



# An integrated, structure- and energy-based view of the genetic code

Henri Grosjean, Eric Westhof

## ► To cite this version:

Henri Grosjean, Eric Westhof. An integrated, structure- and energy-based view of the genetic code. Nucleic Acids Research, 2016, 44 (17), pp.8020 - 8040. 10.1093/nar/gkw608 . hal-01412326

**HAL Id: hal-01412326**

**<https://hal.science/hal-01412326>**

Submitted on 8 Dec 2016

**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.

## SURVEY AND SUMMARY

# An integrated, structure- and energy-based view of the genetic code

Henri Grosjean<sup>1,\*</sup> and Eric Westhof<sup>2,\*</sup>

<sup>1</sup>Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette, France and <sup>2</sup>Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de biologie moléculaire et cellulaire du CNRS, 15 rue René Descartes, 67084 Strasbourg, France

Received May 08, 2016; Revised June 11, 2016; Accepted June 17, 2016

### ABSTRACT

The principles of mRNA decoding are conserved among all extant life forms. We present an integrative view of all the interaction networks between mRNA, tRNA and rRNA: the intrinsic stability of codon–anticodon duplex, the conformation of the anticodon hairpin, the presence of modified nucleotides, the occurrence of non-Watson–Crick pairs in the codon–anticodon helix and the interactions with bases of rRNA at the A-site decoding site. We derive a more information-rich, alternative representation of the genetic code, that is circular with an unsymmetrical distribution of codons leading to a clear segregation between GC-rich 4-codon boxes and AU-rich 2:2-codon and 3:1-codon boxes. All tRNA sequence variations can be visualized, within an internal structural and energy framework, for each organism, and each anticodon of the sense codons. The multiplicity and complexity of nucleotide modifications at positions 34 and 37 of the anticodon loop segregate meaningfully, and correlate well with the necessity to stabilize AU-rich codon–anticodon pairs and to avoid miscoding in split codon boxes. The evolution and expansion of the genetic code is viewed as being originally based on GC content with progressive introduction of A/U together with tRNA modifications. The representation we present should help the engineering of the genetic code to include non-natural amino acids.

### INTRODUCTION

The genetic code is considered as ‘quasi’ universal. Although deviations of the ‘standard’ genetic code have been

identified, they occur in only a few organisms and mainly in organelles (1–3). However, as more organisms are being presently studied, new exceptions will certainly be discovered. Nevertheless, because all extant organisms evolved from a small set of ancestral living organisms that, along evolution, continued to exchange DNA information, the basic molecular principles of decoding (including the genetic code) remained essentially the same in all organisms of the three domains of life (Bacteria, Eukarya and Archaea). These principles are anchored in the physics and chemistry of the interplay between the diverse, and often locally neutral, non-covalent and weak molecular interactions engaged, and in the thermodynamics of formation of the several distinct complexes present at the various stages of the multistep translational process. Far from excluding profound improvements, these principles allow for adaptation and even innovations of many of the elements of the modern translation machinery (including the genetic code) according to physiological needs and/or to cellular niches and environments.

The traditional way to represent the 64 codons, as organized in the standard codon table (4–7) took several years to settle down. In the first row of the codon table are indicated all codons starting with a U at the first codon position, with a C in the second row, and a A or a G for the third and fourth rows, respectively. For each column, the same (base) order is used for the second codon base, as well as within each of the 16 decoding boxes for the classification of the 4 codons ending with a different base. This standard and quasi-universal codon table is divided into boxes with 4 synonymous codons (hereafter designated as 4-codon boxes or unsplit decoding boxes) where the third base can be either U, C, A or G for coding the same amino acid, and boxes with 2 synonymous codons (hereafter designated as 2-codon boxes or split decoding boxes) where the third base codes for a different amino acid depending on its nature, either pyrimidine

\*To whom correspondence should be addressed. Email: E.Westhof@jbmc-cnrs.unistra.fr  
Correspondence may also be addressed to Henri Grosjean. Email: henri4g@me.com

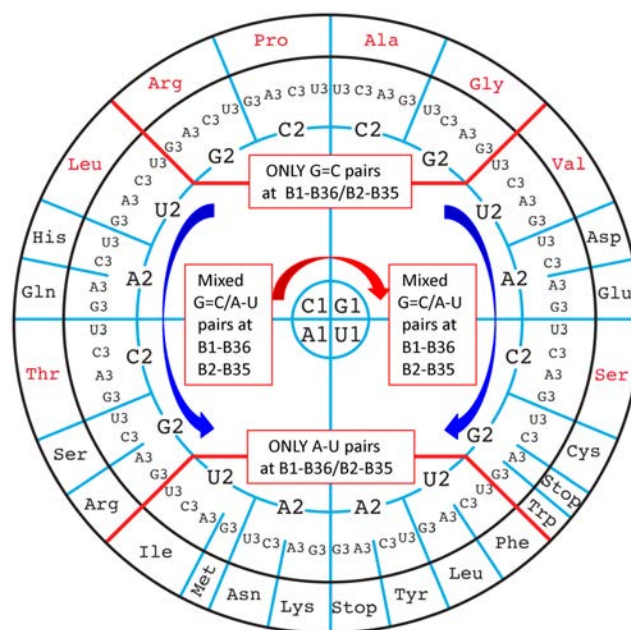
(Y) or purine (R). A third category concerns split 3:1-codon boxes where a single codon ending with guanine means methionine or tryptophan (1-codon box). Within these split 2:2- and 3:1-codon boxes are those meaning STOP for opal UGA, ochre UAA and amber UAG. At the time the table was first proposed, knowledge about the detailed molecular mechanisms underlying the decoding process did not exist. A few research groups argued for a different organization of the initial and now standard codon table (8–11). Arguments were based on various criteria or regularities of the code but never with knowledge of all the elements of the whole decoding process, in particular the large diversity of post-transcriptional nucleotide modifications that exist especially in the tRNA.

Recent progress in identifying modified nucleotides and their functions in tRNA (12–14), and information gained from the recent detailed structural, physicochemical and kinetic studies of ribosomes associated with mRNA/aminoacyl-tRNA, has made clear that selection of a given codon by its cognate aminoacyl-tRNA harboring a complementary or nearly complementary anticodon within the ribosomal decoding center is a complex and intricate process (15–22). The ribosome is not a passive machine but an active component (structurally and kinetically speaking) of the codon selection mechanism. Also, the original wobble ‘hypothesis’ (23) has to be broadened in several cases involving non-standard base pairings at the third base pair of the codon/anticodon helix (see for examples: (24–26)). Indeed, it is now becoming clear that some base–base oppositions avoid usual pairings or even H-bonding while still being accommodated within a Watson–Crick-like helix. Indeed, post-transcriptional modifications of nucleotide-34 can change the physicochemical behavior of the base (frequency of tautomerism) and/or the spatial preference of the nucleotide (*syn/anti*, puckering of the ribose) to precisely allow the third anticodon nucleotide to fit within a mini-helix structure together with the two other base-pairs of the anticodon (15,27–32). In short, wobbling at the third position does occur in various ways, but without movement of the codon base toward the minor groove (see below).

In the present work, we took into consideration all the new elements that we now understand are important for the aminoacyl-tRNA selection at the A site of the decoding center of the ribosome in order to propose a visually useful decoding table embedding multiple structural aspects of translation. This new representation allows one to propose a rationale for the emergence of the contemporary G/C/A/U containing quasi-universal genetic code from an ancient GC-rich code and the multiple constraints on such an evolution.

## THE CODON TABLE

We started from an observation, already made at least partly by others (8,9,33,34), on the distribution of codons within the codon table. The codons within the unsplit 4-codon boxes always have one C in the second or either C or G at the first position of codon leading to C = G (G = C) pair in the codon/anticodon triplets. Among these codon boxes, the two conditions are simultaneously satisfied for four amino acids, Ala, Arg, Gly and Pro. In contrast, codons of the split



**Figure 1.** Circular representation of the genetic code emphasizing the inherent regularities of the decoding recognition process. The codons containing solely G = C pairs at the first two positions are at the top, those containing solely A = U pairs at the bottom, and those with mixed pairs of G = C and A = U either at the first or second pair of the codon/anticodon helix in the middle at the right and left. Thick red lines separate the three main regions. The red arrow indicates the direction of rotation for C1, G1, U1, A1 and the blue arrows the direction of rotation for C2, G2, U2, A2 on the right and left parts of the wheel. The amino acids coded by unsplit 4-codon boxes are indicated in red and those by split 2:2- and 3:1-codon boxes, together with the usual stop codons, are indicated in black. Throughout, the codon positions are numbered B1–B2–B3 and the anticodon nucleotides B34–B35–B36, both from 5' to 3'.

2:2- and 3:1-codon boxes have, instead, A at the second or either A or U at the first position of the codon leading to a A–U (U–A) pair in the codon/anticodon triplets. Among these, the two conditions are simultaneously satisfied for seven amino acids, Asn, Ile, Leu, Lys, Met, Phe and Tyr. The nine remaining amino acids (but 12 codon boxes because Leu, Ser and Arg each have 2 distinct codon boxes) have a mixture of C = G and U–A pairs, with either a C = G pair at the first position followed by a U–A pair or a U–A pair at the first position followed by a C = G pair. The two additional amino acids that are co-translationally inserted, pyrrolysine and selenocysteine, are recoded to replace stop codons, and will be discussed later in our analysis.

In our new representation, we arrange each codon corresponding to the 20 canonical amino acids on a circle (genetic code wheel) as a function of the nature of the first two base pairs of the codon/anticodon triplet. On the circle, the G = C only codons are disposed at the top, the A–U only codons at the bottom, and the mixed ones in-between on the left and the right. The order of third codon base has been chosen so that proximities of purine- and pyrimidine-ending codons are preserved (Figure 1). Note how the bases C, G, U, A rotate around the circle: for the first position the C1 starts in the top left quadrant, for the second position C2, G2, A2, U2 rotates down in the right part in a right-handed fashion and in the left part in a left-handed fashion.

The choice of this unsymmetrical disposition will become more apparent and coherent in the following sections.

We then computed the energies of the codon/anticodon minihelices (5'-B1-B2-B3-3' codon paired to 3'-B36-B35-B34-5' anticodon) following the recent Turner values for dimers (35). First, we considered only the first two base pairs, B1-B2 paired to B36-B35 (Figure 2, box A). The values (indicated as bold red numbers) are distributed on a circular view of the codons. Clearly, from those values, three groups of amino acids can be distinguished. A STRONG (S) group (average  $-3.1$  kcal/mole) with, as above, the four Ala, Arg, Gly, Pro. A WEAK (W) group (average  $-1.0$  kcal/mole) with, again as above, the seven Asn, Ile, Leu, Lys, Met, Phe, Tyr. And finally, an INTERMEDIATE (I) group (average  $-2.2$  kcal/mole) gathers the remaining amino acids (or codon boxes). A similar distribution is observed by computing the triplets of base pairs (B1-B2-B3 with B36-B35-B34) (Figure 2, box B), the corresponding values are in parentheses). Indeed, the distribution from STRONG to WEAK is respected (averages from  $-5.8$ ,  $-4.3$ ,  $-2.7$  kcal/mole). Finally, in order to be slightly closer to biological reality, we considered the U3oG34 and G3oU34 Crick-type of wobbles at the third base pair (Figure 2, boxes C and D, respectively). Again, the three groups segregate well with energies, respectively, of  $-5.0$ ,  $-3.5$  and  $-1.8$  kcal/mole for the U3oG34 wobble and  $-4.6$ ,  $-3.0$ , and  $-1.4$  kcal/mole for the G3oU34 wobble, respectively. Interestingly, the G3oU34 wobble is always less stable than the U3oG34 wobble interactions (see discussion below). There is no overlap between the three categories, in the sense that no INTERMEDIATE codon has a stronger energy than any STRONG codon or a weaker one than any WEAK codon.

Despite the large differences of energy between GC-rich and AU-rich codon–anticodon duplexes, the measured binding constants of natural aminoacyl-tRNAs on 70S ribosomes programmed with mRNA are unexpectedly remarkably similar (20–22,36,37). Likewise the stability of complexes formed between two natural tRNAs harboring complementary anticodons is very much the same whatever the GC/AU composition of the duplex (38,39). Only when non-natural mutant tRNAs are used are large differences of binding constants measured. Obviously, both the ribosome and the built-in features of natural tRNAs have evolved to allow minimization of the differential thermodynamic contributions of the codon–anticodon interactions, allowing tRNA–mRNA interactions to be rather uniform; in other words, to be neither too sticky nor too loose during the process of translation. We will now describe these features, first on the ribosome and then on the transfer RNAs.

## THE RIBOSOMAL DECODING SITE AND THE ANTICODON TRIPLET

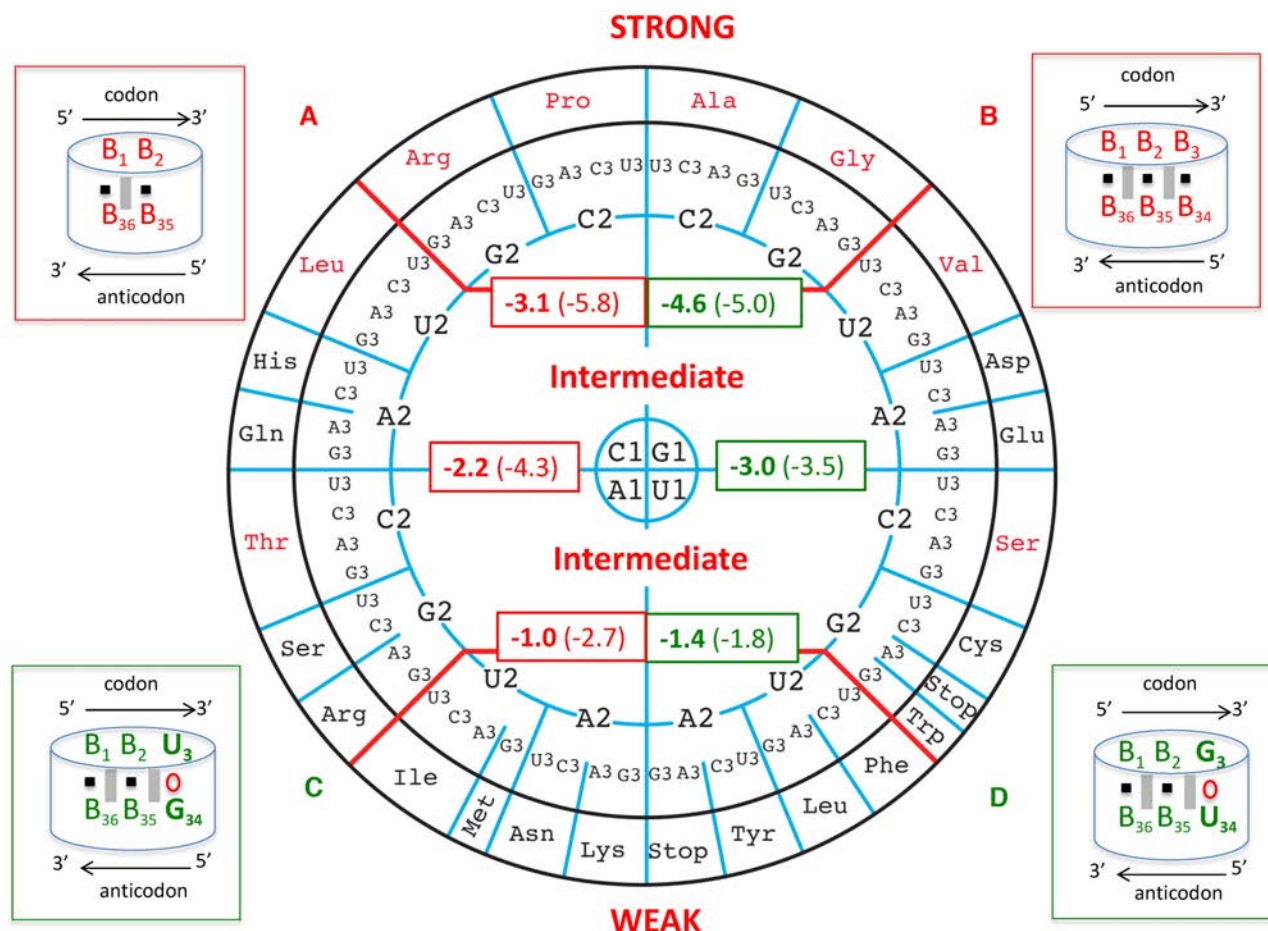
Recent crystallographic data of aminoacyl-tRNAs associated to mRNA programmed ribosomes revealed the contacts made by the ribosomal A-site with the codon/anticodon helix, designated as the ribosomal grip. The two invariant adenine residues A1492 and A1493 of helix 44 of 16S rRNA make A-minor type interactions in the minor groove of the first two base pairs of the

codon/anticodon helix (Figure 3) (15–17,27–29,31,32,40–43). Except for those made to the B1-B36 pair, such contacts are almost identical for all types of Watson–Crick pairs and, therefore, can be considered to contribute similarly to the global interaction energies (44) (Table 1). On the circular representation (Figure 3), the A-minor contacts are equally partitioned (listed in plain red boxes), except, importantly, for those between A1493 and B1-B36 (G36 = C1 or G1 = C36, in dashed red boxes), which are specific to the top half of the circle.

In Table 2 and Figure 3 are indicated the constant and variable contacts that occur within the anticodon loop and with the codon/anticodon minihelix (45) (see also Supplementary Figure S1A and B). The contributions of such contacts to aminoacyl-tRNA (aa-tRNA) binding to the ribosome have been established by various methods (46–48). Recent experiments on the effects of 2'-O-methyl group in the ribose of mRNA show the strongest reduction in translation efficiency when the 2'-O-methyl group is present at the second position, followed by methylation at the third position (25% of unmodified) and the first position (80% of unmodified), confirming further the O2'-hydroxyl group donor contacts at the second and third positions (49). They involve mainly the ribose O2' of U33 (an almost invariant nucleotide) with the Hoogsteen edge of B35 (where the contact U33(O2')...R35(N7) is stronger than the U33(O2')...Y35(C5) contact) (see Tables 1 and 2 and Supplementary Figure S2). Of note is that these contacts therefore slightly disfavor the codon/anticodon interactions with a purine at B2 (Arg, Gly, His/Gln, Asp/Glu, Arg/ser, Cys/Trp) that are circled in green in Figure 3 (discussed below). In short, the ribosomal grip interactions add to the stabilization of mainly codons of the unsplit 4-codon boxes, especially through the A1493(N3) H-bond to the first G = C/C = G pair in the upper part of the circle in Figure 3.

In addition, the crystal structures of ribosomal complexes with mRNA and aminoacyl-tRNAs reveal many contacts between a few amino acids of r-proteins and the RNA sugar–phosphate backbones. The majority of these contacts should be general and applicable to any tRNA and mRNA and, thus, should not make a major contribution to the specificity and fidelity of a specific codon–anticodon binding, at least in the decoding A-site of the ribosome. However, r-proteins S4, S5 and S12 have been demonstrated to be involved in the ribosomal grip by controlling the closure of 30S subunit upon correct codon recognition (50–52). The protein S12 especially closely contacts the decoding site with Ser46 binding to the Hoogsteen edge of A1492 (which monitors the second base pair B35-B2) and Pro44 interacting via a solvent molecule with the hydroxyl group of the ribose of B3, together with G530 and C518 (17,53). Among the many conserved and semi-conserved residues present in the rRNA region of the decoding center, a few are post-transcriptionally modified (54,55). However, none of these modified residues is in direct contact with the codon or anticodon in the A-site of the ribosome and therefore should not contribute directly to specific codon selection mechanism. The main functions of these post-transcriptional modifications appear to reside in the modulation of rRNA structure and ligand interactions leading to indirect and





**Figure 2.** On the circular representation for the code are given the Turner energies calculated for the first two base pairs of the codon/anticodon helix, B1-B2 paired with B36-B35. Most generally, 5'-B1-B2-B3-3' pairs to 5'-B34-B35-B36-3', forming pairs B1-B36, B2-B35 and B3-B34. The free energy values for dimers are in kcal/mole and taken from (35). In the scheme (A) at the top left, the black square symbolizes the H-bonds between A-U or G-C and the grey vertical rectangles the stacking interactions between the base pairs. The average values are given in the red rectangles within the codon wheel: it is -3.1 kcal/mole for the top codon boxes, -2.2 kcal/mole for the middle codon boxes and -1.0 kcal/mole for the bottom ones, with therefore 1.0 kcal/mole difference between the three groups. The reported errors on the free energies are less than 0.1 kcal/mole. The top four groups of un-split family codon boxes will be called 'STRONG' codon boxes, the middle ones encompassing split and un-split codon boxes are designated 'INTERMEDIATE' codon boxes, and the last all split family codon boxes at the bottom as 'WEAK' codon boxes. The values in parentheses represent the same calculations with the consideration for the third base pair assumed to be Watson-Crick and without nucleotide modifications (schematized in (B) at the upper right corner). Although such calculations are not biologically meaningful, they may be relevant when considering the evolution of the decoding system. The differences between the averaged values are of the order of -1.5 kcal/mole and around 3.0 kcal/mole between the STRONG and WEAK codon boxes. The same calculations were done with the third base pair either a U3oG34 or a G3oU34 (schemes (C) and (D) at bottom left and right corners respectively). Note that with B2-B35 being either a G=C or C=G pair, the energies are 2.0 kcal/mole (depending on GoU34 or UoG34, respectively) more stable than with B2-B35 being a A-U or U-A pair. Also, whatever the nature of the B2-B35 pair, G3oU34 is always less stable than U3oG34 by about half a kcal/mole. Correct decoding depends on formation of a short double helix-like structure between the three bases of the codon and the anticodon (symbolized by the dashed red cylinder).

**Table 1.** Regularities in the codon table and the ribosomal grip

**4-codons box:** always either a C = G pair at 2nd position or a C = G/G = C pair at 1st position (Note that the two states occur in Pro, Ala, Arg, Gly).

A1492 ... C2 = G35

A1493 ... C1 = G36/G36 = C1

**2-codons box:** always either a A-U pair at 2nd position or a U-A/A-U pair at 1st position (Note that the two states occur in Phe, Leu, Ile, Met, Asn, Lys, Tyr).

A1492 ... A2-U35

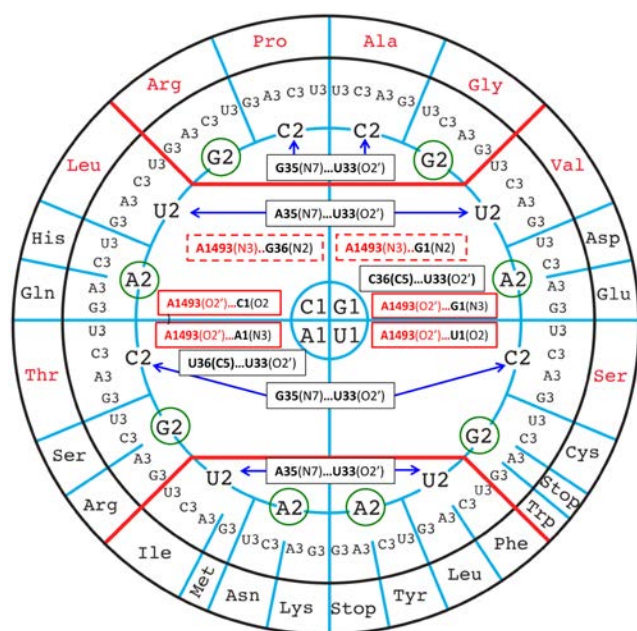
A1493 ... U1-A36/U36-A1

Throughout the codon positions are numbered 1-2-3 and the anticodon nucleotides 34-35-36, both from 5' to 3'. Most generally, they are 5'-B1-B2-B3-3' pairing to 5'-B34-B35-B36-3'.

**Table 2.** Constant and variable or sequence dependent interactions present between the codon/anticodon minihelix and the ribosomal grip or the anticodon loop

Constant interactions with rRNA	Variable interactions with rRNA
<b>B1-B36</b> A1493(N1)... B36(O2') A1493(O2')... B1(O2')	<b>B1-B36</b> A1493(N3)... G36(N2) A1493(N3)... G1(N2) A1493(O2')... Y1(O2) or A1493(O2')... R1(N3)
<b>B2-B35</b> G530(N3)... B35(O2') G530(O2')... B35(O4') G530(N1)... A1492(N1) A1492(N3)... B2(O2') A1492(O2')... B2(O2')*	<b>B2-B35</b> None observed yet.
<b>B3-B34</b> B3(O2')... M... G530/C518/S12 B34 sugar stacking on C1054	
<b>Constant interactions with tRNA</b>	<b>Variable interactions with tRNA</b>
<b>B1-B36</b> U33(N3)... B36(OR)	<b>B1-B36</b> U33(O2)... Y36(C5)
<b>B2-B35</b> U33 Stacking on B35(OR)	<b>B2-B35</b> U33(O2')... R35(N7) or weaker U33(O2')... Y35(C5)

Notice that no variable interaction occurs between B2-B35 and the conserved rRNA nucleotides while no interaction occurs between B1-B36 and the other tRNA nucleotides. The contact indicated by \* is present in usual A-minor interactions, but presents a long distance in structures for the A-site (3.6 Å or longer). Note also that in standard A-minor contacts, the equivalent C1 = G36 is observed more often than G1 = C36 and C1 = G36 more often than G1 = C36 (44).



**Figure 3.** The experimentally identified interactions that are variable between the codon/anticodon base pairs, the ribosomal grip and the anticodon loop nucleotides are shown on the circular representation of the codons. The constant contacts are not shown. The weak contacts involving C–H bonds are also not shown. The rRNA nucleotides are in red and the anticodon loop nucleotides in black. In plain black boxes are the tRNA intra-anticodon loop interactions, in plain red boxes are the interactions of the ribosomal grip with, in dashed red boxes, the contact occurring only in the top half of the wheel. Purine bases at position 2 of codons are circled in green to emphasize the fact that they do not contact the corresponding tRNAs.

general effects on translational fidelity (56). One exception could be the positively charged m<sup>7</sup>G527 in bacteria. It is lo-

cated about 5 Å from Pro44 of S12 and the conserved base G530, possibly creating an electrostatic environment that influences the electrostatic state of modified B34, especially with hypermodification at U34.

In summary, the present organization of the genetic code makes a clear energetic segregation between all GC-rich codons belonging to 4-codon family boxes and all AU-rich codons belonging to 2:2 and 3:1 codon family boxes, with stop-codons being among the latter. The codon boxes with a mixture of G = C and A–U pairs in the first and second positions are intermediate in the various energetic parameters considered. Based on the above criteria, codons of the unsplit 4-codon boxes appear to fulfill enough criteria to bind correctly to their cognate aa-tRNAs within the ribosomal A site. They probably possess enough energy from both the internal codon–anticodon mini helix and the additional external stabilizing grip energies from both the ribosome and the tRNA anticodon loop.

Remarkably, in certain ribosomal contexts (Mycoplasmas, mitochondria), the energetics of the first two base pairs of the codon–anticodon mini helix appears to be sufficient; a single tRNA (containing unmodified U34) is able to decode each of the four GC-rich codons of unsplit decoding boxes (57–60), although with different efficiencies. This is known as the 2 out of 3 decoding rule (26,33). In contrast, a minimum of two tRNAs (one with G34, a second one with modified U34—see below—and eventually a third one with C34) is required to read AU-rich codons of each split 2:2-decoding boxes (57–59). Apparently, below a certain energy threshold, the contributions of the first two base pairs are insufficient to sustain translation and a third base pair has to complement the missing energetics. This is probably why the A/U-rich codons of the split 2:2-decoding boxes, where pyrimidine-ending (Y-ending) codons must be distinguished from purine-ending (R-ending) codons, have been

selected along evolution for expansion of the genetic code with additional amino acids (see below).

## STRUCTURAL CHARACTERISTICS OF THE tRNA ANTICODON HAIRPIN

Transfer RNAs are not passive molecules but active participants during translation. They can modulate their binding and specificity through sequence preferences at key positions, extending beyond the anticodon triplet (61–63), and through very diverse base modifications, particularly in the anticodon loop and stem (14,64). In the following sections, the contributions of the sequence preferences and base modifications to smooth and uniform decoding are analyzed within the framework just described, with a focus restrained to the anticodon stem and loop.

Figure 4 depicts the general distribution of conserved and semi-conserved nucleotides and the presence of modified nucleotides within the anticodon hairpin of 400 native elongator tRNA sequences originating from organisms spanning Bacteria, Eukarya and Archaea (65). Initiator tRNA<sup>Met</sup> is excluded from the analysis because it participates in a different step of translation and its anticodon hairpin is quite distinct from those observed in elongator tRNAs (66).

The distribution of purines and pyrimidines in elongator tRNAs is clearly asymmetric and the presence of a few conserved and semi conserved bases is evident. Some of the bases along the stem are post-transcriptionally modified to  $\Psi$  or simple methylated derivatives, which are known to better stack with neighboring bases, thus favoring a locally more rigid conformation (67). In the middle of the stem, base pair B29-B41 is almost always a Watson-Crick pair, and either a G = C or a C = G pair is usually found at positions B30-B40, attesting to the need for a relatively stable helical conformation in this part of the anticodon stem. Depending on the tRNA species, either a C = G or a A- $\Psi$  pair is preferred at B31-B39; thus the nucleotide entering the stem at the 3'-side is generally either a G39 or a  $\Psi$ 39, two bases which favor the start of a helical stem (35,68). Therefore, all bases of the anticodon stem and particularly the last three base pairs stabilize helical conformations. In the anticodon loop, the first and last residues, B32 and B38, also present a biased distribution that maintains in this case a non-Watson-Crick pair, with B32 almost exclusively a pyrimidine Y (with a preference for a C or a U, again depending on the tRNA species) and B38 mainly a A followed by a less frequently used C (see Table 1 in (69)). Thus, at variance with B31-B39 pair, B32 and B38 tend to form a non-Watson-Crick pair with a single bifurcated contact between exocyclic O2(Y) and exocyclic N6(A) (or N4(C)) (45,70).

At B37 a strictly conserved purine is present (mostly A37 or a modified A37 derivative). When G37 is present, it is invariably modified into N1-methylated G37 or a Wyosine derivative in tRNA<sup>Phe</sup> of eukaryotes and Archaea (Figure 5A). Modifications at A37 depend on the identity of the adjacent nucleotide at position 36; it is generally modified when B36 is either A or U. The purine residue at 37 stacks over the first codon/anticodon base pair (40,71,72) and propagates the stacking continuity to neighboring residues B38, B39 and the anticodon stem (preferential 3'-stacked

conformation). Modifications at the cyclic C2 atom and the exocyclic amine at C6 of A37 (and especially the highly hydrophobic tricyclic derivatives of the Wye-base) reinforce the stacking power of A37 with the neighboring nucleotides (12,72). Likewise, at the first position of the anticodon (B34), a plethora of diverse modified nucleotides are found (Figure 5B, and discussed below). Their functions are mostly related to decoding within the ribosomal decoding site, however some of them, like Cm, Um, s<sup>2</sup>U,  $\Psi$ , ac<sup>4</sup>C, Gm and Q, also contribute to the pre-structuring of the anticodon loop (31,73,74).

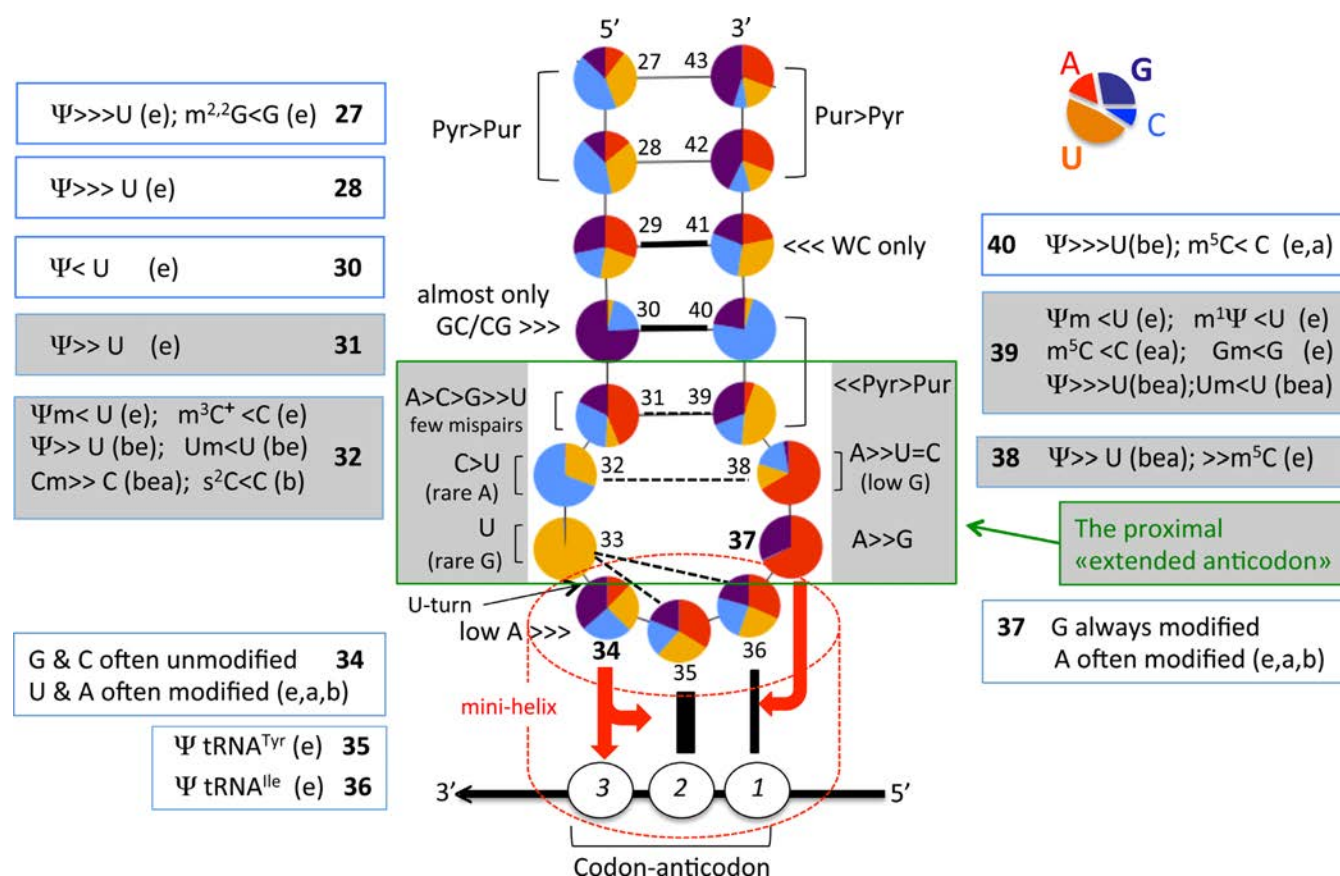
## THE PROXIMAL 'EXTENDED ANTICODON' CONCEPT

More than 30 years ago, Yarus (61) pointed out the strong co-variations between nucleotide B36 of anticodon and several bases of the whole anticodon stem-loop of *E. coli* tRNAs and introduced the 'extended anticodon' concept. Many experiments further confirmed such an apparent but indirect link between the sequence of the anticodon stem and the performance of the experimentally mutated bacterial tRNAs to efficiently translate certain codons both *in vivo* and *in vitro*. However, as more tRNA sequences became available from many different organisms, and more experiments were performed, it became evident that these co-variations were particularly relevant for the nucleotides of the anticodon loop and the last base pair B31-B39 of the anticodon loop (indicated in grey background in Figure 4 and thereafter designated as the proximal 'extended anticodon').

Figure 6A–C corresponds to an update of Yarus' original data in the form of the 'wheel of code' for the 40 elongator tRNAs of the complete *E. coli* repertoire that read all the 61 codons for the 20 canonical amino acids. In Figure 6A, are displayed the distributions of B31-B39 base pair, B32/B38 in Figure 6B and B37 in Figure 6C (including the acronyms of modified nucleotides). The same analysis has been done for the elongator tRNA repertoire of the archaea *Haloferax volcanii* and the eukaryote *S. cerevisiae* and as well as in two extreme situations for which enough sequence information (including for the modified bases) is reasonably well known, *Mycoplasma capricolum* and human mitochondria (see Supplementary Figure S3). A separate analysis is mandatory for each organism because notable differences exist in the decoding strategies (though not in the genetic codes) of distantly evolutionary related organisms, such as those of Bacteria, Eukarya and Archaea (64,69).

The most frequently found B31-B39 base pair (boxed in black) is C = G, especially in tRNAs of the 4-codon family boxes (Figure 6A, upper part of the wheel). In contrast, a A- $\Psi$  pair is predominantly present in tRNAs of the split 2-codon boxes (lower part of the wheel). Concerning B32/B38, the combination Y32/A38 is the most frequently found one, with both U32 and C32 occasionally modified to  $\Psi$  or Um and C32 to Cm or s<sup>2</sup>C, respectively (Figure 6B). These modifications of the base, especially the pseudouridine and the 2'-O-methylation of the ribose, are known to limit significantly the conformational flexibility of the nucleotide and thereby locally rigidify the corresponding portion of the nucleic acid (75).





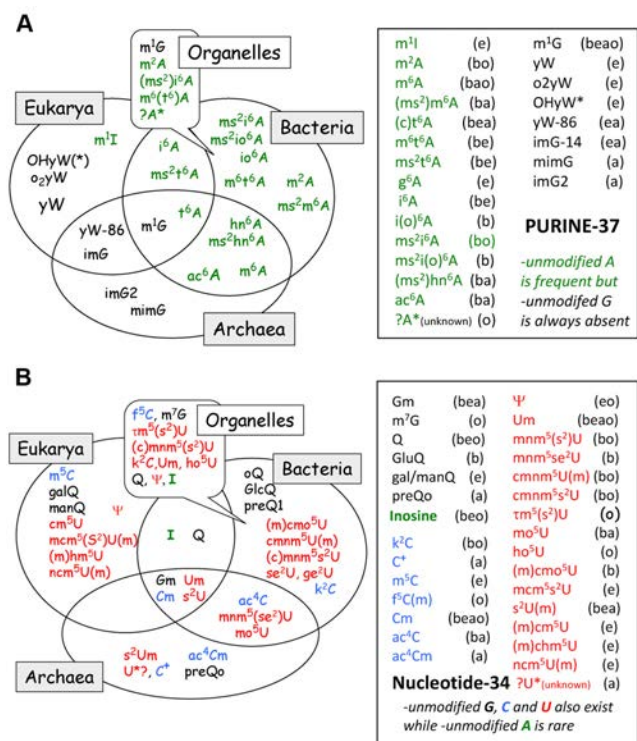
**Figure 4.** Structural characteristics of anticodon hairpin of tRNA. Identity and relative frequency of a nucleotide (including modified ones) are obtained from compilation of 382 elongator tRNAs belonging to the three domains of life (123 from Eubacteria, 55 from Archaea, 204 from Eukaryota). Initiator tRNA<sup>Met</sup>, tRNAs coding for Pyl and Sec, all tRNAs from mitochondria/plastids and bacteriophages/virus were excluded from the analysis. The data set comprises tRNA sequences present in the Modomics database (14) that currently contains all sequences available by 2009 in the tRNA database as in ref. (76). Analysis was performed using the software tool tRNAmoviz (<http://genesilico.pl/trnamoviz>). Distributions of nucleotide residues at each position of the hairpin (positions 27 to 43) are visualized as a pie chart, of which the color code is indicated at the top right corner of the figure. Universal numbering system is used (144). Acronyms of modified nucleotides present in certain isoacceptor tRNA are indicated outside the anticodon hairpin ('b' means bacteria, 'e', eukaryote and 'a', archaea). Those present in the proximal 'extended anticodon' (in green square box, see also text) are in grey background. Modified nucleotides present at positions 37 and 34 are listed in Figure 5A and B, respectively. The red arrow pointing from B34 of anticodon to B3 of the codon and the plateau of B35/B2 and the red arrow from B37 to the plateau of B36/B1 symbolize the stabilizing effects of B34/B37 on codon-anticodon pairings.

One remarkable exception is found in tRNA<sup>Ala</sup> (anticodon GGC) and tRNA<sup>Pro</sup> (anticodon GGG) with A32 instead of the quasi universal pyrimidine, but not in the other isoacceptor tRNA<sup>Ala</sup> (anticodon U\*GC) and tRNA<sup>Pro</sup> (anticodons UGG and CGG, respectively) in *E. coli* (65,76). Base-38 is also occasionally a U or C instead of the usual A38, the uridine-38 being often modified to  $\Psi$ . This situation is found in majority of tRNA specific for Ala, Leu (of the 4-codon boxes), Val, His, Gln, Asp, and Glu, all belonging to decoding systems of the upper half of the codon wheel. In some instances, B32-B38 can even form a Watson-Crick pair, either U32-A38 (Pro, Gly, Thr) or A32-U38 (Ala, Pro). Again, these tRNAs with a potential B32-B38 Watson-Crick pair all belong to the group of strong 4-codons decoding tRNAs (Ala, Pro) and, surprisingly also to tRNA<sup>Thr</sup> belonging to the intermediate group. The case of tRNA<sup>Ala</sup> has been studied in detail (36,77–79). The take-home lesson is that tRNAs harboring too strong a base-pairing binding capacity have been 'tuned down' during evolution for optimal and uniform translation. As dis-

cussed above, B32-B38 should not form a Watson-Crick base pair that antagonizes the canonical conformation of the anticodon loop with the anticodon triplet in a correct helical orientation. With complementary bases at B32 and B38, one expects that a Watson-Crick pair is formed in some tRNA population, leading to a less optimal anticodon conformation for pairing with the codon and, thus, possibly rejection from the ribosomal decoding site. Without a selection for the preformed conformation of the anticodon loop, the strong codon/anticodon pairs would allow miscoding by binding to other G/C-rich near- or even non-cognate aa-tRNAs.

The distributions of B37 in Figure 6C (including post-transcriptional modifications) follow each quadrant and thus, strongly depend on the codon base at the first position. This conclusion is valid for the other tRNA repertoires analyzed (Supplementary Figure S3). The presence of stabilizing modifications such as ct<sup>6</sup>A and ms<sup>2</sup>i<sup>6</sup>A in *E. coli* (or i<sup>6</sup>A, m<sup>1</sup>G, yW, m<sup>6</sup>A in other organisms, Supplementary Figure S3) correlates with the need to stabilize weak neighbor-





**Figure 5.** Phylogenetic distribution of modified and hypermodified nucleosides at (A) B37 and (B) B34 of anticodon hairpin of tRNA from the 3 domains of life. All acronyms are those conventionally used, the corresponding chemical structures, full scientific names and chemical characteristics of most of them can be found in (145) (see also in: Modomics, <http://modomics.genesilico.pl/>). Only a few tRNA from Archaea have been sequenced so far, therefore information concerning this domain (especially for U34 modifications) is incomplete. Meanings of 'b', 'a', 'e' is the same as above in Figure 4 with that of 'o' corresponding to organelles. The data set comprises tRNA sequences present in the Modomics database (14).

ing codon–anticodon base pair A1–U36 or U1–A36, respectively. These modifications, as well as the N1-methylation of G37, which is unable to base pair in a Watson–Crick mode, are also required for limiting ribosome frameshifting at specific AU-rich ‘slippery’ sequences of the mRNA (80,81).

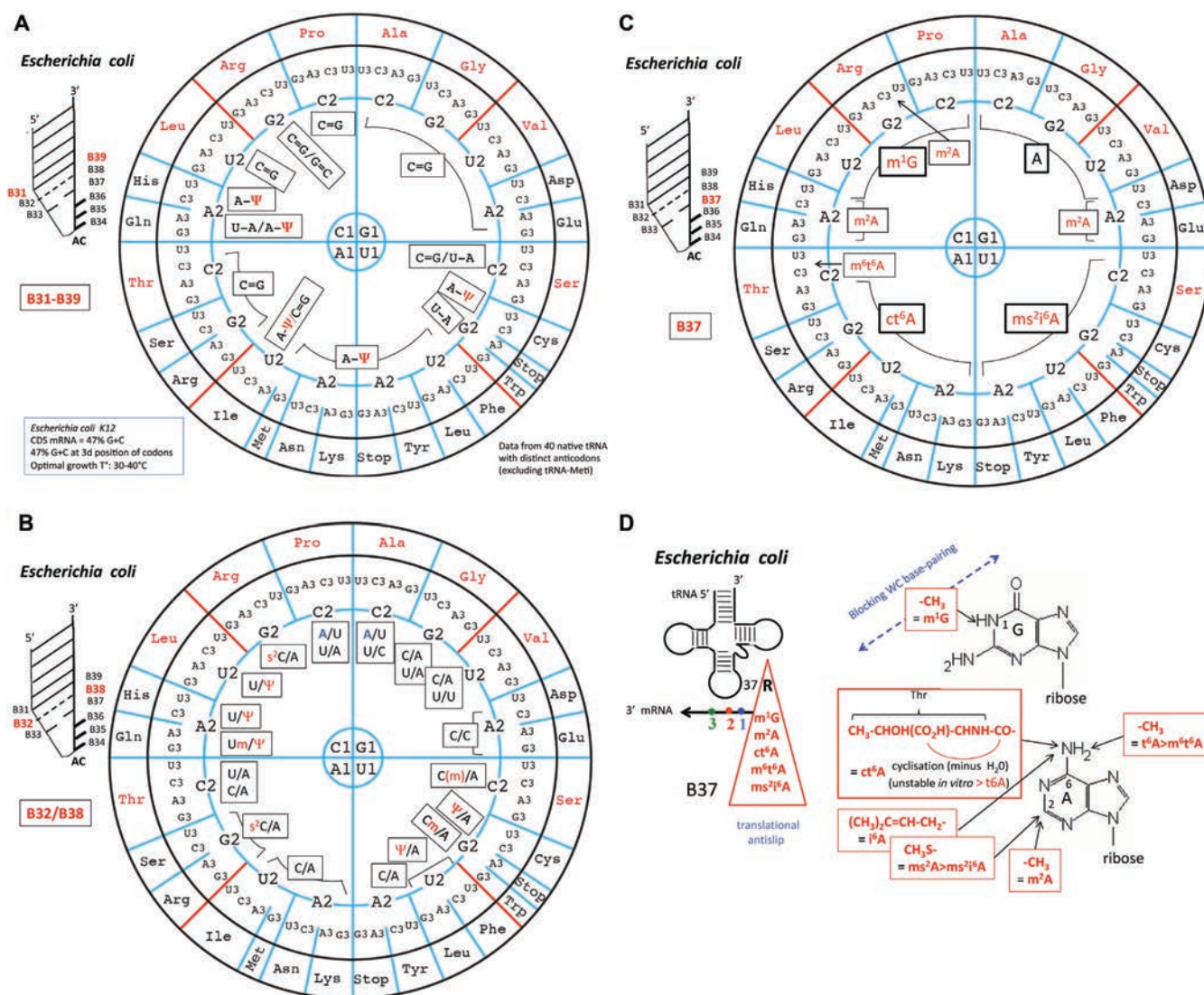
In summary, purine-37, together with B31–B39 and B32/B38, and contacts with U33 (grey background in Figure 4), synergistically contribute to maintaining a conserved pre-formed and optimal anticodon conformation, 3'-stacked with both the rest of the anticodon arm and the last two bases B35 and B36 in the anticodon. Supplementary Table S1 shows an approximate partition of these energetic elements that, together with the distribution of key elements on the wheels, helps explain how these apparently disparate interacting elements (including modified bases and ribose methylation) may cooperate to average the binding energies of tRNAs, whatever the GC or AU content of their anticodons to the mRNA-programmed ribosomes. One advantage of such an intricate networking system is the possibility to fine-tune the aminoacyl-tRNA for its ability to read the various cognate codons with appropriate efficiencies while avoiding reading near- and non-cognate codons.

## CONTRIBUTION OF NUCLEOTIDE MODIFICATIONS AT THE FIRST tRNA ANTICODON POSITION

A large variety of post-transcriptionally modified nucleotides are found at position 34 of cytoplasmic elongator tRNAs (65,76). Partitioning of these nucleotides according to organisms in the three domains of life is given in Figure 5B. The most frequently and diversely modified nucleotide at position 34 is uracil, followed by cytidine and guanine. The only modified form of the adenine is hypoxanthine (inosine). Methylation of 2'-O-ribose also occurs at G34, C34 and U34 (in the latter case, in combination with additional modifications at C5- and/or C2-atom of the base). These chemical adducts to B34 or its ribose alter the nucleotide chemical properties, and thus its polarity, but also the preference for the *syn* or *anti* conformer and the base's tautomeric form. They can also control ribose puckering and, in addition, allow for additional interactions with other elements of the translation machinery (see Supplementary Table S2 for an overview).

The distribution of modifications at B34 (C34, A34, G34) in elongator tRNAs according to the triplet code is depicted in Figure 7A in the case of *E. coli*. The same analysis for *Haloferax volcanii*, *S. cerevisiae*, *Mycoplasma capricolum* and human mitochondria are given in Supplementary Figure S4. Again, a clear segregation exists between tRNAs belonging to the ‘strong binding’ class (top part of the wheel) and those belonging to the ‘weak binding’ class (bottom part of the wheel). Except for inosine-34, which is found only in bacterial tRNA<sup>Arg</sup> of the 4-codon boxes, other modified nucleotides at position 34, like ac<sup>4</sup>C, k<sup>2</sup>C, Cm, Um, Q and gluQ are found only in tRNAs belonging to the 2:2 and single-codon boxes.

The situation concerning modified U\*34 (\*: modified at the C5-atom) and its doubly modified s<sup>2</sup>U\* derivatives constitutes a special case. In the ribosomal grip, the third nucleotide of a codon is constrained by several contacts with the ribosome (specifically with G530, C518 of the 16S rRNA and a residue from RpS12), while the anticodon nucleotide B34 is in a loose stacking contact with C1054. When B34 forms a Watson–Crick pair, this asymmetry in contacts is not discriminatory. However, when forming a GoU wobble pair, the unsymmetrical constraints amplify the non-isostericity of UoG pairs such that G34oU3 wobble pairs can form but U34oG3 pairs form much less readily (for more detail see Supplementary Figure S5). Indeed, in several crystal structures with a U34oG3 opposition, the pair formed is not the standard wobble pair but a Watson–Crick-like pair between tautomeric forms of either G or U (27,29–32). The modifications at U34 are structurally imposed by the constraints on the codon base. Thus, in the non-isosteric GoU pairs, modifications on U34 will be necessary throughout the wheel, especially in the absence of additional C34-containing tRNAs. With C34-containing tRNAs present, the modifications at U34 are less important, since there is no need to form a U34oG3 pair. Despite this, when chemical adducts at the C5-atom of uracil are examined, a clear segregation by type exists depending on whether the modified U34-containing tRNAs belong to exclusively 4-codons (strong) or mainly 2-codons (weak)



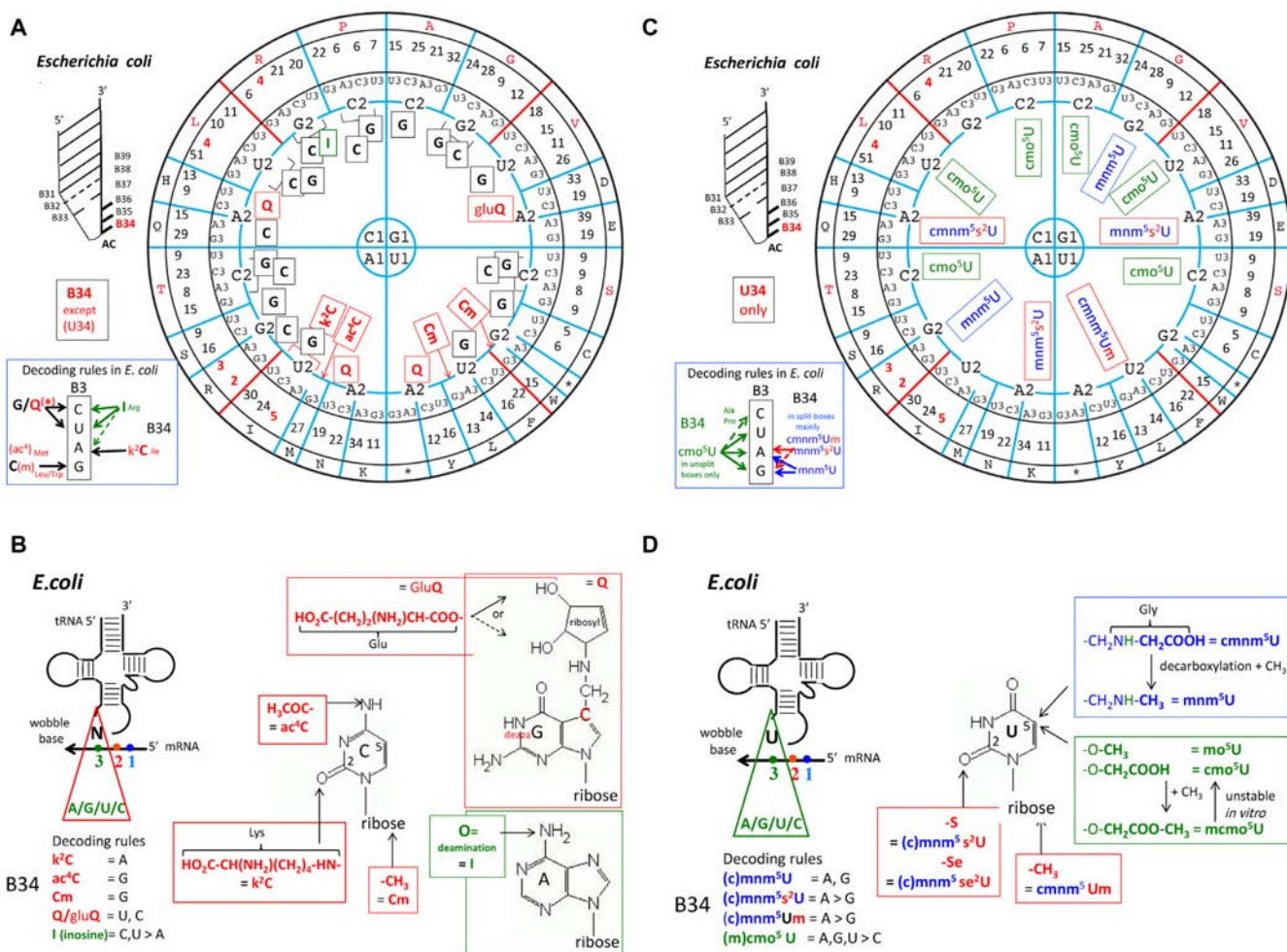
**Figure 6.** Architecture of the proximal extended anticodon loop of *Escherichia coli* tRNAs. On the circular representation for the code are given (A) the base pairs B31-B39, (B) the base opposition B32/B38 and the (C) identity of purine-37 found in the anticodon loop of the various tRNA species corresponding to each of the decoding boxes for the 20 amino acids. Modified nucleotides are indicated in red. Non-random usage of base pairings B31-B39 is apparent, with a frequent use of A31- $\Psi$ 39 in tRNA in the 'weak' decoding boxes. Likewise, the B32/B38 positions are more variable in the top of the wheel. Modification of B37 is clearly dependent on B1, thus of B36 of anticodon that has to base pair with B1 codon. Only in a limited number of isoacceptor species, another modified B37 ( $\text{m}^2\text{A}$  or  $\text{m}^6\text{t}^6\text{A}$ ) are found. Same analysis for tRNAs from *H. volcanii*, *S. cerevisiae*, *M. capricolum* and human mitochondria are shown in Supplementary Figure S3. (D) Chemical structures of the modifications found at position B37. Notice the presence of amino acid as part of  $\text{ct}^6\text{A}$  modification.

boxes (Figure 7C indicated in green and in blue, respectively).

Without exceptions, aminoacyl-tRNAs reading codons of the split 2:2 or 3:1 decoding boxes, as well as those reading codons harboring A and/or U at the B1-B2 codon positions (bottom of the wheel) are the most diversely post-transcriptionally (hyper)-modified at B34, but also at other positions of the anticodon hairpin (compare Figure 6A-C). Evidently, fine-tuning the efficacy and accuracy of mRNA translation is a highly sophisticated process in which the chemistry of each nucleotide (modified or not) within the anticodon hairpin of the various individual natural elongator tRNAs collectively play a pivotal role (61). Figure 8 is a

comparison between the general structures of the anticodon loop and proximal stem of the 10 *E. coli* tRNAs belonging to the STRONG class (panel a) and the 9 other *E. coli* tRNAs belonging to the WEAK class (panel b). The same figures for the cases of *H. volcanii* (archaea), *S. cerevisiae* (eukaryote), *M. capricolum* (minimalist bacteria) and human mitochondria (organelle) are presented in Supplementary Figure S6. Evidently each tRNA has evolved in such a way that substantial differences exist at almost every position of the molecule, the most characteristic ones being of course B37 and B34, especially for those tRNAs harboring A or U at the two last positions of anticodon (B37 and B38).





**Figure 7.** Identity of nucleotides at the first anticodon position (B34) of *E. coli* tRNAs. The global codon usage (after (146)) is inserted between the circle for the third base and that for the amino acid type. Conventional one letter code for amino acid is used. For the sake of clarity, the distribution of G34, I34 and C34 derivatives are shown in (A), and U34 derivatives in (B). Modified G/C/I-34 are indicated in red. All Q-containing tRNAs and those containing modified C\* are found in tRNA belonging to split 2:2-codon boxes. For the modified U34 (shown in (C)), two chemically distinguishable types of modified residues are found, one harboring an oxyacetic acid group (sometimes methylated) at the C5 atom of uracil ( $\text{cmo}^5\text{U}$  or  $\text{mcmo}^5\text{U}$ ), are indicated as green. These U34 derivatives are found only in isoacceptor tRNA belonging to the unsplit 4-codon family boxes. The second type of modified U34 derivatives harbors a methylaminomethyl (sometimes carboxymethylated) group at the C5 atom of uracil ( $\text{mnm}^5\text{U}$  or  $\text{cmnm}^5\text{U}$ ). They are indicated as blue, some of which are also hypermodified into 2-thiolated derivatives ( $\text{s}^2\text{U}^*$ ) or methylated on the 2'-hydroxyl ribose ( $\text{U}^*\text{m}$ ). They are found in all split 2:2-codon boxes, and in the Arg/Gly 4-codon boxes (after (88)). Same analysis for tRNAs from *H. volcanii*, *S. cerevisiae*, *M. capricolum* and human mitochondria are shown in Supplementary Figure S4. (D) Chemical structures of the modifications found at positions B34. Notice the presence of amino acid as part of a few modifications.

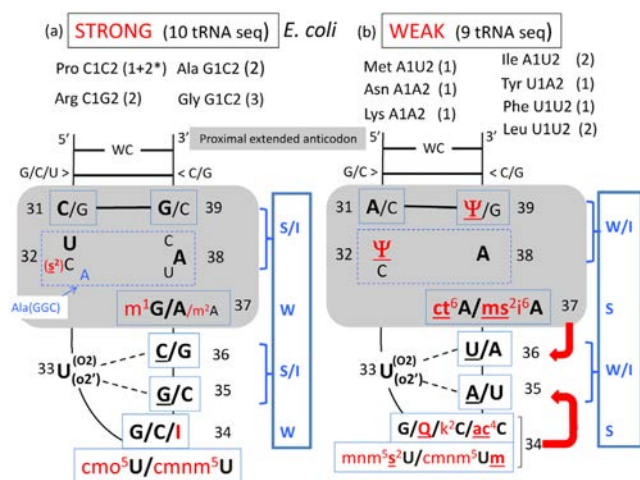
## SHORT SUMMARY OF THE STRUCTURAL ASPECTS

Overall, the three-dimensional structure of the anticodon loop is characterized by two helical stacks interrupted at the phosphate 3' of the invariant U33 residue: the 5'-stack in continuity with the 5'-strand of the anticodon stem stops at residue U33 (the U-turn), while the 3'-stack extends from residue B34 to B38 in continuity with B39 to B43 in the 3'-strand of the anticodon stem. Although residues B34 to B38 form a helical stack with standard rotation angles between nucleotides, the 3'-strand conformation does not lead to enough rotation to allow B38 to form a Watson–Crick pair with B32 (70). Indeed, formation of a Watson–Crick pair between B32 and B38 would require a change in the conformation of the rest of the loop. B39, however, has rotated enough and is able to form a Watson–Crick pair with

B31. Evidently, various combinatorial structural solutions exist to generate a canonical architecture with a balance between flexibility and rigidity of the anticodon hairpin that allows the elongator tRNA to perform a common decoding function on the ribosome. This common functional architecture in elongator tRNAs depends very much on the conserved residue U33. Only in exceptional cases, as in many initiator tRNA<sup>Met</sup> in eukaryotes (not discussed here) or in the tRNA<sup>Ser</sup> of most genus of *Candida* species harboring a leucine anticodon CUG, have differences at this positions been found (C33 or G33, respectively) (65,76).

Thus, the deciphering of AU-rich codons within the decoding A site of the ribosome requires extra stabilization sources that of course have to arise mainly from the tRNA. Along the course of evolution, various strategies based,





**Figure 8.** Modulation of codon-anticodon binding according to G+C (STRONG) or A+U (WEAK) binding capability of selected *E. coli* tRNA species. In both cases, the anticodon hairpin is schematically represented with all the nucleotides of the 5' branch in continuous stacking up to B34 and the complementary nucleotides of the 3' branch in continuous stacking up to B32. The U33 turn is indicated with its links to R35 and/or Y36 (underlined). At positions B32 and B38, various combinations of base opposition are found (boxed with dashed lines), the most frequent ones being indicated in bold letters. In red are the modified nucleotides; underlining emphasizes that the chemical adduct reinforces the stacking power of the base with the neighboring nucleotides. Modulation of the strength of codon-anticodon binding occurs by the anticodon loop constraints that mainly depends on the choice of the B32-B38 base opposition, additional interactions with the conserved U33 and the identity of the chemical adducts on B37 and B34 that stabilize the B36-B1 and the B35-B2 interactions, respectively (schematized by red arrows). The number 2 with an asterisk for Pro, means that information came only from tDNA sequence, the corresponding matured transcripts have not been sequenced yet. On the right of each diagram, an approximate energy scheme is displayed (in a blue rectangle) with S standing for STRONG, I for INTERMEDIATE, and W for WEAK. Same analysis for tRNAs from *H. volcanii*, *S. cerevisiae*, *M. capricolum* and human mitochondria are shown in Supplementary Figure S6.

as in any complex molecular system, on the interplay between many weak interactions together with a few strong interactions have offered alternatives to overcome this decoding problem. In the proximal 'extended anticodon', every single nucleotide between 31 to 39 in the anticodon loop contributes to uniform decoding by adapting its structure to the strength of the codon/anticodon pairings and the contacts within the ribosomal grip. Some combinations of B31-B39 and B32-B38 favor the canonical anticodon triplet conformation, contributing favorably to the free energy of binding, while others can be either too floppy or too rigid (with even the formation of an additional base pair at B32-B38), contributing unfavorably the free energy of binding. The stacking contacts of the modified base B37 sandwiched between B36 and B32-B38 can be extensive or minimal and contribute at the proper interface between the codon/anticodon triplet and the anticodon stem. Finally, the critical role of modifications of B34 for promoting non-standard base pairs has been emphasized.

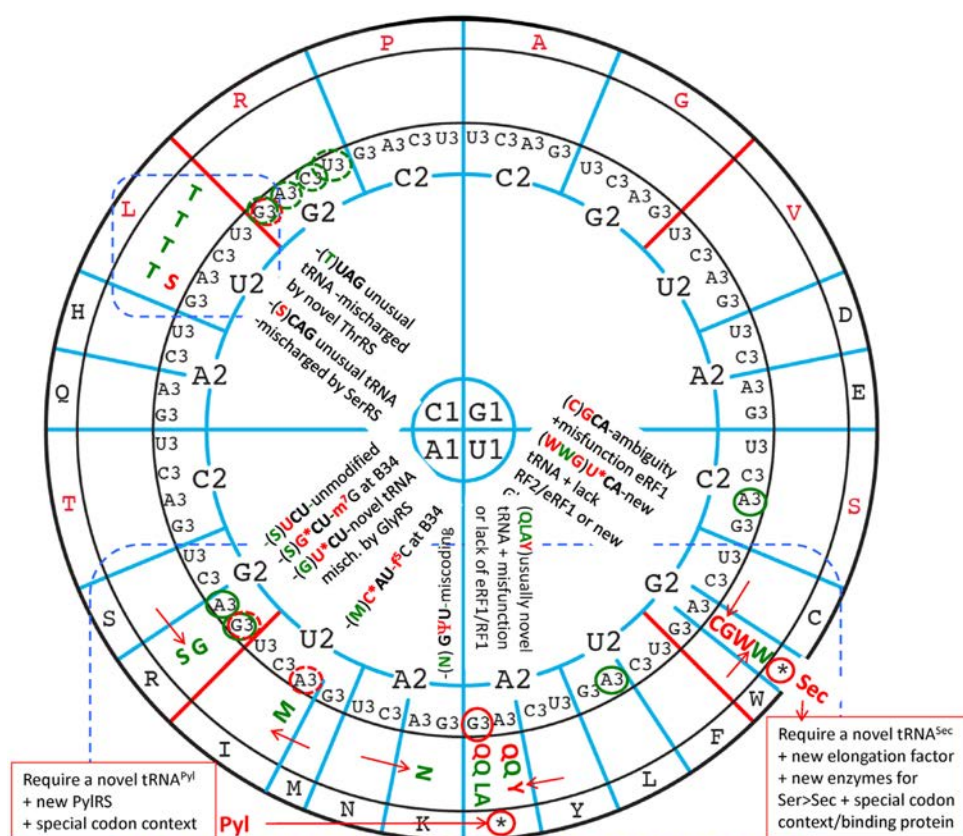
## DEVIATIONS FROM THE STANDARD GENETIC CODE

Despite the various decoding strategies existing in Bacteria, Archaea, Eukarya and organelles, the meaning of each of the individual 64 triplets of the genetic code remains remarkably universal. However, the genetic code is not frozen and a few deviations have been identified in the nuclear genomes of a limited number of species (fungi, green algae, ciliates, diplomonads and firmicutes), though mostly in the small A+T rich genomes of mitochondria (metazoan, fungi, green plants and red algae) (1,82,83). Particular codon reassignments appeared to have taken place repeatedly and independently in the different lineages, attesting that certain codons are more prone than others to this evolutionary process. Also, two proteogenic amino acids, selenocysteine (Sec) and pyrrolysine (Pyl), have been added to the standard 20-member amino acid alphabet (natural expansion of the genetic code). These two amino acids are inserted in a few proteins in response to stop codons UGA and UAG, respectively, of cytoplasmic mRNAs of a few bacteria, eukaryotes and even archaea (84,85).

Figure 9 summarizes the best-characterized genetic code deviations identified so far. The most frequently encountered deviations in both nuclear and mitochondrial genomes concern the switch of terminator codons to sense codons (UGA to Trp, Cys, Gly or Sec, UAG to Gln, Leu, Ala or Pyl, and UAA to Gln or Tyr). Identity changes of sense codons are less frequent and found mainly in mitochondrial genomes (except in one case). They are AUA-Ile to Met; AAR-Lys to Asn, AGA/G-Arg to Ser or Gly, all belonging to 2-codons family boxes for which the first base of the codon is either A1 or U1. The few cases where a C1-containing codon is involved are CUY/R-Leu to Thr or Ala in mitochondria of a few filamentous *Saccharomycetaceae* and the CUG-Leu to Ser in the cytoplasm of *Candida* and *Debaryomyces* species. In these latter cases, the replacement of one amino acid by another results mainly from mischarging of a tRNA rather than of misreading of a codon during translation. Among the sense codons, those starting with G1 appear most resistant to amino acid reassignment. Also, as a direct consequence of strongly biased codon usage in mitochondria, a few codons are either not used (remain unassigned) or possibly used as alternative stop codons (2).

Evidently, the most frequent codon reassignments are found in A/U-rich weaker codon binders and correspond mostly to switch between two amino acids of the same pair of split codon family boxes. These codons are also the most liable to translational miscoding (see for example (86-88)). Similarly, the expansion of the genetic code to additional selenocysteine (Sec) and pyrrolysine (Pyl) occurs mainly, but not exclusively at stop codons, again at the same more frequently AU-rich reassigned codons (85,89).

Changing the meaning of a codon in a genome is challenging. It implies a few coordinated adjustments of various elements of the translation apparatus that necessarily result from a multistep evolutionary process. Among these steps are: (i) mutation in a gene coding for a tRNA (or a duplicated variant) to allow the emergence of a 'novel' tRNA able to efficiently read a given codon (or a set of near cognate codons) that is (are) distinct from the original



**Figure 9.** Deviations from the standard, almost universal genetic code. The most frequent deviations concern terminator codons unexpectedly efficiently translated as sense codons for amino acids Trp, Gly, Cys, Gln, Tyr, Leu or Ala. Conventional one letter code for amino acid is used. In red are shown amino acid reassignments observed in nuclear genomes while in green are those related to mitochondrial genomes. The less frequently encountered reassignment of sense codons for a non-standard amino acid have been observed essentially in mitochondrial genome (indicated in green), except in one case (a Leu-codon coding for Ser in *Candida* and *Debaryomyces* species (83). For more details and references of original papers (in addition to those cited in text), (1,82,147). The codons remaining unassigned and possibly playing the role of occasional alternative stop codon in certain organisms are indicated by a red (nuclear) or green (mitochondria) circle around B3, dotted line means avoided codons, plain lines means unassigned, possibly stop codons. The red arrows between 2 boxes indicate a switches between amino acids within 2:2 decoding boxes. Outside the circle, are indicated the reassignment of stop codon UGA and UAG into Sec (21st proteogenic amino acid) and Pyl (22nd amino acid), respectively (2). The most frequent codons reassignments mostly occur within the blue dotted boxes.

wild type tRNA species; (ii) mutations in the gene coding for an aminoacyl-tRNA synthetase (or a duplicated variant) to allow the mutant enzyme to charge efficiently the newly evolved tRNA with an amino acid distinct to that expected from its anticodon; (iii) mutations in the gene coding for a modification enzyme (or a duplicated variant) to allow (or anneal) specific base modification that can influence the decoding property of the novel tRNA; (iv) mutations in the gene coding for potential competing tRNA (or even its elimination) that may read the same reassigned codon(s); (v) mutations in the termination factor (or even its elimination) to avoid competition with the novel functional suppressor tRNA (discussed in (90)); (vi) sequence rearrangement in mRNA for signaling with stop codon (or sense codon) has to be recoded (or reassigned); (vii) invention of a new elongation factor for selectively bringing the Sec-tRNA<sup>Sec</sup> into the decoding site of the translating ribosome; (viii) possibly mutations in the rRNA or r-proteins of the decoding site of the ribosome to allow optimal translation of the newly reassigned codon by its new cognate (or near cognate) aa-tRNA. Not all of these adjustments of the

translation apparatus are simultaneously required. However, depending on the case, at least a combination of them has to occur. In this way, each time a reassigned codon, or a stop codon with the appropriate neighboring sequence is encountered in mRNA, the newly assigned amino acid is incorporated following standard decoding mechanism on the ribosome. This is at variance with the phenomena of translational missense error or stop-codon 'readthrough' that are mediated by a standard aa-tRNA<sup>aa</sup> harboring non-cognate anticodon or after occasional mischarging of a tRNA with a wrong amino acid, which *in fine* give rise to occasional misincorporation of an amino acid into the growing polypeptide. The yields of such naturally occurring mischarging and misreading are usually very low (below  $10^{-5}$  to  $10^{-3}$ ), unless, again, mischarging is genetically programmed by multiple sequential mutations in the tRNA, mRNA, various translational proteins and/or even the rRNA or proteins of the ribosome in order to eventually reach the degree of perfection which has been achieved for the reassignment or recoding of codons in a few organisms (91).

Among all the above combinatory strategies for codon reassignments, the more intriguing ones are those involving posttranscriptional modification machineries. For example, in mitochondria of several metazoa and budding yeasts, C34 of the unique Met-tRNA<sup>Met</sup> is post-transcriptionally modified to 5-formyl-cytidine (f<sup>5</sup>C34) by a unique mitochondrial 5-formyl-cytosine synthase that allows the mature Met-tRNA<sup>Met</sup> to translate both codons AUG and AUA codons as Met (92,93). The completeness of such AUA-Ile-to-Met reassignment depends on the absence of a competing lysidine-containing Ile-tRNA<sup>Ile</sup> (anticodon k<sup>2</sup>CAU). However, in mitochondria of a totally different lineage (Ascidian *Halocynthia roretzi*), instead of a f<sup>5</sup>C34, a 5-taurine-2-thio-methyl-2-thiouridine (τm<sup>5</sup>s<sup>2</sup>U34) and an exceptionally unmodified U37 were found in Met-tRNA<sup>Met</sup> (94). In the mitochondria of the squid *Loligo bleekeri*, an unmodified C34 and a rather normal anticodon loop were found but differences in other parts of the Met-tRNA<sup>Met</sup> were required for reading the AUA codon as Met (95). Evidently, different independent solutions have been developed along mitochondrial evolution for solving the same problem. Several other examples of this sort have been reported for the reassignments AAA-Lys to Asn (dependence on a peculiar pseudouridine-35 in Asn-tRNA<sup>Asn</sup>) and AGA/AGG-Arg to Ser or Gly (dependence on a unique m<sup>7</sup>G34 in Ser-tRNA<sup>Ser</sup> or τm<sup>5</sup>U34 in a B34-mutant of the same Ser-tRNA<sup>Ser</sup>) (1). Most of these reassignments depend on base modifications and are prevalent for codons belonging to the A1-quarter of the wheel code and, therefore, frequently depend on the decoding property of a U34-containing tRNAs, for which the panoply of chemical modifications is huge (see discussion above about the non-isostericity of UoG pairs and Figure 5B and Supplementary Figure S5).

The presence of nucleotide modifications, like m<sup>6</sup>A, m<sup>1</sup>A, I or Ψ in mRNAs can now be measured and assessed in translation (49,96–98). The presence of m<sup>6</sup>A has moderate effects on the measured translation efficacy, the effect being strongest with m<sup>6</sup>A at the first position (49,99). The effects of O6-methylG depend strongly on the codon position, with incorporation at the second position B2 leading surprisingly to stalling of the ribosomes (98). O6-methylG, because of the absence of a hydrogen atom at N1 position of the base, cannot form a Watson–Crick-like base pair with C, but does so with U. Consequently, a decrease in accuracy at the first and third B1 and B3 positions was observed but, unexpectedly, not at the second position. Incorporation of Ψ in bacterial mRNAs led to a 30% repression of translation, with the strongest effect at the B3 position (49). However, in this respect, the conversion of nonsense into sense codons by the presence of Ψ in the mRNA is particularly striking (42,100) (Supplementary Figure S7 shows that the observed miscodings occur at the bottom of the wheel and especially in the U1 quarter). The miscoding effects of Ψ35, in which the absence of a queuosine (Q34) promote misreading of G34oA3 or G34oG3 (101) and have been discussed above. In the published crystal structures, Ψ1 leads to the formation of non-Watson–Crick *cis* Hoogsteen–Watson–Crick G2oA35 or A2oG35 pairs, with the mRNA base in the *syn* conformation (42). More crystal structures are necessary to complete our understanding of the molecular basis of these intriguing observations (102).

Because the activity of modification enzymes usually depends on available cofactors of the basic cellular metabolism, we anticipate that the efficiency of such type of codon reassignments depends on the physiological state of the cell and may constitute a regulatory device. They possibly correspond also to early evolutionary steps toward less dependent and more permanent amino acid reassignment.

Briefly, the most frequently encountered code deviations, recoding and codon reassignment, occur in the U1/A1 quadrants, while the G1 quadrant is still unaffected to date. The most frequent miscoding, i.e. decoding errors and code ambiguities also occur in the bottom half of the wheel (together with the Glu/Asp and His/Gln ambiguities) (86,87). Such errors are highly dependent on the tRNA modifications and constitute a rich source of post-transcriptional regulation. Recoding of stop codons (Sec and Pyl) require, not only new enzymes for the new metabolism (4 for Sec and 2 for Pyl), but also several new factors and cofactors and, especially structures in the mRNA for signaling a given stop codon for recoding. Codon reassignments require less specific changes in the translational apparatus (aminoacyl tRNA synthetases, absence of some release factors, etc.) but appear very efficient without competition and should probably be favored for further studies and applications.

## THE EVOLUTION OF THE GENETIC CODE

If the organization of the genetic code is, as we suggest, anchored in the structural contacts with the associated energetics correctly integrated within the networks of interactions at the ribosomal decoding grip, it should also reflect historical aspects of the evolution of the genetic code. Nowadays, it is understood that there is not a single explanation for the emergence and expansion of the genetic code as it stands to date. The current most plausible hypothesis is, however, that the earliest proto-biosynthetic system originated from RNA:RNA duplexes with the most stable complementary GC-rich triplets coding for small polypeptides composed of alternate hydrophobic Ala and hydrophilic Gly, the very first amino acids encoded. As more adenine and uracil was introduced into the RNA, this minimalist proto-biosynthetic system later progressively evolved to a more systematic use of repeating GNC, GNY, GNS or SNS triplets (N, Y and S denote respectively: any of the four canonical nucleotides; C or U; and C or G) (103–106). The earliest amino acids used to synthesize ancestral polypeptides were found in the prebiotic soup and selected according their specific interactions with the ancestral codons (the ‘stereochemical hypothesis’: (107,108)). These steps were subsequently expanded through the co-evolution with the invention of biosynthetic pathways for new amino acids (the ‘amino acid metabolism hypothesis’, initially pointed out by: (109); reviewed and updated: by (110,111)) and the emergence of the corresponding primordial aminoacyl-tRNA synthetases able to fix these new amino acids on appropriated proto-tRNAs (‘co-evolution with tRNA aminoacylation systems’: (112–114)). Constant refinement at both the replication and translation levels allow to progressively minimize the impact of coding errors and to increase the diversity and functionality of proteins that can be made with a larger amino acid alphabet (the



'error minimizing code hypothesis': (107,115,116)). Finally, as mentioned in the previous section, the code can further evolve by reassignment of unused, temporarily ambiguous or less used codons for another canonical or even totally new amino acids (the 'codon capture theory': (117,118)). Finally, early horizontal transfer and collective evolution of the code through different subspecies has been emphasized (see for examples (119,120)). In other words, the present day genetic code did not necessarily result solely from divergent evolution, but also from collective evolution via the development of an innovation-sharing process that allows the emergence of a quasi universal genetic code among populations of species 'speaking the same language' (121).

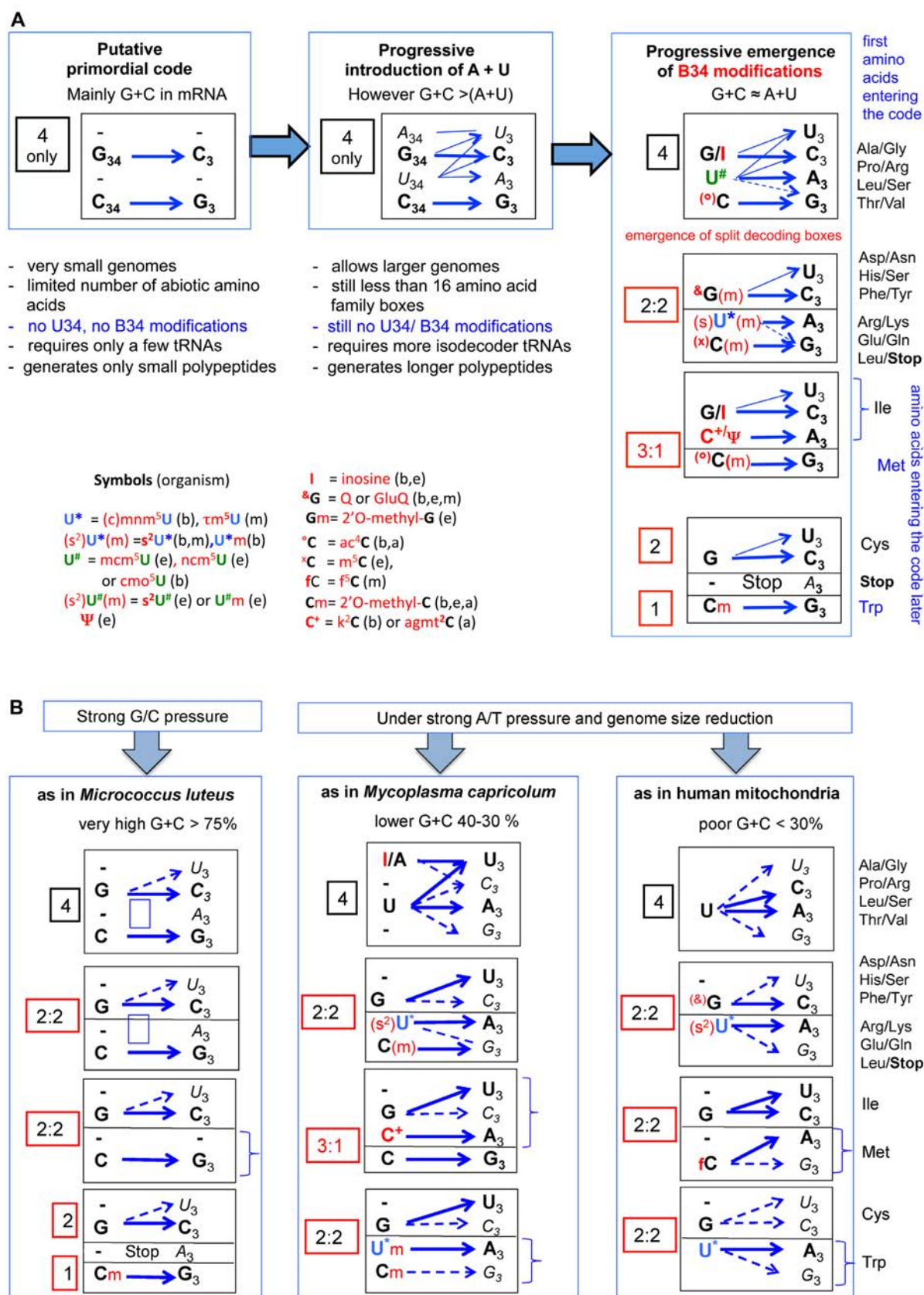
The take-home lesson of all the above information is that along the expansion of the genetic code, an optimal stability of complementary codon-anticodon pairs appears to have been the main evolutionary force. Starting with a G/C-rich RNA coding for simple abiotic amino acids like Gly, Ala, Pro and probably a precursor of Arg allows the subsequent development of more complex modern proteosynthetic machineries as we know to date. As argued by Trifonov (122), the temporal order of the sequential acquisition of new amino acids into the code seems to result from a 'descending' thermostability of the respective codon:anticodon pairs, with initiation AUG codon among the latest codons to enter the code, together with UAG/UAA/UGA codons which remain unassigned and serve as termination codons. As explained above, this 'descending' thermostability was obviously compensated along evolution by a better organization of the tRNA anticodon hairpin and of the ribosomal grip within the ribosome. The main differences in reading the 'ancient sense strong codons' of the 4-codon boxes as compared to the subsequently acquired 'new, weak, sense codons' of the 2:2-codon and 3:1-codon family boxes is that the former are less dependent than the latter on stabilizing modified nucleotides within the proximal extended anticodon (Figures 6 and 7).

The progressive introduction of A and U into the primordial G/C-rich RNA (both in proto-mRNA and proto-tRNA) was of great advantage for expanding the decoding potentiality of the proteosynthesis machinery (Figure 10A). However, as mentioned in a preceding section, the use of U-rich codons and U34-containing tRNA is promiscuous, both for miscoding and frameshifting of the comma-free mRNA during translation. It was possible to solve this dilemma only by avoiding usage of such U-rich 'slippery' codons and by developing specific enzymatic systems for posttranscriptional modifications of base of the anticodon hairpin, mainly at B34 (Figure 7) and B37 (Figure 6C) (64,123–125). Because frequent frameshift events are probably more deleterious to the cell than frequent miscoding of certain codons, it is possible that emergence of B37 modifications systems preceded modifications at B34 (126–128). The advantage of restricting the coding potentiality of certain tRNA by modifications of B34 is that it allows the introduction of new amino acids into the code, which, however, require duplication of a tRNA gene and of the emergence of the corresponding aminoacyl-tRNA synthetases (co-evolution of tRNA genes, tRNA maturation enzymes and tRNA aminoacylation apparatus).

The earliest of these modification systems would have appeared before the divergence of Woese's three major domains (129), perhaps very early or at least before completion of the genetic code. These modifications would be universally distributed today (130), while those more recently invented occur either specifically in archaea, bacteria and/or eukaryotes (Figure 5A–B, discussed in (64)). Lateral transfer of genes coding for a given modified enzymes present in bacteria, organelles and/or the cytoplasm of eukaryotes could also have occurred through endosymbiosis or virus infection. Remarkably, the chemical adducts of several modified nucleotides (B34 and B37) contain an amino acid, as in N6-threonylcarbamoyladenine ( $t^6A37$ ), N6-hydroxynorvalinecarbamoyladenine ( $hn^6A37$ ), N6-glycinecarbamoyl adenine ( $g^6A37$ ), 5-taurinomethyl-2-thiouridine ( $tm^5s^2U34$ ) and 3-(3-amino-3 carboxypropyl)-uridine ( $acp^3U47$ ), lysidine ( $k^2C34$ ), agmatidine ( $agm^2C34$ ) and glutamyl-queuosine ( $gluQ34$ ). These modifications might be relics of ancient code, when the RNA hairpin structures harboring the primordial anticodons (the ancestor of tRNA) were strongly associated, and possibly even 'charged' with amino acids (131–133).

As a rule, codons ending with pyrimidines U3 or C3 are rarely distinguishable in translation and always behave as synonymous codons, limiting the number of tRNAs able to read four codons to a maximum of three species (Figure 10A). Under special genomic selection conditions, such as strong A/T or G/C pressure, whether coupled or not with genome size reduction, the proteosynthetic machinery can further evolve toward a simpler version than the one found in majority of extant biological cells. In the parasitic *Mycoplasmas* like *M. capricolum* and more evidently in mitochondria, both with A/T-rich genome (Figure 10B), fewer isodecoder tRNAs (28 and 35, respectively, instead of 40–45 in most cells) are needed to translate mRNA coding for 20 canonical amino acids. The reduction in the number of distinct tRNAs concerns mainly C34-containing tRNAs in 4-codon boxes and eventually 2:2 decoding boxes with the consequence that now both Met and Trp become encoded by 2 synonymous codons as in mitochondria. More extensive tRNA reduction can occur with A34/G34-containing tRNA as long as the uracil of the remaining U34-containing tRNA is not post-transcriptionally modified (loss of corresponding maturation enzymes). Accordingly, the codon usage in these cells or organelles is strongly biased, with avoidance of NNG codons (see Supplementary Figure S4 and (59,60,117)). Conversely in some *Micrococcus* strains like *M. luteus*, with a G/C-rich genome (Figure 10B), reduction of tRNA applies only to U34-containing tRNAs and the usage of codons is almost exclusively NNC and NNG. The main difference between *M. luteus* (Figure 10B) and primordial cells is that 20 amino acids, and thus 2:2 codon boxes are used, while in the ancestral cells only a few amino acids, corresponding to only a few G/C-containing codons, prevailed.

Briefly, after an initial cooptation of a few abiotic amino acids into the code, further evolution of the primordial RNA with amino acids occurred through intricate but interdependent stepwise expansion of the amino acid biosynthetic pathways, coupled with the evolution of aminoacyl-tRNA synthetases and tRNA modification enzymes. As a



**Figure 10.** Hypothetical stepwise evolution of the genetic code and translation machinery in Bacteria, Eukarya, Archaea and mitochondria. (A) Starting from highly G+C rich small RNA pieces and a few abiotic amino acids, the primordial translation system evolved by extending its coding capacity by

result, codon:anticodon pairs with intrinsically weaker interactions provided greater diversity in amino acids (chemical structures and global polarities). These pairs also concentrate a large number of deviations from the genetic code or codon reassignments, and possibly correspond to best candidates for further insertions of additional non-canonical amino acids of special technical interest. The elucidation of mechanisms and the development of tools for the recoding of sense codons are currently actively being researched in order to be able to produce proteins with a variety of non-natural amino acids at several positions (134,135). The rules delineated here should, with further work, help in this development.

## CONCLUSIONS

Cells from different organisms, or the same organism but from different tissues, generally have different population of functional tRNAs. Both the total number of tRNA species with distinct anticodons (tRNA repertoire or tRNome) and the relative amounts of each individual isoacceptor species (isodecoders) differ much from one type of cell to another (69,136). Similarly, the frequency with which each codon is translated (codon usage or codon preference) varies depending on the origin of cell and on the levels of gene expression. This is due partly to varying amino acid composition of the proteins synthesized and most importantly to the different use of synonymous codons (codon bias) that is dependent on the global G+C content of the mRNAs (125,137–142).

Inspection and comparison of the mappings of the extended anticodon concept wheels illustrates how, despite the constraining process of decoding and the use of an identical genetic code (except for Met and Trp in certain types of cells like mycoplasma and mitochondria), a large variety of tRNA repertoires and codon usage is still allowed and prevalent. The interpretation of this biological diversity in molecular terms is a daunting task. However, it should be

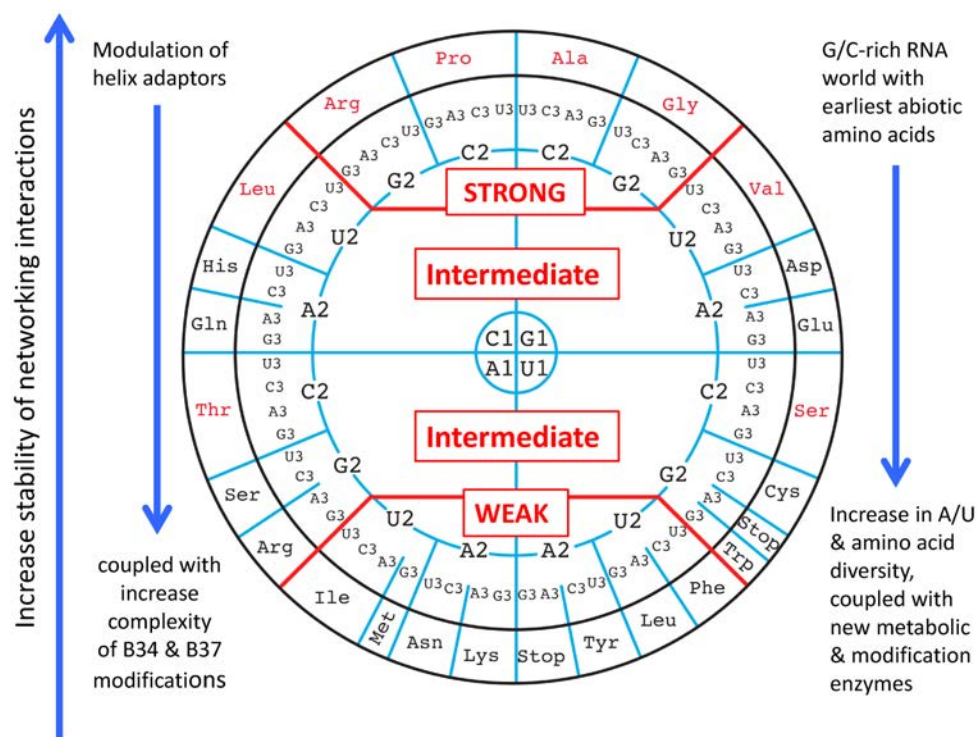
clear that evaluating the frequency of a given codon in individual mRNAs or, as is usually done for statistical reasons, of bulk cellular mRNAs, is not sufficient, and that the exact composition of the tRNA repertoire in a given cell or organism (see below) has also to be determined. Complications arise from the many diverse underlying evolving molecular mechanisms that preside over the biogenesis, stability and functional regulation of each component (tRNAs, mRNAs, snoRNA in Eukarya and Archaea, possibly regulatory RNAs) of the translation machinery.

The majority of individual codons in mRNAs can be translated by either a cognate aa-tRNA harboring the three complementary base pairs or by one or several near-cognate aa-tRNA carrying the same amino acid differing by the base-34 or any other important base of the extended anticodon region. The main advantage of representing data related to such complex decoding system in the form of a wheel of genetic code is that all sequence co-variances can be visualized within an internal energy-based logic, for each anticodon of the 61 almost universally used sense codons in a given organism. The concomitant emergence of machineries allowing post-transcriptional enzymatic modifications of bases and ribose at position 34 of especially tRNA species of the split 2:2- and 3:1-codon boxes also play a determinant role, not only to help codon–anticodon stabilization but also to guarantee a better discrimination between the two sets of Y-ending and R-ending codons. Furthermore, for smooth and regular decoding and ribosomal progression along the correct frame of mRNAs, optimal and uniform binding of the various aminoacyl-tRNAs independently of codon types is required (Figure 11).

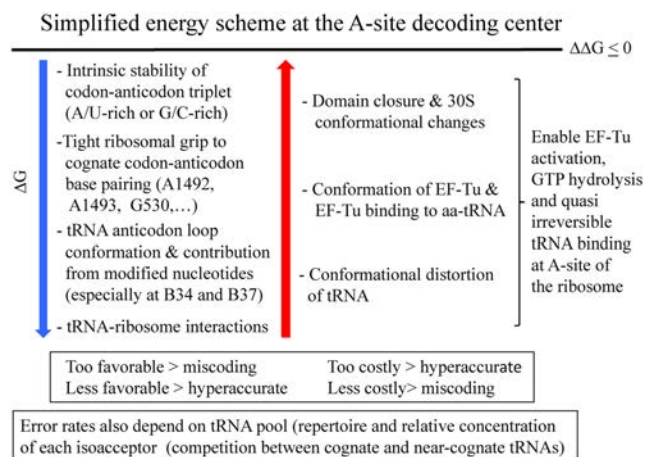
The codon usage, the number of tRNAs in each isoacceptor families and the numbers of isodecoder tRNAs all vary with cell type and cellular differentiation state but, despite the overwhelming complexity and potential ambiguity they introduce, as an ensemble they still have to comply with the networks of interactions which act at the decoding A site (Figure 12).

progressive introduction of A and U in both the template and decoder RNAs (see also (122)). The final decoding machinery as we know to date (and illustrated on the left panel of the figure) results from a long and complex intricate stepwise coevolution with the emergence of metabolically generated amino acids (up to 20), duplication and speciation of proto-tRNAs with distinct anticodons (up to 40–45 to date), post-transcriptional modification enzymes (up to more than 100 known to date), new amino acid tRNA synthetases (up to 20), complexification of the ribosomal architecture (rRNAs and r-proteins) and the introduction of additional protein factors allowing the extension and ultimate tuning the efficacy and accuracy of the decoding capability of 61/62 sense (cognate and near cognate) codons with in addition 2 to 3 terminators for 20 proteogenomic natural amino acids. Pyrrolysine and selenocysteine have been excluded from our analysis, as well as the situation of the special tRNA<sup>Met</sup> involved in the initiation of protein synthesis that obviously arose later during cell evolution. Emphasis is given to the importance of B34 modifications (indicated in red) that allows the segregation of the A/U-rich 4-codon boxes into split decoding boxes 2:2 and 3:1 with subsequent additional amino acids to enter the code. (B) Following genomic selection conditions such as directional constraints or mutational pressure on codons (strong G/C as in *M. luteus* (148) or strong A/T combined with drastic genome size reduction as in mammalian mitochondria (57,60) or the minimalist bacteria *M. capricolum* (58,59,149), simplifications of the usual translational decoding system are evident, while preserving the split decoding boxes because of the need to encode for 20 amino acids. Less tRNA species with distinct anticodons (isodecoders) are required (22 for human mitochondria, 28 in *M. capricolum* and 29 in *M. luteus*, again not including tRNA<sup>Met</sup> initiator). However, remarkably the only remaining B34 modifications found in natural tRNAs are those related to the split 2:2 and/or 3:1 decoding boxes. More information in relation to the corresponding tRNA modification enzymes of a given tRNA repertoire in the subgroup of Mollicutes (mainly mycoplasmas) can be found in (150). The situation of *M. luteus* is remarkable for its total absence of U34-containing tRNAs while in mycoplasma and mitochondria U34-containing tRNA are critical. Codon usage correlates with tRNA repertoire and amino acid type (125,137–142). Symbols for each acronym of the most important modified nucleotides are indicated within the figure. In parenthesis b, e, a, m means bacteria, eukaryotes, archaea and mitochondria respectively. More details about their chemical structures are shown in Figure 7 and Supplementary Figure S4. Within each decoding boxes, blue arrows correspond to the most dominant codon:anticodon pairs. Only B34 are indicated. When symbols of B34 modification are in parenthesis, it means that the unmodified version of B34 is used only in certain decoding boxes, while the modified version(s) is (are) used in other decoding boxes (for details see in Figure 7 for *E. coli* and Supplementary Figure S4 for *H. volcanii*, *S. cerevisiae*, *M. capricolum* and human mitochondria, respectively). When B3 is in bold, it means that the codon usage corresponding to the particular codon is dominant over the other near cognate codons of the same 4- or 2-decoding boxes, while B34 indicated in regular italics correspond to rare codons (for details see as above, Figure 7 for *E. coli* and Supplementary Figure S4 for the other organisms analyzed).





**Figure 11.** Summary scheme illustrating the mapping of the energetics and evolutionary history on the wheel organization of the genetic code. Right vertical arrows: from bottom to top, there is an increase in the strengths of the networking interactions that is coupled with an increase of base modifications at B34 and B37 from top to bottom. Left vertical arrows at the right and left sides: the evolution from the primordial G/C-rich to A/U-rich codon/anticodons triplets required base modifications and the related metabolic enzymatic activities.



**Figure 12.** Summary figure: an approximate and simplified energy scheme at the A site decoding center illustrates how the favorable and costly contributions top the free energies and compensate to maintain a smooth and regular translation process with minor final variations in free energy. On the wheel representation, the energies at the left can be mapped. However, not all energies can be mapped; for example, the interactions between the ribosome and parts other than the anticodon hairpin, the conformational distortions or alternative states of tRNAs, and the energies associated with ribosomal movements.

The multiplicity, diversity and constrained neutrality of the non-covalent and weak molecular interactions between the mRNA and tRNAs within the ribosomal grip allow for such a continuing and divergent evolution toward either

complexity or simplification in the genetic code that clearly cannot be viewed as frozen. Changing the meaning of a codon in a genome implies a few coordinated adjustments of various elements of the translation apparatus that necessarily result from a multistep evolutionary process. Thorough rationalization and understanding of the molecular mechanisms underlying the decoding process and the errors in decoding fidelity should contribute to further developments and progress in engineering of the genetic code (85,143).

The main insights gained by the present new organization of the genetic code can be summarized in the following way.

- The new organization of the genetic code clearly segregates 4-codon boxes from 2:2- and 3:1-codon boxes.
- The 2:2 codon boxes present the weakest intrinsic stabilities and, among them, the STOP codons are the least stable.
- The absence of symmetry in the wheel representation is not arbitrary. It correlates with all known parameters acting at the decoding site: the structure of the anticodon loop and stem, the intrinsic stability of the codon/anticodon interactions, the molecular grip exerted by the ribosome at the decoding site and the presence of characteristic modified nucleotides.
- The co-variances of base-pair 32–38 and base-opposition 31–39 modulate the adaptation of the codon/anticodon triplet to the anticodon stem and the body of the tRNA.
- The complexity of the chemical modifications at base 34 and 37 presents a correlation with AU-richness in the

codon/anticodon pairs. The system requires fewer modifications with GC-rich codon/anticodon pairs.

- The structurally based network of interactions results in an uniform decoding of all tRNAs that can adapt to the cellular constraints.
- The structural and unsymmetrical role is played by residue 34 by exploiting a great diversity of mechanisms: tautomeric or (de)protonation forms, *syn/anti* conformational changes, and non-Watson–Crick pairs.
- Modification of U34 is mandatory for tRNAs belonging to split 2:2-codon boxes: it allows reading only R-ending codons and possibly also the elimination of a C34-containing tRNA that reads the cognate G-ending codon. Conversely, all four codons of an unsplit 4-codon box can be translated by a single tRNA harboring a non-modified U34, but with a preference for A- and U-ending codons (biased codon usage).
- Anticodon and codon are two sides of the same coin; the relative usage is necessarily correlated or co-adapted.
- In the evolution of the triplet code, a major role was played by thermostability of codon–anticodon interactions. The malleability and, thus, the evolution of the genetic code is primarily based on genomic GC-content, with the progressive introduction of U/A together with tRNA modifications (especially at U34). This leads to a great diversity of number and type of the tRNA repertoire (and thus also of codon usage) depending on GC-content of the genome.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

Drawings of the genetic code as a wheel in various formats are available at <http://www-ibmc.u-strasbg.fr/spip-arn/rubrique138?lang=fr> or upon requests to the authors. The authors would like to dedicate this article to two friends and colleagues recently deceased, Knut Nierhaus (Berlin) and Mathias Sprinzl (Bayreuth), who both contributed significantly to our understanding of ribosomal translation. The authors wish to thank numerous colleagues for constructive discussions on decoding during the preparation of this article, for help with the drawings, and for critical reading of the MS (J. Chihade, P. Farabaugh, A. Krol, D. Moras, M. Ryckelynck, M. Sissler, G. Yusupova, M. Yusupov).

## FUNDING

LABEX: ANR-10-LABX-0036\_NETRINA and benefits from a funding from the state managed by the French National Research Agency as part of the Investments for the future program (to E.W.). Funding for open access charge: LABEX: ANR-10-LABX-0036\_NETRINA.

*Conflict of interest statement.* None declared.

## REFERENCES

1. Watanabe, K. and Yokobori, S. (2011) tRNA modification and genetic code variations in animal mitochondria. *J. Nucleic Acids*, **2011**, 623095.
2. Ling, J., O'Donoghue, P. and Söll, D. (2015) Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. *Nat. Rev. Microbiol.*, **13**, 707–721.
3. Bezerra, A.R., Guimaraes, A.R. and Santos, M.A. (2015) Non-standard genetic codes define new concepts for protein engineering. *Life (Basel)*, **5**, 1610–1628.
4. Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F. and O'Neal, C. (1965) RNA codewords and protein synthesis. VII. On the general nature of the RNA code. *Proc. Natl. Acad. Sci. U.S.A.*, **53**, 1161–1168.
5. Crick, F.H. (1968) The origin of the genetic code. *J. Mol. Biol.*, **38**, 367–379.
6. Woese, C.R., Dugre, D.H., Saxinger, W.C. and Dugre, S.A. (1966) The molecular basis for the genetic code. *Proc. Natl. Acad. Sci. U.S.A.*, **55**, 966–974.
7. Agris, P.F., Vendeix, F.A. and Graham, W.D. (2007) tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.*, **366**, 1–13.
8. Lehmann, J. (2000) Physico-chemical constraints connected with the coding properties of the genetic system. *J. Theor. Biol.*, **202**, 129–144.
9. Jimenez-Montano, M.A. (2009) The fourfold way of the genetic code. *Biosystems*, **98**, 105–114.
10. Danckwerts, H.J. and Neubert, D. (1975) Symmetries of genetic code-doublets. *J. Mol. Evol.*, **5**, 327–332.
11. Wilhelm, T. and Nikolajewa, S. (2004) A new classification scheme of the genetic code. *J. Mol. Evol.*, **59**, 598–605.
12. Agris, P.F. (2008) Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep.*, **9**, 629–635.
13. Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A., Fabris, D. and Agris, P.F. (2011) The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res.*, **39**, D195–D201.
14. Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M. *et al.* (2013) MODOMICS: a database of RNA modification pathways–2013 update. *Nucleic Acids Res.*, **41**, D262–D267.
15. Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. and Yusupova, G. (2012) A new understanding of the decoding principle on the ribosome. *Nature*, **484**, 256–259.
16. Jenner, L., Demeshkina, N., Yusupova, G. and Yusupov, M. (2010) Structural rearrangements of the ribosome at the tRNA proofreading step. *Nat. Struct. Mol. Biol.*, **17**, 1072–1078.
17. Ogle, J.M., Brodersen, D.E., Clemons, W.M. Jr, Tarry, M.J., Carter, A.P. and Ramakrishnan, V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science*, **292**, 897–902.
18. Ogle, J.M. and Ramakrishnan, V. (2005) Structural insights into translational fidelity. *Annu. Rev. Biochem.*, **74**, 129–177.
19. Yusupova, G.Z., Yusupov, M.M., Cate, J.H. and Noller, H.F. (2001) The path of messenger RNA through the ribosome. *Cell*, **106**, 233–241.
20. Wohlgemuth, I., Pohl, C., Mittelstaet, J., Konevega, A.L. and Rodnina, M.V. (2011) Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Philos. Trans. R Soc. Lond. B Biol. Sci.*, **366**, 2979–2986.
21. Fahlman, R.P., Dale, T. and Uhlenbeck, O.C. (2004) Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Mol. Cell*, **16**, 799–805.
22. Zhang, J., Jeong, K.W., Johansson, M. and Ehrenberg, M. (2015) Accuracy of initial codon selection by aminoacyl-tRNAs on the mRNA-programmed bacterial ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 9602–9607.
23. Crick, F.H. (1966) Codon–anticodon pairing: the wobble hypothesis. *J. Mol. Biol.*, **19**, 548–555.
24. Nasvall, S.J., Chen, P. and Björk, G.R. (2007) The wobble hypothesis revisited: uridine-5-oxyacetic acid is critical for reading of G-ending codons. *RNA*, **13**, 2151–2164.
25. Yamada, Y., Matsugi, J. and Ishikura, H. (2003) tRNA<sup>I</sup>Ser(G34) with the anticodon GGA can recognize not only UCC and UCU codons but also UCA and UCG codons. *Biochim. Biophys. Acta*, **1626**, 75–82.
26. Lagerkvist, U. (1981) Unorthodox codon reading and the evolution of the genetic code. *Cell*, **23**, 305–306.

27. Murphy, F.V.t., Ramakrishnan, V., Malkiewicz, A. and Agris, P.F. (2004) The role of modifications in codon discrimination by tRNA(Lys)UUU. *Nat. Struct. Mol. Biol.*, **11**, 1186–1191.
28. Murphy, F.V.t. and Ramakrishnan, V. (2004) Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat. Struct. Mol. Biol.*, **11**, 1251–1252.
29. Weixlbaumer, A., Murphy, F.V.t., Dziergowska, A., Malkiewicz, A., Vendeix, F.A., Agris, P.F. and Ramakrishnan, V. (2007) Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nat. Struct. Mol. Biol.*, **14**, 498–502.
30. Kurata, S., Weixlbaumer, A., Ohtsuki, T., Shimazaki, T., Wada, T., Kirino, Y., Takai, K., Watanabe, K., Ramakrishnan, V. and Suzuki, T. (2008) Modified uridines with C5-methylene substituents at the first position of the tRNA anticodon stabilize U-G wobble pairing during decoding. *J. Biol. Chem.*, **283**, 18801–18811.
31. Vendeix, F.A., Murphy, F.V.t., Cantara, W.A., Leszczynska, G., Gustilo, E.M., Sproat, B., Malkiewicz, A. and Agris, P.F. (2012) Human tRNA(Lys3)(UUU) is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. *J. Mol. Biol.*, **416**, 467–485.
32. Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M. and Yusupova, G. (2015) Structural insights into the translational infidelity mechanism. *Nat. Commun.*, **6**, 7251.
33. Lagerkvist, U. (1978) 'Two out of three': an alternative method for codon reading. *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 1759–1762.
34. Lehmann, J. and Libchaber, A. (2008) Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon. *RNA*, **14**, 1264–1269.
35. Chen, J.L., Dishler, A.L., Kennedy, S.D., Yildirim, I., Liu, B., Turner, D.H. and Serra, M.J. (2012) Testing the nearest neighbor model for canonical RNA base pairs: revision of GU parameters. *Biochemistry*, **51**, 3508–3522.
36. Olejniczak, M. and Uhlenbeck, O.C. (2006) tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie*, **88**, 943–950.
37. Wohlgenuth, I., Pohl, C. and Rodnina, M.V. (2010) Optimization of speed and accuracy of decoding in translation. *EMBO J.*, **29**, 3701–3709.
38. Grosjean, H.J., de Henau, S. and Crothers, D.M. (1978) On the physical basis for ambiguity in genetic coding interactions. *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 610–614.
39. Houssier, C. and Grosjean, H. (1985) Temperature jump relaxation studies on the interactions between transfer RNAs with complementary anticodons. The effect of modified bases adjacent to the anticodon triplet. *J. Biomol. Struct. Dyn.*, **3**, 387–408.
40. Jenner, L.B., Demeshkina, N., Yusupova, G. and Yusupov, M. (2010) Structural aspects of messenger RNA reading frame maintenance by the ribosome. *Nat. Struct. Mol. Biol.*, **17**, 555–560.
41. Selmer, M., Dunham, C.M., Murphy, F.V.t., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
42. Fernandez, I.S., Ng, C.L., Kelley, A.C., Wu, G., Yu, Y.T. and Ramakrishnan, V. (2013) Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature*, **500**, 107–110.
43. Schmeing, T.M., Voorhees, R.M., Kelley, A.C. and Ramakrishnan, V. (2011) How mutations in tRNA distant from the anticodon affect the fidelity of decoding. *Nat. Struct. Mol. Biol.*, **18**, 432–436.
44. Doherty, E.A., Batey, R.T., Masquida, B. and Doudna, J.A. (2001) A universal mode of helix packing in RNA. *Nat. Struct. Biol.*, **8**, 339–343.
45. Auffinger, P. and Westhof, E. (2001) An extended structural signature for the tRNA anticodon loop. *RNA*, **7**, 334–341.
46. von Ahsen, U., Green, R., Schroeder, R. and Noller, H.F. (1997) Identification of 2'-hydroxyl groups required for interaction of a tRNA anticodon stem-loop region with the ribosome. *RNA*, **3**, 49–56.
47. Khade, P.K., Shi, X. and Joseph, S. (2013) Steric complementarity in the decoding center is important for tRNA selection by the ribosome. *J. Mol. Biol.*, **425**, 3778–3789.
48. Fahlman, R.P., Olejniczak, M. and Uhlenbeck, O.C. (2006) Quantitative analysis of deoxynucleotide substitutions in the codon-anticodon helix. *J. Mol. Biol.*, **355**, 887–892.
49. Hoernes, T.P., Clementi, N., Faserl, K., Glasner, H., Breuker, K., Lindner, H., Huttenhofer, A. and Erlacher, M.D. (2016) Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res.*, **44**, 852–862.
50. Zaher, H.S. and Green, R. (2010) Hyperaccurate and error-prone ribosomes exploit distinct mechanisms during tRNA selection. *Mol. Cell*, **39**, 110–120.
51. Demirci, H., Wang, L., Murphy, F.V.t., Murphy, E.L., Carr, J.F., Blanchard, S.C., Jögl, G., Dahlberg, A.E. and Gregory, S.T. (2013) The central role of protein S12 in organizing the structure of the decoding site of the ribosome. *RNA*, **19**, 1791–1801.
52. Agarwal, D., Gregory, S.T. and O'Connor, M. (2011) Error-prone and error-restrictive mutations affecting ribosomal protein S12. *J. Mol. Biol.*, **410**, 1–9.
53. Ogle, J.M., Murphy, F.V., Tarry, M.J. and Ramakrishnan, V. (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, **111**, 721–732.
54. Polikanov, Y.S., Melnikov, S.V., Söll, D. and Steitz, T.A. (2015) Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat. Struct. Mol. Biol.*, **22**, 342–344.
55. Sharma, S. and Lafontaine, D.L. (2015) 'View From A Bridge': A new perspective on eukaryotic rRNA base modification. *Trends Biochem. Sci.*, **40**, 560–575.
56. Jiang, J., Seo, H. and Chow, C.S. (2016) Post-transcriptional modifications modulate rRNA structure and ligand interactions. *Acc. Chem. Res.*, **49**, 893–900.
57. Bonitz, S.G., Berlani, R., Coruzzi, G., Li, M., Macino, G., Nobrega, F.G., Nobrega, M.P., Thalenfeld, B.E. and Tzagoloff, A. (1980) Codon recognition rules in yeast mitochondria. *Proc. Natl. Acad. Sci. U.S.A.*, **77**, 3167–3170.
58. Samuelsson, T., Guindy, Y.S., Lustig, F., Boren, T. and Lagerkvist, U. (1987) Apparent lack of discrimination in the reading of certain codons in *Mycoplasma mycoides*. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 3166–3170.
59. Andachi, Y., Yamao, F., Muto, A. and Osawa, S. (1989) Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in *Mycoplasma capricolum*. Resemblance to mitochondria. *J. Mol. Biol.*, **209**, 37–54.
60. Suzuki, T. and Suzuki, T. (2014) A complete landscape of post-transcriptional modifications in mammalian mitochondrial tRNAs. *Nucleic Acids Res.*, **42**, 7346–7357.
61. Yarus, M. (1982) Translational efficiency of transfer RNAs: uses of an extended anticodon. *Science*, **218**, 646–652.
62. Cochella, L. and Green, R. (2005) An active role for tRNA in decoding beyond codon:anticodon pairing. *Science*, **308**, 1178–1180.
63. Dale, T. and Uhlenbeck, O.C. (2005) Amino acid specificity in translation. *Trends Biochem. Sci.*, **30**, 659–665.
64. Grosjean, H., de Crécy-Lagard, V. and Marck, C. (2010) Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. *FEBS Lett.*, **584**, 252–264.
65. Machnicka, M.A., Olchowik, A., Grosjean, H. and Bujnicki, J.M. (2014) Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol.*, **11**, 1619–1629.
66. Barraud, P., Schmitt, E., Mechulam, Y., Dardel, F. and Tisne, C. (2008) A unique conformation of the anticodon stem-loop is associated with the capacity of tRNA<sup>fMet</sup> to initiate protein synthesis. *Nucleic Acids Res.*, **36**, 4894–4901.
67. Motorin, Y. and Helm, M. (2010) tRNA stabilization by modified nucleotides. *Biochemistry*, **49**, 4934–4944.
68. Davis, D.R. (1995) Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.*, **23**, 5020–5026.
69. Marck, C. and Grosjean, H. (2002) tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA*, **8**, 1189–1232.
70. Auffinger, P. and Westhof, E. (1999) Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes. *J. Mol. Biol.*, **292**, 467–483.
71. Westhof, E., Dumas, P. and Moras, D. (1983) Loop stereochemistry and dynamics in transfer RNA. *J. Biomol. Struct. Dyn.*, **1**, 337–355.
72. Konevega, A.L., Soboleva, N.G., Makhno, V.I., Semenov, Y.P., Wintermeyer, W., Rodnina, M.V. and Katunin, V.I. (2004) Purine



- bases at position 37 of tRNA stabilize codon-anticodon interaction in the ribosomal A site by stacking and Mg<sup>2+</sup>-dependent interactions. *RNA*, **10**, 90–101.
73. Kawai, G., Yamamoto, Y., Kamimura, T., Masegi, T., Sekine, M., Hata, T., Iimori, T., Watanabe, T., Miyazawa, T. and Yokoyama, S. (1992) Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry*, **31**, 1040–1046.
  74. Davis, D.R. and Durant, P.C. (1999) Nucleoside modifications affect the structure and stability of the anticodon of tRNA(Lys,3). *Nucleosides Nucleotides*, **18**, 1579–1581.
  75. Davis, D.R. (1998) Biophysical and conformational properties of modified nucleotides in RNA. In: Grosjean, H. and Benne, R. (eds). *Modification and Editing of RNA*. American Society for Microbiology Press, Washington D.C., pp. 85–102.
  76. Juhling, F., Morl, M., Hartmann, R.K., Sprinzl, M., Stadler, P.F. and Putz, J. (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.*, **37**, D159–D162.
  77. Ledoux, S., Olejniczak, M. and Uhlenbeck, O.C. (2009) A sequence element that tunes Escherichia coli tRNA(Ala)(GGC) to ensure accurate decoding. *Nat. Struct. Mol. Biol.*, **16**, 359–364.
  78. Murakami, H., Ohta, A. and Suga, H. (2009) Bases in the anticodon loop of tRNA(Ala)(GGC) prevent misreading. *Nat. Struct. Mol. Biol.*, **16**, 353–358.
  79. Shepotinovskaya, I. and Uhlenbeck, O.C. (2013) tRNA residues evolved to promote translational accuracy. *RNA*, **19**, 510–516.
  80. Curran, J.F. and Yarus, M. (1987) Reading frame selection and transfer RNA anticodon loop stacking. *Science*, **238**, 1545–1550.
  81. Waas, W.F., Druzina, Z., Hanan, M. and Schimmel, P. (2007) Role of a tRNA base modification and its precursors in frameshifting in eukaryotes. *J. Biol. Chem.*, **282**, 26026–26034.
  82. Knight, R.D., Freeland, S.J. and Landweber, L.F. (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat. Rev. Genet.*, **2**, 49–58.
  83. Miranda, I., Silva, R. and Santos, M.A. (2006) Evolution of the genetic code in yeasts. *Yeast*, **23**, 203–213.
  84. Ambrogelly, A., Palioura, S. and Söll, D. (2007) Natural expansion of the genetic code. *Nat. Chem. Biol.*, **3**, 29–35.
  85. Mukai, T., Englert, M., Tripp, H.J., Miller, C., Ivanova, N.N., Rubin, E.M., Kyrpides, N.C. and Söll, D. (2016) Facile recoding of selenocysteine in nature. *Angew. Chem. Int. Ed. Engl.*, **55**, 5337–5341.
  86. Parker, J. (1989) Errors and alternatives in reading the universal genetic code. *Microbiol. Rev.*, **53**, 273–298.
  87. Manickam, N., Joshi, K., Bhatt, M.J. and Farabaugh, P.J. (2016) Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Res.*, **44**, 1871–1881.
  88. Björk, G.R. and Hagervall, T.G. (2014) Transfer RNA modification. *EcoSal Plus*, **6**, doi:10.1128/ecosalplus.ESP-0007-2013.
  89. Brocker, M.J., Ho, J.M., Church, G.M., Söll, D. and O'Donoghue, P. (2014) Recoding the genetic code with selenocysteine. *Angew. Chem. Int. Ed. Engl.*, **53**, 319–323.
  90. Lehman, N. (2001) Molecular evolution: Please release me, genetic code. *Curr. Biol.*, **11**, R63–66.
  91. Namy, O., Rousset, J.P., Naphine, S. and Brierley, I. (2004) Reprogrammed genetic decoding in cellular gene expression. *Mol. Cell*, **13**, 157–168.
  92. Takemoto, C., Spremulli, L.L., Benkowski, L.A., Ueda, T., Yokogawa, T. and Watanabe, K. (2009) Unconventional decoding of the AUA codon as methionine by mitochondrial tRNA<sup>Met</sup> with the anticodon f5CAU as revealed with a mitochondrial in vitro translation system. *Nucleic Acids Res.*, **37**, 1616–1627.
  93. Cantara, W.A., Murphy, F.V.t., Demirci, H. and Agris, P.F. (2013) Expanded use of sense codons is regulated by modified cytidines in tRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 10964–10969.
  94. Suzuki, T., Miyauchi, K., Suzuki, T., Yokobori, S., Shigi, N., Kondow, A., Takeuchi, N., Yamagishi, A. and Watanabe, K. (2011) Taurine-containing uridine modifications in tRNA anticodons are required to decipher non-universal genetic codes in ascidian mitochondria. *J. Biol. Chem.*, **286**, 35494–35498.
  95. Ohira, T., Suzuki, T., Miyauchi, K., Suzuki, T., Yokobori, S., Yamagishi, A. and Watanabe, K. (2013) Decoding mechanism of non-universal genetic codes in Loligo bleekeri mitochondria. *J. Biol. Chem.*, **288**, 7645–7652.
  96. Schwartz, S., Bernstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., Leon-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S. *et al.* (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell*, **159**, 148–162.
  97. Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M. *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, **485**, 201–206.
  98. Hudson, B.H. and Zaher, H.S. (2015) O6-Methylguanosine leads to position-dependent effects on ribosome speed and fidelity. *RNA*, **21**, 1648–1659.
  99. Hudson, G.A., Bloomington, R.J. and Znosko, B.M. (2013) Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. *RNA*, **19**, 1474–1482.
  100. Karijolic, J. and Yu, Y.T. (2011) Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature*, **474**, 395–398.
  101. Tomita, K., Ueda, T. and Watanabe, K. (1999) The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. *Nucleic Acids Res.*, **27**, 1683–1689.
  102. Svidritskiy, E., Madireddy, R. and Korostelev, A.A. (2016) Structural basis for translation termination on a pseudouridylated stop codon. *J. Mol. Biol.*, **428**, 2228–2236.
  103. Eigen, M. and Winkler-Oswatitsch, R. (1981) Transfer-RNA, an early gene? *Naturwissenschaften*, **68**, 282–292.
  104. Eigen, M. and Winkler-Oswatitsch, R. (1981) Transfer-RNA: the early adaptor. *Naturwissenschaften*, **68**, 217–228.
  105. Ikehara, K. (2002) Origins of gene, genetic code, protein and life: comprehensive view of life systems from a GNC-SNS primitive genetic code hypothesis. *J. Biosci.*, **27**, 165–186.
  106. Trifonov, E. and Berezovsky, I. (2002) Molecular evolution from abiotic scratch. *FEBS Lett.*, **527**, 1–4.
  107. Yarus, M., Caporaso, J.G. and Knight, R. (2005) Origins of the genetic code: the escaped triplet theory. *Annu. Rev. Biochem.*, **74**, 179–198.
  108. Yarus, M., Widmann, J.J. and Knight, R. (2009) RNA-amino acid binding: a stereochemical era for the genetic code. *J. Mol. Evol.*, **69**, 406–429.
  109. Wong, J.T. (1975) A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. U.S.A.*, **72**, 1909–1912.
  110. Ronneberg, T.A., Landweber, L.F. and Freeland, S.J. (2000) Testing a biosynthetic theory of the genetic code: fact or artifact? *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 13690–13695.
  111. Di Giulio, M. (2008) An extension of the coevolution theory of the origin of the genetic code. *Biol. Direct*, **3**, 37.
  112. Klipcan, L. and Safo, M. (2004) Amino acid biogenesis, evolution of the genetic code and aminoacyl-tRNA synthetases. *J. Theor. Biol.*, **228**, 389–396.
  113. Delarue, M. (2007) An asymmetric underlying rule in the assignment of codons: possible clue to a quick early evolution of the genetic code via successive binary choices. *RNA*, **13**, 161–169.
  114. Smith, T.F. and Hartman, H. (2015) The evolution of Class II Aminoacyl-tRNA synthetases and the first code. *FEBS Lett.*, **589**, 3499–3507.
  115. Freeland, S.J., Wu, T. and Keulmann, N. (2003) The case for an error minimizing standard genetic code. *Orig. Life Evol. Biosph.*, **33**, 457–477.
  116. Buhrman, H., van der Gulik, P.T., Klau, G.W., Schaffner, C., Speijer, D. and Stougie, L. (2013) A realistic model under which the genetic code is optimal. *J. Mol. Evol.*, **77**, 170–184.
  117. Osawa, S., Jukes, T.H., Watanabe, K. and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol. Rev.*, **56**, 229–264.
  118. Schultz, D.W. and Yarus, M. (1996) On malleability in the genetic code. *J. Mol. Evol.*, **42**, 597–601.
  119. Smets, B.F. and Barkay, T. (2005) Horizontal gene transfer: perspectives at a crossroads of scientific disciplines. *Nat. Rev. Microbiol.*, **3**, 675–678.
  120. Shackleton, L.A. and Holmes, E.C. (2008) The role of alternative genetic codes in viral evolution and emergence. *J. Theor. Biol.*, **254**, 128–134.

121. Vetsigian, K., Woese, C. and Goldenfeld, N. (2006) Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 10696–10701.
122. Trifonov, E.N. (2000) Consensus temporal order of amino acids and evolution of the triplet code. *Gene*, **261**, 139–151.
123. Atkins, J.F. and Baranov, P.V. (2010) The distinction between recoding and codon reassignment. *Genetics*, **185**, 1535–1536.
124. van der Gulik, P.T. and Hoff, W.D. (2011) Unassigned codons, nonsense suppression, and anticodon modifications in the evolution of the genetic code. *J. Mol. Evol.*, **73**, 59–69.
125. Novoa, E.M., Pavon-Eternod, M., Pan, T. and Ribas de Pouplana, L. (2012) A role for tRNA modifications in genome structure and codon usage. *Cell*, **149**, 202–213.
126. Björk, G.R., Jacobsson, K., Nilsson, K., Johansson, M.J., Bystrom, A.S. and Persson, O.P. (2001) A primordial tRNA modification required for the evolution of life? *EMBO J.*, **20**, 231–239.
127. Thiaville, P.C., El Yacoubi, B., Kohrer, C., Thiaville, J.J., Deutsch, C., Iwata-Reuyl, D., Bacusmo, J.M., Armengaud, J., Bessho, Y., Wetzel, C. *et al.* (2015) Essentiality of threonylcarbamoyladenine (t(6) A), a universal tRNA modification, in bacteria. *Mol. Microbiol.*, **98**, 1199–1221.
128. Sample, P.J., Koreny, L., Paris, Z., Gaston, K.W., Rubio, M.A., Fleming, I.M., Hinger, S., Horakova, E., Limbach, P.A., Lukes, J. *et al.* (2015) A common tRNA modification at an unusual location: the discovery of wyosine biosynthesis in mitochondria. *Nucleic Acids Res.*, **43**, 4262–4273.
129. Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U.S.A.*, **87**, 4576–4579.
130. McKenney, K.M. and Alfonzo, J.D. (2016) From prebiotics to probiotics: The evolution and functions of tRNA modifications. *Life (Basel)*, **6**, doi:10.3390/life6010013.
131. Di Giulio, M. (1998) Reflections on the origin of the genetic code: a hypothesis. *J. Theor. Biol.*, **191**, 191–196.
132. Grosjean, H., de Crecy-Lagard, V. and Björk, G.R. (2004) Aminoacylation of the anticodon stem by a tRNA-synthetase paralog: relic of an ancient code? *Trends Biochem. Sci.*, **29**, 519–522.
133. Szathmari, E. (1999) The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet.*, **15**, 223–229.
134. Liu, C.C. and Schultz, P.G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.*, **79**, 413–444.
135. Chin, J.W. (2014) Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.*, **83**, 379–408.
136. Goodenbour, J.M. and Pan, T. (2006) Diversity of tRNA genes in eukaryotes. *Nucleic Acids Res.*, **34**, 6137–6146.
137. Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13–34.
138. Grosjean, H. and Fiers, W. (1982) Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene*, **18**, 199–209.
139. Rocha, E.P. (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res.*, **14**, 2279–2286.
140. Ran, W. and Higgs, P.G. (2010) The influence of anticodon-codon interactions and modified bases on codon usage bias in bacteria. *Mol. Biol. Evol.*, **27**, 2129–2140.
141. Quax, T.E., Claassens, N.J., Söll, D. and van der Oost, J. (2015) Codon bias as a means to fine-tune gene expression. *Mol. Cell*, **59**, 149–161.
142. Wald, N., Alroy, M., Botzman, M. and Margalit, H. (2012) Codon usage bias in prokaryotic pyrimidine-ending codons is associated with the degeneracy of the encoded amino acids. *Nucleic Acids Res.*, **40**, 7074–7083.
143. Iwane, Y., Hitomi, A., Murakami, H., Katoh, T., Goto, Y. and Suga, H. (2016) Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. *Nat. Chem.*, **8**, 317–325.
144. Steinberg, S., Misch, A. and Sprinzl, M. (1993) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **21**, 3011–3015.
145. Limbach, P.A., Crain, P.F. and McCloskey, J.A. (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res.*, **22**, 2183–2196.
146. Nakamura, Y., Gojobori, T. and Ikemura, T. (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.*, **28**, 292.
147. Beier, H. and Grimm, M. (2001) Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res.*, **29**, 4767–4782.
148. Kano, A., Andachi, Y., Ohama, T. and Osawa, S. (1991) Novel anticodon composition of transfer RNAs in *Micrococcus luteus*, a bacterium with a high genomic G + C content. Correlation with codon usage. *J. Mol. Biol.*, **221**, 387–401.
149. de Crecy-Lagard, V., Marck, C., Brochier-Armanet, C. and Grosjean, H. (2007) Comparative RNomics and modomics in Mollicutes: prediction of gene function and evolutionary implications. *IUBMB Life*, **59**, 634–658.
150. Grosjean, H., Breton, M., Sirand-Pugnet, P., Tardy, F., Thiaucourt, F., Citti, C., Barre, A., Yoshizawa, S., Fourmy, D., de Crecy-Lagard, V. *et al.* (2014) Predicting the minimal translation apparatus: lessons from the reductive evolution of mollicutes. *PLoS Genet.*, **10**, e1004363.