

# Identity Elements of tRNA as Derived from Information Analysis

Gabriel S. Zamudio<sup>1</sup> · Marco V. José<sup>1</sup>

Received: 3 April 2017 / Accepted: 9 June 2017 /  
Published online: 28 June 2017  
© Springer Science+Business Media B.V. 2017

**Abstract** The decipherment of the tRNA's operational code, known as the identity problem, requires the location of the sites in the tRNA structure that are involved in their correct recognition by the corresponding aminoacyl-tRNA synthetase. In this work, we determine the identity elements of each tRNA isoacceptor by means of the variation of information measure from information theory. We show that all isoacceptors exhibit sites associated with some bases of the anticodon. These sites form clusters that are scattered along the tRNA structure. The clusters determine the identity elements of each tRNA. We derive a catalogue of clustered sites for each tRNA that expands previously reported elements.

**Keywords** Identity elements · Operational code · Anticodon code · tRNA evolution · Information theory

## Introduction

The correct implementation of the genetic code comprehends an extensive and complex set of biological interactions inside the cell. At the core of the system lies the transfer RNA (tRNA), a key molecule driving the translation process. To preserve this intricate process, the tRNA is equipped with two codes along its structure. The anticodon code that reads the codons of a messenger RNA (mRNA), and a “second” “operational code (De Duve 1988) which is commonly associated to the acceptor stem of the tRNA (Hou and Schimmel 1988). Since the dawn of the mini-helix structure of tRNA (Tamura 2015) 3.5 billion years ago, both codes have coevolved and they are encrypted in its structure. tRNA molecules share a common

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s11084-017-9541-6) contains supplementary material, which is available to authorized users.

---

✉ Marco V. José  
marcojose@biomedicas.unam.mx

<sup>1</sup> Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, CDMX, C.P. 04510 Ciudad de México, Mexico

structure that is recognized by other molecules of the biosynthetic pathways (Arnez and Moras 1997). In the translation process, the tRNA is the only participant which recognizes the codons of a mRNA directly (Cusack 1997). A mature tRNA has a canonical length of 76 bases (Altman 1993), although some isoacceptors (tRNAs with different anticodons for the same amino acid) differ in length, mainly at the variable loop. The anticodon triplet is located at bases 34, 35, and 36. The terminal CCA motif at 3' end, which is added post-transcriptionally (Hou 2010), is the place at which amino acids are attached through an esterification reaction with its cognate aminoacyl-tRNA synthetase (aaRS) that has been previously charged with an amino acid (Arnez and Moras 1997). The “second” genetic code, also called “operational code” (De Duve 1988; Rodin and Ohno 1997), directs the correct identification of tRNA isoacceptors with its respective aaRSs by stereochemical means. The problem of deciphering the operational code is known as “the identity problem”, since the molecules of aaRS must identify the correct tRNA from a pool of similar molecules, without necessarily interacting with the anticodons.

Two protein subfamilies compose the set of aaRSs, Class I, and Class II, with ten proteins each (Eriani et al. 1990). The 20 different aaRSs account for the 20 canonical amino acids, and so, the operational code is nondegenerate (Eriani et al. 1990). In contrast the anticodon code consists of 48 anticodons, since triplets starting with adenine are absent of this code (Guimarães et al. 2008; José et al. 2014). The two classes of aaRSs recognize the acceptor helix of the tRNAs by different approaches: Class I recognizes the minor groove and charges the amino acid at the 2'OH group of the terminal adenosine, while Class II access from the major groove and charges the amino acid in the 3'OH group (Eriani et al. 1990). It has been shown that the specificities between tRNAs and aaRSs coevolved during the formation of the genetic code and they were driven by the hydrophathy of anticodons (Farias et al. 2014a).

Information theory has been used to analyze genetic sequences (Adami 2004, 2012). It has also been used to predict the secondary structure of RNA (Durbin et al. 1998). In this work, we use the measure of variation of information, from information theory, to determine the specific sites in the tRNA structure that are highly related to the anticodon. This measure uses the variation in the gene sequences of a tRNA isoacceptor to locate the sites that contribute to the degeneracy of the isoacceptor's anticodon code. These identity elements determine the recognition process of tRNAs by their respective aaRSs in the translation process.

## Data Sources

The database (Abe et al. 2014) contains curated tRNA genes from the three kingdoms of life. We selected those sequences whose lengths matched the canonical length of 76 nucleotides discarding the variable loop. Redundant sequences were also omitted in order to avoid duplicated sequences that could alter the statistical results. Overall, we analyzed a total of 13,093 gene sequences including all isoacceptors.

## Methods

Information theory is devoted to the quantification, transmission and storage of information. In particular, given two messages, it is possible to determine the information shared by them, and consequently compare them. The variation of information is a measure that determines the

information distance between two messages. This measure is used as a clustering algorithm for data and for comparing different clusters, of the same data (Meila 2003, 2007). The variation of information is a measure that gives a distance between two messages  $X$  and  $Y$ . It is given by the equation  $VI(X, Y) = H(X) + H(Y) - 2I(X, Y)$  where  $H(X)$  is the Shannon's entropy, and the term  $I(X, Y)$  accounts for the mutual information shared between  $X$  and  $Y$  (Meila 2003). The random variables  $X$  and  $Y$ , describe the distribution of characters or symbols in any given message. The results are given in bit units. This measure captures the information needed to describe one variable from previously knowing the other. As the mutual information function is a factor of the variation of information, the pairs of sites close to each other have a general dependence, not necessarily a linear dependence that can be obtained with the correlation function (Li 1990). Dividing the tRNA genes by isoacceptors, and using the variation of information per site, the distribution of nucleotides in a single site of the tRNA was considered as a random variable. Then, it was possible to compute the variation of information distance between any two sites of the same isoacceptor, for each of the 20 amino acids. For the calculations, the terminal CCA was removed from the sequences, as it is a constant motif in all of them. For small sample sizes a correction is applied to the entropy and mutual information function to account for the bias. The approximated error is based on the number of states and the sample size (Li 1990). For the variation of information, the error is approximated by  $\overline{VI(X, Y)} - VI(X, Y) \approx \frac{K^2}{N}$ , where  $\overline{VI(X, Y)}$  stands for the true variation of information, while  $VI(X, Y)$  is the calculated variation of information, the number of states is denoted by  $K$  and  $N$  is the sample size. This error is applied uniformly to all calculated distances. As the error is applied for each isoacceptor, regardless of the sample size, it can be neglected. If the error is considered, the methodology would require obtaining the minimum distances and the same results would be attained, since this distance would coincide with the estimated error.

## Results

If two sites  $x_1$  and  $x_2$  are at distance zero, it means that those sites are clustered and so, the state or occurrence of a base in the site  $x_1$  is completely predictable by the state of  $x_2$ , and vice versa. Hence, the sites involved with the correct aminoacylation of tRNAs, i.e., the identity elements, would be those whose distance with any base with the anticodon is zero. These sites form clusters that are involved in the proper recognition of the corresponding aaRSs. The nucleotide bases present in the sites of a cluster are coordinated. This means that each nucleotide present in a site is derived by the presence of a nucleotide found in another site of the cluster. Table 1 enlists the sets of sites that are fully clustered forming a unit dividing the isoacceptors by its aaRSs class; anticodon positions are marked in red. Notice that some isoacceptors contain multiple sets of predictable sites. Also notice that all the sites in a single set are at distance zero of each other. The presence of multiple clusters in an individual isoacceptor is apparent. Some form Watson-Crick pairs (marked in green) which are relevant to maintain the stability of the secondary and tertiary structure of the tRNA molecules. Other clusters are used to avoid misidentification and mischarging of the tRNA (Giegé et al. 1998). As an example, the clusters for tRNA<sup>Gln</sup> are colored on the secondary structure (Fig. 1). The sites that belong to the same cluster set are colored with the same color. Notice that the tRNA<sup>Gln</sup> possesses three clusters: i) The cluster with the 35, 36 anticodon bases associated with five bases at the TΨC-loop; ii) a

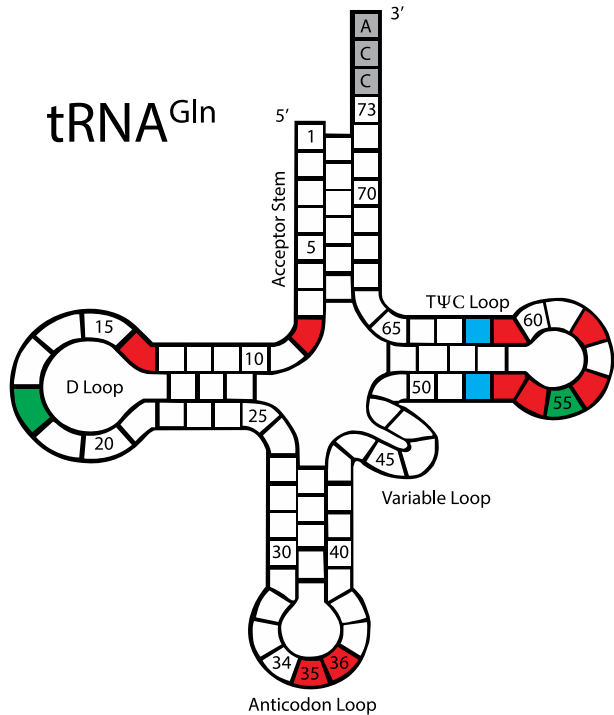
**Table 1** Table of clusters for each tRNA. Each set conforms a cluster of positions. Anticodon bases are in red. Positions forming Watson-Crick pairs are marked in green. The sites that do not present a Watson-Crick pair are marked in black. The symbol Nstands for the sample size of each isoacceptor

tRNA	Class I	
	Unity Clusters	N
tRNA <sup>Arg</sup>	{8,14, <b>35</b> ,54,58}	1249
tRNA <sup>Cys</sup>	{1,8,14,18, <b>34,35,36</b> ,53,54,56,58,61,72}, {2,71}, {3,70}, {5,68}, {7,66}, {10,25}, {52,62}	131
tRNA <sup>Gln</sup>	{8,14, <b>35,36</b> ,53,54,56,58,61}, {18,55}, {52,62}	354
tRNA <sup>Glu</sup>	{ <b>35,36</b> ,58}, {53,61}	721
tRNA <sup>Ile</sup>	{8,10,14,18,22,25, <b>35,36</b> ,53,54,56,58,61}, {7,66}, {11,24}, {12,23}, {28,42}, {29,41}, {50,64}	32
tRNA <sup>Leu</sup>	{8,14,18, <b>35</b> ,53,54,56,58,61}, {52,62}	731
tRNA <sup>Met</sup>	{8,33, <b>34,35,36</b> ,53,54,56,58,61}	559
tRNA <sup>Phe</sup>	{8,10, <b>12,21,23</b> ,33, <b>36</b> ,53,56}, {34,35}	812
tRNA <sup>Trp</sup>	{18,33, <b>35,36</b> ,37,53,55,58,61}	552
tRNA <sup>Val</sup>	{18, <b>35,36</b> }, {12,23}	774

tRNA	Class II	
	Unity Clusters	N
tRNA <sup>Ala</sup>	{18, <b>35,36</b> ,54,58}, {53,61}	1563
tRNA <sup>Asn</sup>	{ <b>35,36</b> ,53,58}, {12,23}, {18,56}, {52,62}, {31,39}	579
tRNA <sup>Asp</sup>	{1,17,33, <b>34,35,36</b> ,53,54,58,61,72}, {11,24}, {50,64}, {51,63}	126
tRNA <sup>Gly</sup>	{33, <b>35,36</b> ,54,58}, {50,64}, {53,61}	804
tRNA <sup>His</sup>	{18, <b>34,35,36</b> ,53,54,55,56,61}, {12,23}, {29,41}, {51,63}	584
tRNA <sup>Lys</sup>	{18,33, <b>35,36</b> }, {37,38}, {53,61}	1490
tRNA <sup>Pro</sup>	{ <b>35,36</b> ,53,54,55,58,61,73}, {1,72}, {2,71}, {3,70}, {18,33}, {29,41}	179
tRNA <sup>Ser</sup>	{1,33,53,54,55,56,60,61}, { <b>35,36</b> }, {5,68}, {29,41}, {52,62}	55
tRNA <sup>Thr</sup>	{ <b>35,36</b> ,37}, {3,70}, {53,61}	1779
tRNA <sup>Tyr</sup>	{18,21,32,33, <b>34,35,36</b> ,53,54,56,58,60,61,73}, {30,40,55}, {2,71}, {3,70}, {4,69}, {5,68}, {8,14}, {10,25}, {12,23}, {28,42}, {31,39}, {49,65}, {52,62}	19

base in the D-loop; the base 8 that links the acceptor stem and the D-loop. The second cluster is a Watson-Crick pair at the TΨC-loop. A third cluster displays bases at opposite sides of the tRNA, the base 18 at the D-loop, and the nucleotide 55 in the TΨC-loop. Altogether, these 3 clusters represent the identity elements of the operational code for tRNA<sup>Gln</sup>. The isoacceptors that correspond to Class I more generally present the nucleotide 8 from the D-loop, and in some cases the base 18 is also present. Class I often present the nucleotide 8 from the D-loop clustered with the anticodon, along with base 14. In both classes, the sites 53, 54, 58, and 61 from the TΨC-loop, are all generally associated with the anticodon when some base from that side is present, marking a diffuse pattern. The tRNA with the largest number of clusters is tRNA<sup>Tyr</sup> but this observation must be taken with caution due to the small sample size. We strongly recommend increasing the sample size for tRNA<sup>Tyr</sup>. The wobbling base 34 is present only in Cys, Met, Phe, Asp, His, and Tyr. There is no relation between the codonicity and the location of the identity elements. The remaining figures for all isoacceptors can be found in Appendix A.

**Fig. 1** Identity elements of tRNA<sup>Gln</sup>. The figure portrays the tRNA secondary structure of tRNA<sup>Gln</sup>. Clusters are marked with different colors



## Discussion

The present work expands the current catalogue of identity elements of the 20 canonical tRNA isoacceptor groups. Every isoacceptor possess a set of sites (clusters) that includes at least one of the anticodon bases. This is in agreement with the association of the anticodon as an identity element for all tRNAs (Giegé et al. 1998). The sites related to the anticodon are present along all the structure and are different for each tRNA group. Thus, our results are in agreement with the idiosyncratic hypothesis of identity elements (Loftfield 1972) and with its distribution on the molecule (Goddard 1977). The information theoretical approach for detecting identity elements was initiated by Durbin et al. (2002) and followed by Adami (2004, 2012). The identity elements found in yeast tRNA<sup>Phe</sup> by Durbin et al. (2002), using the mutual information function, are practically the same as the ones reported in the present work.

The hypothesis of the anticodon as common regulator was later rejected with in-vitro experiments on tRNA<sup>Ser</sup> that showed no evidence of recognition between the anticodon and the corresponding aaRS (Sundaharadas et al. 1968). Data from yeast and *E. coli* provided clues about the presence of identity elements in the acceptor stem, the position 73 (which is the last before the CCA), the anticodon, the variable loop and the D stem (Goddard 1977). Such results showed no specific sites that could answer the current hypothesis of universal recognition sites, hence, the sites should be idiosyncratic for each isoacceptor (Loftfield 1972). Later, experiments revealed the discriminator base pair G3:U70, which was involved in the correct recognition of tRNA<sup>Ala</sup> (Vargas-Rodriguez and Musier-Forsyth 2014). Further experimental work on particular species, consisted in single modifications of nucleotides along the whole tRNA in order to detect concrete positions that decrease the aminoacylation reaction, both in-

vivo and in-vitro (Giegé et al. 1998). These experiments revealed that each tRNA isoacceptor holds specific sites involved in their correct recognition. The discriminator site 73, was recognized as an identity element in conjunction with the anticodon bases. This base is present at the anticodon cluster for the tRNA<sup>Pro</sup> and tRNA<sup>Tyr</sup> isoacceptors. This base has been reported for Tyr (Bonfond et al. 2005). The long variable loop of tRNA<sup>Ser</sup> has been reported to be determinant for its correct recognition (Wu and Gross, 1993). It has been reported the existence of positions for positive or negative recognition that participate in the correct aminoacylation or that prevent false recognition and mischarging, respectively (Giegé et al. 1998). There are also sites with different forces of recognition, being strong or weak sites. It has also been reported that in a single organism, tRNAs from different isoaccepting groups are more similar to each other than to their isoaccepting counterparts (Saks et al. 1994). They argued that this could be due to an accumulation of neutral mutations that are blind to tRNA recognition. It has also been shown that anticodon mutations could lead to changes in the isoacceptor group and they are highly tolerated. Bioinformatic analysis has shown the discriminator base pair 1:72 for tRNA<sup>Tyr</sup> and tRNA<sup>Trp</sup> (Mukai et al. 2017). Some clusters that contain the anticodon are accompanied by the bases 33 and 37. This has been proposed as an extended anticodon since the bases surrounding the anticodon are recognized by the corresponding aaRSs (Yarus 1982). The entropy per site of each tRNA has been calculated and it has been shown that, in average, there is no difference between the entropy profiles between the major groove and the minor groove (José et al. 2016). This result is in agreement with the sense-antisense complementary hypothesis of a common origin of tRNAs (Rodin and Rodin 2008; Carter et al. 2014). The division of genetic code, by the aaRSs class, has been shown to be an important factor in its evolutionary process (Rodin and Ohno 1977, Carter et al. 2014; Zamudio and José 2017; José et al. 2017),

Our informational approach provides a new insight for determining the identity elements of the operational code. Our results suggest further experimental work for testing the proposed sets of identity elements.

It is widely accepted the early relevance of the acceptor mini-helix in the evolutionary development of tRNA molecules (Schimmel et al. 1993; Schimmel 1995; Rodin et al. 1996). It has been proposed that the amino acid-accepting stem emerged before the anticodon loop of tRNAs, so that the first codification obeyed an operational code where amino acids were attached to their respective tRNA without the need of anticodon loop recognition (Park and Schimmel 1988; Hou and Schimmel 1988; Schimmel et al. 1993; Ribas de Pouplana et al. 1998). Indeed, the origin of the operational code is directly related to the absence of the anticodon loop in tRNAs, which enabled the first peptides to be synthesized in the absence of a genetic code (Belousoff et al. 2010). It has also been suggested that the primitive ribosome worked to synthesize peptides randomly, without the need of a code (Belousoff et al. 2010). However, Shimizu (1995) showed that small tRNAs with the portion of the anticodon loop could bind the amino acid as well. Hence, the anticodon loop was already present in primitive tRNA and should have been important in establishing specific interactions between tRNAs and their corresponding aminoacyl-tRNA synthetases.

Farias et al. (2014a, b) suggested that the coding system was assembled by co-evolution between tRNAs and aminoacyl-tRNA synthetases, being driven by changes in the second base of the anticodon of tRNA, which in turn changed the hydrophathy of the anticodon, and this pressure guided the diversification of aminoacyl-tRNA synthetases. Therefore, the recognition of the anticodon was shown to be essential for the development of the encoding system and acted as a selective pressure for the diversification of aaRSs. The two domains of the L-shaped

tRNA would have arisen independently, with the acceptor branch appearing first. In a later stage in history, the catalytic cores of synthetases emerged independently in their class I and II versions. Co-evolution of catalytic cores of synthetases and accepting hairpins led to an operational RNA code that associated specific amino acids with hairpin structures (Park and Schimmel 1988; Schimmel 1995). The anticodon domain of tRNA and the additional domains of synthetases appeared later in evolution. Anticodon domains brought the link between the RNA operational code and the correlated tRNA recognition by synthetases with the anticodon-dependent recognition by mRNA.

It has been shown (Farias et al. 2017) that the initial existence of an operational code was due to the agglutination capacity of tRNAs without the presence of a genetic code in the Peptidyl Transferase Center (PTC). This suggests that the anticodon loop initially increased the specificity between tRNAs and amino acids, and after the emergence of the proto-genes, by an exaptation process, the anticodon loops were co-opted to interact with the proto-genes, and thus the genetic code emerged and started to decode the biological information. In this manner, the emergence of the operational code and the genetic code occurred simultaneously and both systems played complementary roles in the origin and evolution of the translation system (Farias et al. 2017).

The peptides synthesis without the need of a code, indicates that in the early stages of the process the anticodon had other functions: for example, they could have been involved in the establishment of the assignments between amino acids and tRNAs. Thus, the emergence of the first genes or proto-genes must have reorganized the modes of interaction between tRNAs and PTC, which defined the sites of interaction A, P, and E in the modern ribosome. Hence, the anticodons were co-opted to stabilize the binding with the proto-gene and from this secondary interaction, the genetic correspondences between codons and amino acids were gradually established (Farias et al. 2017). The emergence of the genetic code must have occurred as an exaptation process, where anticodons would initially have the function of increasing the specificity between amino acids and tRNAs. The appearance of proto-genes would have been co-opted to establish a correlation between the information contained in the nucleic acids to proteins. The origin of the translation system is a major evolutionary transition because it enabled the establishment of the ribonucleoprotein world. The tRNAs molecules played a central role during the origin of translation, bridging RNA and (RNA + proteins) worlds. These tRNAs may offer clues towards the elucidation of the origin of the genetic code. Farias et al. (2014a, b) reconstructed the ancestral sequences of tRNAs, and when they compared concatamers of these tRNAs with the sequence of PTC from *Thermus thermophilus*, a similarity of 50.52% was found. Therefore, they suggested that the PTC arose from the junction of proto-tRNAs. It has been shown that the three-dimensional structure of the ancestral PTC, built by concatamers of ancestral sequences of tRNAs, had interactions with tRNAs anticodon loop (Farias et al. 2017).

The tRNA core hypothesis accounts for the transition from the RNA world to the Ribonucleoprotein world, where proto-tRNAs molecules possessing similar folds to those observed in modern tRNAs guided the evolutionary process of the genetic code and the translation, and enabled its fixation (Farias et al. 2014a, b). The tRNA core hypothesis (Farias et al. 2014a, b) places the tRNA at the origin of translation, i.e., tRNA molecules played a significant role in the organization of the first codified biological system, established the information storage system, and participated in coding and decoding this information. They were the protagonists of the translation system via chemical interactions with amino acids, which are inherent to the transition between the RNA world to a ribonucleoprotein world. In the tRNA core hypothesis (Farias et al. 2016), the first genes were derived from tRNA by

structural changes (tRNA-like-mRNA structure), and they enabled other tRNAs (cloverleaf tRNA canonical structure) to bind this sequence to the loop of the proto-anticodon. The tRNA molecules with the proto-tRNA (anticodon stem/loop) could interact with the amino acids (as cofactor or riboswitches), which were present in the prebiotic environment; hence, the binding between tRNA and amino acids was established. The tRNAs binding to amino acids could interact with other tRNAs in open conformation (tRNA-like mRNA structure). This interaction stabilized the complex cloverleaf tRNA canonical structure-tRNA-like mRNA structure.

**Acknowledgements** Gabriel S. Zamudio is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and a fellowship recipient from Consejo Nacional de Ciencia y Tecnología (CONACYT) (number: 737920). Marco V. José was financially supported by PAPIIT-IN224015, UNAM, México.

## References

- Abe T, Inokuchi H, Yamada Y et al (2014) tRNADB-CE: tRNA gene database well-timed in the era of big sequence data. *Front Genet* 5:114
- Adami C (2004) Information theory in molecular biology. *Phys Life Rev* 1:3–22. doi:10.1016/j.plev.2004.01.002
- Adami C (2012) The use of information theory in evolutionary biology. *Ann N Y Acad Sci* 1256:49–65. doi:10.1111/j.1749-6632.2011.06422.x
- Altman RB (1993) Probabilistic structure calculations: a three-dimensional tRNA structure from sequence correlation data. *Proc Int Conf Intell Syst Mol Biol* 1:12–20
- Arnez JG, Moras D (1997) Structural and functional considerations of the aminoacylation reaction. *Trends Biochem Sci* 22:211–216
- Belousoff MJ, Davidovich C, Bashan A, Yonath A (2010) On the development towards the modern world: a plausible role of uncoded peptides in the RNA world. In: Ruiz-Mirazo K, Luisi PL (eds) *Origins of life and evolution of biospheres*. Springer, New York City, pp 415–419
- Bonnefond L, Giegé R, Rudinger-Thirion J (2005) Evolution of the tRNA<sup>Tyr</sup>/TyrRS aminoacylation systems. *Biochimie*. 87:873–883
- Carter CW, Li L, Weinreb V et al (2014) The Rodin-Ohno hypothesis that two enzyme superfamilies descended from one ancestral gene: an unlikely scenario for the origins of translation that will not be dismissed. *Biol Direct* 9:11. doi:10.1186/1745-6150-9-11
- Cusack S (1997) Aminoacyl-tRNA synthetases. *Curr Opin Struct Biol* 7:881–889. doi:10.1016/S0959-440X(97)80161-3
- De Duve C (1988) The second genetic code. *Nature* 333:117–118. doi:10.1038/333117a0
- Durbin R, Eddy SR, Krogh A, Mitchison G (1998) *Biological sequence analysis*. Cambridge University Press, Cambridge
- Durbin R, Eddy S, Krogh A, Mitchison G (2002) *Biological sequence analysis*. Cambridge University Press, Probabilistic models of proteins and nucleic acids
- Eriani G, Delarue M, Poch O et al (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Lett to Nat* 347:203–206. doi:10.1038/346183a0
- Farias ST, Rêgo TG, José MV (2014a) Evolution of transfer RNA and the origin of the translation system. *Front Genet* 5:303. doi:10.3389/fgene.2014.00303
- Farias ST, Rêgo TG, José MV (2014b) Origin and evolution of the peptidyl transferase center from proto-tRNAs. *FEBS Open Bio* 4:175–178
- Farias ST, Rêgo TG, José MV (2016) tRNA core hypothesis for the transition from the RNA world to the ribonucleoprotein world. *Life (Basel, Switzerland)* 6:15. doi:10.3390/life6020015
- Farias ST, Rêgo TG, José MV (2017) Peptidyl transferase center and the emergence of the translation system. *Life (Basel, Switzerland)* 7:21. doi:10.3390/life7020021
- Giegé R, Sissler M, Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* 26:5017–5035
- Goddard JP (1977) The structures and functions of transfer RNA. *Prog Biophys Mol Biol* 32:233–308. doi:10.1016/0079-6107(78)90021-4



- Guimarães RC, Moreira CHC, de Farias ST (2008) A self-referential model for the formation of the genetic code. *Theory Biosci* 127:249–270. doi:[10.1007/s12064-008-0043-y](https://doi.org/10.1007/s12064-008-0043-y)
- Hou YM (2010) CCA addition to tRNA: implications for tRNA quality control. *IUBMB Life* 62:251–260
- Hou Y-M, Schimmel P (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* 333:140–145. doi:[10.1038/333140a0](https://doi.org/10.1038/333140a0)
- José MV, Morgado ER, Guimarães RC et al (2014) Three-dimensional algebraic models of the tRNA code and 12 graphs for representing the amino acids. *Life (Basel, Switzerland)* 4:341–373. doi:[10.3390/life4030341](https://doi.org/10.3390/life4030341)
- José MV, Zamudio GS, Farias ST (2016) Evolution of tRNAs was driven by entropic forces. Evolution of the protein synthesis machinery and its regulation. In: Jagus R, Hernández G (eds). Springer International Publishing Switzerland, pp 1–7. Book ID: 370644\_1\_En. doi: [10.1007/978-3-319-39468-8](https://doi.org/10.1007/978-3-319-39468-8)
- José MV, Zamudio GS, Morgado ER (2017) A unified model of the standard genetic code. *Royal Society Open Science* 4:160908. doi:[10.1098/rsos160908](https://doi.org/10.1098/rsos160908)
- Li W (1990) Mutual information versus correlation functions. *J Stat Phys* 60:823–837. doi:[10.1007/BF01025996](https://doi.org/10.1007/BF01025996)
- Lofffield RB (1972) The mechanism of Aminoacylation of transfer RNA. *Prog Nucleic Acid Res Mol Biol* 12: 87–128. doi:[10.1016/S0079-6603\(08\)60660-1](https://doi.org/10.1016/S0079-6603(08)60660-1)
- Meila M (2003) Comparing clusterings by the variation of information. *Learn theory Kernel Mach 16th Annu Conf Learn Theory 7th Kernel Work COLT/Kernel 2003*, Washington, DC, USA, August 24–27, 2003 Proc 173. doi: [10.1007/978-3-540-45167-9\\_14](https://doi.org/10.1007/978-3-540-45167-9_14)
- Meilã M (2007) Comparing clusterings-an information based distance. *J Multivar Anal* 98:873–895. doi:[10.1016/j.jmva.2006.11.013](https://doi.org/10.1016/j.jmva.2006.11.013)
- Mukai T, Reynolds N, Crnković A, Söll D (2017) Bioinformatic analysis reveals archaeal tRNA<sup>Tyr</sup> and tRNA<sup>Trp</sup> identities in bacteria. *Life (Basel, Switzerland)* 7:8. doi:[10.3390/life7010008](https://doi.org/10.3390/life7010008)
- Park SJ, Schimmel P (1988) Evidence for interaction of an aminoacyl transfer RNA synthetase with a region important for the identity of its cognate transfer RNA. *J Biol Chem* 15:16527–16530
- Ribas de Pouplana L, Turner RJ, Steer BA, Schimmel P (1998) Genetic code origins: tRNAs older than their synthetases? *Proc Natl Acad Sci U S A* 15:11295–11300
- Rodin SN, Ohno S (1997) Four primordial modes of tRNA-synthetase recognition, determined by the (G,C) operational code. *Proc Natl Acad Sci U S A* 94:5183–5188. doi:[10.1073/pnas.94.10.5183](https://doi.org/10.1073/pnas.94.10.5183)
- Rodin SN, Rodin AS (2008) On the origin of the genetic code: signatures of its primordial complementarity in tRNAs and aminoacyl-tRNA synthetases. *Heredity (Edinb)* 100:341–355. doi:[10.1038/sj.hdy.6801086](https://doi.org/10.1038/sj.hdy.6801086)
- Rodin SN, Rodin AS, Ohno S (1996) The presence of codon-anticodon pairs in the acceptor stem of tRNAs. *Proc Natl Acad Sci U S A* 93:4537–4542. doi:[10.1073/pnas.93.10.4537](https://doi.org/10.1073/pnas.93.10.4537)
- Saks ME, Samposn JR, Abelson JN (1994) The transfer RNA identity problem: A search for rules. *Science* 263: 191–197
- Schimmel P (1995) An operational RNA code for amino acids and variations in critical nucleotide sequences in evolution. *J Mol Evol* 40:531–536. doi:[10.1007/BF00166621](https://doi.org/10.1007/BF00166621)
- Schimmel P, Giégé R, Moras D, Yokoyama S (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc Natl Acad Sci U S A* 90:8763–8768. doi:[10.1073/pnas.90.19.8763](https://doi.org/10.1073/pnas.90.19.8763)
- Shimizu M (1995) Specific aminoacylation of C4N hairpin RNAs with the cognate aminoacyl-adenylates in the presence of a dipeptide: origin of the genetic code. *J Biochem* 117:23–26
- Sundaharadas G, Katze JR, Söll D et al (1968) On the recognition of serine transfer RNA's specific for unrelated codons by the same seryl-tRNA synthetase. *Proc Natl Acad Sci U S A* 61:693–700
- Tamura K (2015) Origins and early evolution of the tRNA molecule. *Life* 5:1687–1699. doi:[10.3390/life5041687](https://doi.org/10.3390/life5041687)
- Vargas-Rodríguez O, Musier-Forsyth K (2014) Structural biology: wobble puts RNA on target. *Nature* 510:480–481
- Wu XQ, Gross HJ (1993) The long extra arms of human tRNA(Ser)<sup>sec</sup> and tRNA<sup>Ser</sup> function as major identity elements for serylation in an orientation-dependent, but not sequence-specific manner. *Nucleic Acids Res* 21:5589–5594. doi:[10.1093/nar/21.24.5589](https://doi.org/10.1093/nar/21.24.5589)
- Yarus M (1982) Translational efficiency of transfer RNA's: uses of an extended anticodon. *Science* 218:646–52. doi:[10.1126/science.6753149](https://doi.org/10.1126/science.6753149)
- Zamudio GS, José MV (2017) On the uniqueness of the standard genetic code. *Life* 7:7. doi:[10.3390/life7010007](https://doi.org/10.3390/life7010007)