants (Duchêne et al. 2005) where the second gene product is destined for import into both organelles. However, an examination of the substrate tRNAs encoded in the nuclear genome and in the organelle genomes, showed that the cytosolic tRNAs were not compatible with the hitherto accepted universal identity elements. The major element A20 in the D-loop, which has so far been a consistent feature of cytosolic arginyl-tRNA synthetase recognition, had been replaced by C20 or U20. To allow recognition of such tRNAs the cytosolic-type of arginyl-tRNA synthetase would have required the evolution of a mitochondrial-type of tRNA binding structure, as seen e.g. in S. cerevisiae, whose arginylation activity is insensitive to the nature of the base at position 20 (Sissler et al. 1996). A comparison of the Bacillariophyta enzymes with the crystal structures (Delaguette et al. 2000; Konno et al. 2009; Stephen et al. 2018), which pinpointed the amino acids responsible for forming the D-loop binding pocket, showed that for each Diatom one gene product carried conserved amino acid changes which would be compatible tRNA recognition in the absence of A20. One should, however, also be aware that in Diatom cytosolic tRNAs additional, currently unrecognized, identity elements may have emerged as auxiliary elements, as was the case in metazoan mitochondrial arginyl-tRNA synthetase (Igloi and Leisinger 2014). Consistent with the notion that such an altered enzyme was destined for cytosolic protein synthesis, was the finding that the second alignment cluster was predicted to be targeted to the organelles.

Nevertheless, a striking exception to the presumed universal identity rules was found to be maintained throughout Diatoms (Bacillariophyta). Of the 26 species for which arginyl-tRNA synthetase sequences could be recovered...
horizontal gene transfers in all three genomic compartments (Armbrust et al. 2004; Bowler et al. 2008; Benoiston et al. 2017; Guillory et al. 2018). At which point the acquisition of tRNA^Arg genes coding for entities possessing C20 or U20 nucleotides took place remains a mystery. In modern-day organisms such characteristics in tRNA^Arg are rarely found outside Fungi and Microsporidia so that any concept regarding the origin of Diatoms would have to take the appearance of this facet of the genome into account.

Non-metazoans could be a source of other alternative mechanisms to match recognition elements (Online Resource 2). For example, within the Rhizaria, the phylum Endomyxa, with the data from only three species available, may provide, at least conceptually, another example of how recognition rules could be adapted to accommodate the available tRNAs. All three species have well-defined cytosolic as well as mitochondrial enzymes (Online Resource 1). However, the tRNA^Arg isoacceptors provided by the mitochondrial genome with A20 would require recognition by the cytosolic enzyme, whereas the nuclear tRNAs with U20 or C20 are destined for recognition by the mitochondrial enzyme. In this case differential targeting of the matching enzyme would provide a solution to the recognition problem. Unfortunately, in this case, because of the uncertainty of the N terminus derived from genomic data, no clear-cut targeting prediction could be obtained with the available algorithms.

The coevolution of aminoacyl-tRNA synthetases with their tRNA partners evidently relies not only on the structural plasticity of tRNA molecules (Giegé and Eriani 2021) but also on the adaptation of aminoacyl-tRNA synthetases to the binding of recognition domains present in ancestral tRNAs as has been pointed out for metazoan mitochondrial aminoacyl-tRNA synthetases (Neuenfeldt et al. 2013). This is in line with the hypothesis that emerging aminoacyl-tRNA synthetases adapted to an already established tRNA (De Pouplana et al. 1998) and is consistent with the prediction that tRNAs, as relics of the RNA world (Kühnlein et al. 2021), preceded their synthetases (Nagel and Doolittle 1995).

Methods

For compiling the arginyl-tRNA synthetase collection, genomic (wgs) and transcriptome (tsa) NCBI and other databases were searched using TBLASTN. Accession numbers of the entries containing the corresponding arginyl-tRNA synthetase genes are given in Online Resource 3. Putative full length protein sequences were extracted from genomic hits manually by homology-based alignment FGENESH+ (http://www.softberry.com/) by scanning for protein similarity using the corresponding or closely related organism-specific gene-finding parameters. For non-metazoans the available parameters are somewhat limited making N-terminal predictions, in particular, less reliable in the absence of transcriptome data. Multiple alignments were created with CLUSTALQ (Madeira et al. 2019) and depicted in GENEDOC (Nicholas and Nicholas 1997). Sequences were classified as being of the cytosolic- or mitochondrial-type by visual inspection of the signature regions; GDYQ, KFKT (for cytosolic) and 5∆MSTR (for mitochondrial) (Igloi 2020a). Organelle target prediction was performed online with TARGETP (Emanuelsson et al. 2000), DEEPLOC (Almagro Armenteros et al. 2017), MULOCDEEP (Jiang et al. 2021), HECTOR (Gschloessl et al. 2008) and BUSCA (Savojardo et al. 2018).

For tRNA sequences of non-metazoans neither the badly outdated tRNA database (Jühling et al. 2009) nor the limited genomic tRNA database (Chan and Lowe 2016) was found to be adequate. tRNAs for each organism were therefore recovered from annotated NCBI entries or by BLASTN followed by tRNAscan-SE (Chan et al. 2021). After MUSCLE alignment (Edgar 2004) of tRNA isoacceptors, sequences clustering with annotated mitochondrial or plastid tRNAs were re-examined. Plastid and mitochondrial tRNA isoacceptors form clusters so that one can identify potentially mis-annotated tRNAs. One such example is Chaetoceros neograncilis transcriptome HBTS01037129 which is identical to its annotated plastid genome (MW004650). Where no organelle genome data is available, this clustering approach is a tool to distinguish between organelle-encoded and mis-annotated nuclear-encoded tRNAs.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00239-022-10053-5.

Funding Open Access funding enabled and organized by Projekt DEAL. No funding was received to assist with the preparation of this manuscript.

Declarations

Conflict of interest The author has no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.
References

Aldinger CA, Leisinger A-K, Igloi GL (2012) The influence of identity elements on the aminoacylation of tRNA(Arg) by plant and E.coli arginyl-tRNA synthetases. FEBS J 279:3622–3638. https://doi.org/10.1111/j.1742-4658.2012.08722.x

Almagro Armenteros JJ, Sønderby CK, Sønderby SK et al (2017) DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33:3387–3395. https://doi.org/10.1093/BIOINFORMATICST/BTX431

Armbrust EV, Berges JA, Bowler C et al (2004) The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. Science 306:79–86. https://doi.org/10.1126/SCIENCE.1101156

Benoist AS, Ibarbalz FM, Bittner L et al (2017) The evolution of diatoms and their biogeochemical functions. Philos Trans R Soc B Biol Sci. https://doi.org/10.1098/RSTB.2016.0397

Benzer S, Weisblum B (1961) On the species specificity of acceptor RNA and attachment enzymes. Proc Natl Acad Sci U S A 47:1149–1154. https://doi.org/10.1073/PNAS.47.8.1149

Bi K, Zheng Y, Gao F et al (2014) Crystal structure of E. coli arginyl-tRNA synthetase and ligand binding studies revealed key residues in arginine recognition. Protein Cell 5:151–159. https://doi.org/10.1007/s13328-013-0012-1

Borner J, Burmester T (2017) Parasite infection of public databases: a data mining approach to identify apicomplexan contaminations in animal genome and transcriptome assemblies. BMC Genomics 18:1–12. https://doi.org/10.1186/S12864-017-3504-1/FIGURES/4

Bowler C, Allen AE, Badger JH et al (2008) The phaeodactylumthaliacea. Proc Natl Acad Sci USA 102:16484–16489. https://doi.org/10.1073/pnas.0504682102

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016. https://doi.org/10.1006/jmbi.2000.3903

Falkowski PG, Katz ME, Knoll AH et al (2004) The evolution of modern eukaryotic phytoplankton. Science 305:354–360. https://doi.org/10.1126/SCIENCE.1095964

Fender A, Gaudry A, Jühlung F et al (2012) Adaptation of aminoacylation identity rules to mammalian mitochondria. Biochimie 94:1090–1097. https://doi.org/10.1016/j.biochi.2012.02.030

Geslain R, Bey G, Cavarelli J, Eriani G (2003) Limited set of amino acid residues in a class Ia arginyl-tRNA synthetase is crucial for tRNA binding. Biochemistry 42:15092–15101. https://doi.org/10.1021/bi035581u

Giegé R, Eriani G (2021) Transfer RNA recognition and aminoacylation by synthetases. eLS, Wiley, Hoboken, pp 1007–1030

Giegé R, Sissler M, Florentz C et al (1998) Universal rules and idiosyncratic features in tRNA identity. Nucl Acids Res 26:5017–5035. https://doi.org/10.1093/nar/26.22.5017

Gile GH, Moog D, Slamovits CH et al (2015) Dual organellar targeting of aminoacyl-tRNA synthetases in diatoms and cryptophytes. Genome Biol Evol 7:1728–1742. https://doi.org/10.1093/GBE/EVV095

Gschloessl B, Guermeur Y, Cock JM (2008) HECTAR: a method to predict subcellular targeting in heterokonts. BMC Bioinform 9:1–13. https://doi.org/10.1186/1471-2105-9-393/FIGURES/4

Guiou L, Miranda M (2005) Determinants in tRNA for activation of arginyl-tRNA synthetase: evidence that tRNA flexibility is required for the induced-fit mechanism. Biochemistry 44:16540–16548. https://doi.org/10.1021/bi051575h

Guillory WX, Onyshchenko A, Ruck EC et al (2018) Recurrent loss, horizontal transfer, and the obscure origins of mitochondrial introns in diatoms (bacillariophyta). Genome Biol Evol 10:1504–1515. https://doi.org/10.1093/gbe/evy103

Huson DH, Scombravacca C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol 61:1061–1067. https://doi.org/10.1093/sysbio/sys062

Igloi GL (2020a) Molecular evidence for the evolution of the eukaryotic mitochondrial arginyl-tRNA synthetase from the prokaryotic suborder cystobacterineae. FEBS Lett 594:951–957. https://doi.org/10.1002/1742-4683.13665

Igloi GL (2020b) Gene organization and phylum-specific attributes of eukaryotic arginyl-tRNA synthetases. Gene Reports 20:100778. https://doi.org/10.1016/j.genrep.2020.100778

Igloi GL (2021) The evolutionary fate of mitochondrial aminoacyl-tRNA synthetases in amitochondrial organisms. J Mol Evol 89:484–493. https://doi.org/10.1007/s00239-021-10019-z

Igloi GL (2022) Compilation and Alignment of Eukaryotic Arginyl-tRNA Synthetases. In: Mendeley Data. https://data.mendeley.com/datasets/ls4jibw9nft.3. Accessed 28 Jan 2022

Igloi GL, Leisinger A-K (2014) Identity elements for the aminoacylation of metazoan mitochondrial tRNA Arg have been widely conserved throughout evolution and ensure the fidelity of the AGR codon reassignment. RNA Biol 11:1313–1323. https://doi.org/10.1080/15476286.2014.996094

Jiang Y, Wang D, Yao Y et al (2021) MULocDeep: a deep-learning framework for protein subcellular and suborganellar localization prediction with residue-level interpretation. Comput Struct Bio- technol J 19:4825–4839. https://doi.org/10.1016/j.csbj.2021.08.027

Jühlung F, Mörl M, Hartmann RK et al (2009) tRNAseq 2009: compilation of tRNA sequences and tRNA genes. Nucl Acids Res 37:D159–D162. https://doi.org/10.1093/nar/gkn772
Karlberg O, Canbäck B, Kurland CG, Andersson SGE (2000) The dual origin of the yeast mitochondrial proteome. Yeast 1:170–187. https://doi.org/10.1111/j.1049-323X.2000.957406

Konno M, Sumida T, Uchikawa E et al (2009) Modeling of tRNA-assisted mechanism of Arg activation based on a structure of Arg-tRNA synthetase, tRNA, and an ATP analog (ANP). FEBS J 276:4763–4779. https://doi.org/10.1111/j.1742-4658.2009.01718.x

Kühnlein A, Lanzmich SA, Braun D (2021) tRNA sequences can assemble into a replicator. Elife. https://doi.org/10.7554/ELIFE.63431

Kumazawa Y, Yokogawa T, Hasegawa E et al (1989) The ami- tocidal activity of a mitochondrial phenylalanine tRNA from mitochondria and various nonmitochondrial sources by bovine mitochondrial phenylalanyl-tRNA synthetase. J Biol Chem 264:13005–13011

Kumazawa Y, Himeno H, Miura K, Watanabe K (1991) Unilateral aminoacylation specificity between bovine mitochondria and eubacteria. J Biochem 109:421–427

Ling J, Daoud R, Lajoie MJ et al (2014) Natural reassignment of CUU and CUA sense codons to alanine in ashyba mitochondria. Nucleic Acids Res 42:499–508. https://doi.org/10.1093/nar/gkt842

Loffeld RB, Eigner EA, Nobel J (1968) Inter-phylogenetic specificity in the bonding of amino acids to tRNA. Biol Bull 135:181–192. https://doi.org/10.2307/1539627

Lovato MA, Chihade JW, Schimmel P (2001) Translocation within the acceptor helix of a major tRNA identity determinant. EMBO J 20:4846–4853. https://doi.org/10.1093/emboj/20.17.4846

Madeira F, Park YM, Lee J et al (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641. https://doi.org/10.1093/nar/gkz268

Makituichi T, Nozaki T (2014) Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa. Biochimie 100:3–17. https://doi.org/10.1016/j.biochi.2013.11.018

Mallick B, Chakrabarti J, Sahoo S et al (2005) Identity elements of archaean tRNA. DNA Res 12:235–246. https://doi.org/10.1093/dnares/dsi008

McClain WH, Foss K (1988) Changing the acceptor identity of a transfer RNA by altering nucleotides in a “variable pocket.” Science 241:1804–1807

McShane A, Hok E, Tomberlin J et al (2016) The enzymatic paradox of yeast ArgInyRNA synthetase: exclusive arginine transfer controlled by a flexible mechanism of tRNA recognition. PLoS ONE 11:1–14. https://doi.org/10.1371/journal.pone.0148460

Nagel GM, Doolittle RF (1995) Phylogenetic analysis of the aminoa- cyl-tRNA synthetases. J Mol Biol 264:13005–13011

Nameki N, Asahara H, Tamura K et al (1995) Similarities and differences in tRNA identity between Escherichia coli and Saccharomy- ces cerevisiae: evolutionary conservation and divergence. Nucleic Acids Symp Ser 34:205–206

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 Res 41:2698–2708. https://doi.org/10.1093/nar/gks1322

Nicholas KB, Nicholas HBJ (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by author Norman J, Abelson J (1989) tRNA Identity. Annu Rev Biochem 58:1029–1049

Nomura J, Ogden RC, Horvath SJ, Abelson J (1986) Changing the identity of a transfer RNA. Nature 321:213–219. https://doi.org/10.1038/321213a0

Sampson JR, DiRenzo AB, Behlen LS, Uhlenbeck OC (1989) Nucleo- tides in yeast tRNAphe required for the specific recognition by its cognate synthetase. Science 243:1363–1366. https://doi.org/10.1126/science.2646717

Savojardo C, Martelli PL, Fariselli P et al (2018) BUSCA: an inte- grative web server to predict subcellular localization of proteins. Nucleic Acids Res 46:W459–W466. https://doi.org/10.1093/nar/gky320

Schulman L, Pelka H (1989) The anticodon contains a major element of the identity of arginine transfer RNAs. Science 246:1595–1597. https://doi.org/10.1126/science.2688091

Sekine SL, Shimada A, Nureki O et al (2001) Crucial role of the high-loop lysine for the catalytic activity of arginyl-tRNA synthetase. J Biol Chem 276:3723–3726. https://doi.org/10.1074/jbc.C0007 56200

Shimada A, Nureki O, Goto M et al (2001) Structural and mutational studies of the recognition of the arginine tRNA-specific major identity element, A20, by arginyl-tRNA synthetase. Proc Natl Acad USA 98:13357–13354. https://doi.org/10.1073/pnas. 231267998

Sims PA, Mann DG, Medlin LK (2006) Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia 45:361–402. https://doi.org/10.2216/05-22.1

Sissler M, Giege R, Florentz C (1996) Arginine aminoacylation identity is context-dependent and ensured by alternate recognition sets in the anticodon loop of accepting tRNA transcripts. EMBO J 15:5069–5076

Srinivis M, Horn C, Brown M et al (1998) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 26:148–153. https://doi.org/10.1093/NAR/26.1.148

Stehlin C, Burke B, Yang F et al (1998) Species-specific differences in the operational RNA code for aminoacylation of tRNAPro. Biochemistry 37:8605–8613. https://doi.org/10.1021/bi980364s

Stephen P, Ye S, Zhou M et al (2018) Structure of Escherichia coli arginyl-tRNA synthetase in complex with tRNAArg: pivotal role of the D-loop. J Mol Biol 430:1590–1606. https://doi.org/10. 1016/j.jmb.2018.04.011

Tamura K, Himeno H, Asahara H et al (1992) In vitro study of E.coli tRNA(Arg) and tRNA(Lys) identity elements. Nucl Acids Res 20:2335–2339

Ueda T, Watanabe K (1993) The evolutionary change of the genetic code as restricted by the anticodon and identity of transfer RNA. Orig Life Evol Biosph 23:345–364. https://doi.org/10.1007/BF015 82085

Xue H, Shen W, Giegé R, Wong JT (1993) Identity elements of tRNA(Trp). Identification and evolutionary conservation. J Biol Chem 268:9316–9322

Yuan J, Gogakos T, Babina AM et al (2011) Change of tRNA identity leads to a divergent orthogonal histidyl-tRNA synthetase/tRNA- Hispair. Nucleic Acids Res 39:2286–2293. https://doi.org/10. 1093/nar/gkq1176

Zheng Q-Y, Peng G-X, Li G et al (2019) The G3–U70-independent tRNA recognition by human mitochondrial alanyl-tRNA synthetase. Nucleic Acids Res 47:3072–3085. https://doi.org/10. 1093/nar/gkz078