# The novel Ideal Symmetry Genetic Code table – Common purine-pyrimidine symmetry net for all RNA and DNA species

**Rosandić, Marija; Paar, Vladimir**

Journal of Theoretical Biology 524 (2021) 110748

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

# The novel Ideal Symmetry Genetic Code table – Common purine-pyrimidine symmetry net for all RNA and DNA species

Marija Rosandić [a,b,*], Vladimir Paar [b,c]

[a] University Hospital Centre Zagreb (ret.), Zagreb, Croatia
[b] Croatian Academy of Sciences and Arts, 10000 Zagreb, Croatia
[c] Department of Physics, Faculty of Science, University of Zagreb, Croatia

ARTICLE INFO

*Article history:*
Received 30 November 2020
Revised 22 April 2021
Accepted 27 April 2021
Available online 29 April 2021

*Keywords:*
DNA symmetry
Genetic code symmetry
Mitochondrial code
Nuclear code
Standard genetic code

ABSTRACT

Ever since Nirenberg's discovery in 1961 in which codons code individual amino acids, numerous scientists searched for symmetries within the genetic code. The standard genetic code (SGC) table is an alphabetic artificial construct based on the U-C-A-G ordering of nucleotides without natural symmetries. Up to the present, complete symmetry in the genetic code has not been found, leaving doubt as to whether the symmetrical nature as the protector of order even exists. Our novel Ideal Symmetry Genetic Code (ISyGC) table reflects a unique fundamental physicochemical purine-pyrimidine symmetry net for all more than thirty known variations of nuclear and mitochondrial genetic codes. The nuclear genetic code for RNA and DNA viruses also contains the same purine-pyrimidine symmetry net. We show that the ISyGC table leads to automatic transformation into a DNA sequence akin to the 5′3 codon and 3′5 anticodon patterns. As a result of purine-pyrimidine symmetries between codons in the ISyGC table, algorithms of the first two bases as well of the third base of codons show how tRNA cognate anticodons can recognize synonymous codons during mRNA decoding. We show that the ISyGC purine-pyrimidine net with its physicochemical properties represents an evolutionary common "frozen accident" at the onset of each genetic code creation and RNA to DNA evolution. As such, during all of evolution the unique fundamental purine-pyrimidine symmetry net of all genetic codes remains unchangeable. In this way, evolution is a road paved with symmetries.


## 1. Introduction

It has been a 60-year challenge for biology to find optimal symmetry of the genetic code after initially deciphering the entire genetic code identifying amino acid or stop signals corresponding to each one of the 64 nucleotide codons (Gamow, 1954; Nirenberg and Matthaei, 1961). The fundamental role of symmetry in the genetic code is to decrease disorder between codons and to preserve the integrity of systems (Rosandić et al. 2013, Rosandić et al. 2019, Rosandić and Paar 2014). Namely, information transfer from one nucleic acid to another is governed by stereochemistry, in that it relies exclusively on the symmetry between codons.

Symmetries within the genetic code were on the whole researched giving emphasis to codons and amino acids distribution in triangular, rectangular, circular or torus form, but also to the distribution of nucleotides within codons or their binary transformation (Ahmed et al., 2010; Berleant et al., 2009; Grosjean and Westhof, 2016a,b; Jose et al., 2017; Koonin and Novozhilov, 2009; Kubyshkin et al., 2018; Lenstra, 2014; Michel, 2013; Michel and Pirillo, 2010; Nemzer, 2017; Seligmann and Warthi, 2017; Shu, 2017; Štambuk and Konjevoda, 2020; Castro-Chavez, 2012; Štambuk, 2000). There are several references that link the structure of the genetic code to the structure of codons regarding the distribution of types of nitrogenous bases throughout the codon (Seligmann and Warthi, 2017). Unfortunately, they all suffer from an inability to illustrate a symmetrical nature of the genetic code. Their results conveyed little about the functional physicochemical relationships and symmetries between codons in the code and its evolution.

The algebraic approach to the Standard Genetic Code (SGC) was proposed with the aim of explaining the degeneracies encountered in SGC as a result of a sequence of symmetry breakings (Hornos and Hornos, 1993; Stewart, 1994; Kent et al., 1998; Hornos et al., 2004; Antoneli et al., 2010). The algebraic sp(6) - model was favored, although there are a few other possibilities (Hornos et al., 1999; Antoneli and Forger, 2011). In this way, the genetic code degeneracies were obtained, but these results also do not

* Corresponding author at: Croatian Academy of Sciences and Arts, 10000 Zagreb, Croatia.
*E-mail addresses:* rosandic@hazu.hr (M. Rosandić), paar@hazu.hr (V. Paar).

https://doi.org/10.1016/j.jtbi.2021.110748
0022-5193/

explain the functional physicochemical relationships and symmetries between codons in the genetic code.

There are more than $10^{84}$ possible alternative code tables, but only one in every million random alternative code is more efficient than the SGC table (Freeland and Hurst, 1998; Koonin, 2017). Despite this, all biology and genetics textbooks which present the SGC table suffer from an inability to show the complete physicochemical symmetry between codons.

How large the scientific challenge of finding a relationship between symmetry and the evolution of the genetic code is testified in the comment by Eugene W. Koonin (2009): It seems that the two pronged fundamental question "why is the genetic code the way it is and how did it come to be?", that was asked over fifty years ago, at the dawn of molecular biology, might remain pertinent even in another fifty years.

A better understanding of symmetry and the essential role it has played in codon formation will improve our understanding of the nature of coding activity and evolution. We demonstrate that the discovery of a fundamental purine – pyrimidine symmetry net in our ISyGC table, common for all the various genetic codes, can facilitate that improvement.

## 2. Material and methods

### 2.1. Differences between the SGC and ISyGC tables

The full set of relationships between codons and amino acids or stop signals is called the genetic code. It is usually summarized in a table called the genetic code table. Different types of genetic code tables can be assigned to each genetic code.

The random alternative codes consist of 64 codons randomly assigned to 20 natural amino acids with no physicochemical or symmetry relationship between codons and amino acids.

The genetic code is degenerate because more than one type of codon (2, 3, 4, or 6) may encode a single amino acid.

The SGC table is the structure of 61 codons and 3 stop signals. It seems to have universal tendencies among organisms such as prokaryotes and eukaryotes and provides a link between inheritance and development and appears optimized for resistance to various types of error and fixation of an optimal genetic code (Freeland et al., 2000). Even although the bases are grouped as pyrimidines (U, C) and purines (A, G), the SGC table is structured alphabetically in a horizontal and vertical array of U-C-A-G bases with partial symmetry and no physicochemical relationship between them (Fig. 2B)).

The main difference between the SGC table and our ISyGC table is the purine – pyrimidine symmetry net with its physicochemical properties which constraints underlying amino acids assignment in the ISyGC table. The ISyGC table consists of leading and nonleading groups of codons, where all four columns, each of the 16 codons, have an identical relationship between purines and pyrimidines (Rosandić and Paar 2014, Rosandić et al. 2019). At the same time, pairs of codon rows along the whole ISyGC table also have the same purine-pyrimidine pattern. (Figs. 1 and 2).

In the leading group, the serin plays a central role with its six codons. Namely, in the ISyGC table, only serin encompasses two neighboring vertical boxes, whose codons are in a direct ↔ complement relationship. Transforming horizontally purine ↔ purine and pyrimidine ↔ pyrimidine, serin positions the whole box from four of six codons of leucine. The remaining two codons displace to a non-leading group and in this way codons with their third base lead to a positioning of the corresponding codons in a non-leading group. The sextet assigned to arginine also encompasses two boxes in the leading group. However, they are not in a direct ↔ complement relationship but continue vertically to serin.

Each of the four columns with 16 codons has the same arrangement of purines and pyrimidines: 010, 010, 011, 011 – 101,101, 100, 100; 000, 000, 001, 001 – 111, 111, 110, 110. Pairs of codons in the same box also have the same relationship between purines and pyrimidines. Within each column, direct (italic) – complement (bold) boxes alternate (Fig. 1). This symmetry base construction of the ISyGC table is automatically placing the AUG start signal at the beginning of the code. Due to purine–pyrimidine symmetries, the localization of each codon in the ISyGC table is strictly defined: codons assigned for amino acids are arranged in continuo and are not scattered as in the SGC table. The grouping of similar amino acids within the same column of the code table immediately indicates that the code has a degree of robustness to mutational and translational errors (Koonin, 2017). We point out that our ISyGC table is highly compatible with this statement.

The stereochemical theory postulates that the structure of the code is determined by a physicochemical affinity between amino acids and codons or anticodons (Koonin, 2017). In the ISyGC table, amino acids are also characterized by polarity, acid base property and an aromatic ring, approximately equally distributed between leading and nonleading groups of amino acids (Rosandić et al. 2019). The ISyGC table in DNA form (Fig. 3(a)) automatically represents a sequence like the 5′3 codon and 3′5 anticodon pattern.

As a result of all these characteristics, the Ideal Symmetry Genetic Code table is named as such for good reason.

The ISyGC table is structured arraying 16 characteristic boxes, each containing 4 codons. All codons within the same box have an equal pair of the first and second base. There are four different groups of pairs of boxes, which are mutually in a direct-complement relationship (vertically), and in purine–purine or pyrimidine–pyrimidine (horizontally) strictly symmetrically distributed bases within the ISyGC table. Such an arrangement of codons between boxes in the ISyGC table creates a new symmetry resembling a spiral-like shape (Fig. 1) and universal common physicochemical purine–pyrimidine symmetry net (Fig. 2a and b).

Their structure enables an automatic transformation of the ISyGC table into a DNA-type sequence with Watson-Crick pairing as well as its form analogous to the 5′3 codon and the 3′5 anticodon (Fig. 3a). This transformation appears automatically by linearizing the first four rows of direct codons from the ISyGC table one after the other in a single line. Analogous linearization is performed for complement codons from the second rows of boxes (Fig. 3(a)A). In this way, the molecular structure of DNA is automatically created with direct and complement codons of ISyGC table (Fig. 1). The same pattern appears for codons from the third and fourth rows of boxes from ISyGC table (Fig. 3(a)B). To our knowledge such symbiosis between genetic code and DNA molecule was not known.

We have shown (Rosandić et al., 2016, 2019) that the basic building structure, even of the smallest DNA genomes such as, for example, symbionts *Carsonela hodgkinia cicadicola* (143795 bp) and *Candidatus carsonela ruddi* (162589 bp), consist of 10 A + T rich and 10 C + G rich quadruplets with strong purine-pyrimidine symmetry from our trinucleotide classification.

The same 10 A + U rich and 10 C + G rich quadruplets of the codon's classification with substitution T in U (Fig. 4) are incorporated in the ISyGC table. (Fig. 3(b). For each trinucleotide, the corresponding quadruplet consists of its direct - reverse complement – complement – reverse, structured on fundamental purine–pyrimidine symmetry. It could be said that the ISyGC table with a purine-pyrimidine symmetry net (Fig. 2A) represents the smallest common denominator of 20 amino acids and their corresponding 61 codons and 3 stop signals in each DNA genome. In this way, the ISyGC table and the DNA genome connect the same fundamental purine – pyrimidine symmetry. From an evolutionary point of view, with the appearance of the first DNA species, their genetic code table

| | Leading group | | | | | | Nonleading group | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Box | Amino acid | I. Codons | Pu/Py | Pu/Py | II. Codons | Amino acid | Amino acid | III. Codons | Pu/Py | Pu/Py | IV. Codons | Amino acid |
| Direct boxes 1-4 | Met/Start | *AUG* | 010 | 010 | GCA | Ala | Thr | *ACA* | 010 | 010 | GUG | Val |
| | Ile | *AUA* | 010 | 010 | GCG | | | ACG | 010 | 010 | *GUA* | |
| | | *AUC* | 011 | 011 | GCU | | | *ACU* | 011 | 011 | GUC | |
| | | *AUU* | 011 | 011 | GCC | | | ACC | 011 | 011 | *GUU* | |
| Comple-ment boxes 5-8 | Tyr | *UAC* | 101 | 101 | CGU | Arg | Cys | *UGU* | 101 | 101 | CAC | His |
| | | *UAU* | 101 | 101 | CGC | | | UGC | 101 | 101 | *CAU* | |
| | Stop | *UAG* | 100 | 100 | CGA | | Stop | *UGA* | 100 | 100 | CAG | Gln |
| | Stop | *UAA* | 100 | 100 | CGG | | Trp | *UGG* | 100 | 100 | *CAA* | |
| Direct boxes 9-12 | Glu | GAG | 000 | 000 | *AGA* | Arg | Gly | GGA | 000 | 000 | *AAG* | Lys |
| | | *GAA* | 000 | 000 | AGG | | | GGG | 000 | 000 | *AAA* | |
| | Asp | GAC | 001 | 001 | *AGU* | Ser | | GGU | 001 | 001 | *AAC* | Asn |
| | | *GAU* | 001 | 001 | *AGC* | | | GGC | 001 | 001 | *AAU* | |
| Comple-ment boxes 13-16 | Leu | CUC | 111 | 111 | *UCU* | Ser | Pro | CCU | 111 | 111 | *UUC* | Phe |
| | | *CUU* | 111 | 111 | UCC | | | CCC | 111 | 111 | *UUU* | |
| | | CUG | 110 | 110 | *UCA* | | | CCA | 110 | 110 | *UUG* | Leu |
| | | *CUA* | 110 | 110 | UCG | | | CCG | 110 | 110 | *UUA* | |

**Fig. 1.** The Ideal symmetry genetic code (ISyGC) table; 0 pu: purine, 1 py: pyrimidine; italic: A + U rich codons; bold: C + G rich codons; dark blue: no split CG codon boxes (vertical pair of complementary boxes); light blue: no split mixed codon boxes (horizontal pair of boxes with purine-purine and pyrimidine-pyrimidine transformation); dark yellow: split AU codon boxes (vertical pair of complementary boxes); light yellow: split mixed codon boxes (horizontal pair of boxes with purine-purine and pyrimidine - pyrimidine transformation). Within the ISyGC table, vertical and horizontal pairs of boxes are also symmetrically arranged. In this way, the third dimension of symmetry position of codons in a spiral form with a clockwise direction is created. The leading group of codons: columns I and II; non-leading group of codons: columns III and IV. In our code table, three symmetries are present. First, the purine-pyrimidine structure of all four codon columns within the same row is identical and the consecutive pairs of codon rows also have an identical purine-pyrimidine profile (see also Fig. 2). Second, boxes 5–8 and 13–16 are complements of direct boxes 1–4 and 9–12, respectively, and *vice versa*. Third, A + U rich (italics) and C + G rich (bold) codons alternate between pairs of codon columns. This genetic code table is modified with respect to refs. (Rosandić and Paar 2014, Rosandić et al. 2019) by exchanging the position of the third and fourth codon columns. The purine-pyrimidine symmetry net in the ISyGC table does not change because the purine and pyrimidine arrangement of all four column codons is identical.

is simultaneously structured with its fundamental purine – pyrimidine symmetry net. We can say that the unique ISyGC table net supports the "frozen accident" hypothesis (Crick 1968) of the genetic code evolution. Namely, Nobel laureate Francis H. Crick considered the SGC table an optimal solution for the organization of codons and amino acids and search for symmetries. He found only partial direct – complement symmetry for the first two bases of codons in boxes whose four codons code the same amino acid. The third base can be any of the four possible and have no role in symmetry creation. He concluded that various factors could have contributed to the initial codon assignments, but once the choice is made, it gets frozen and only minor and rare change may be allowed.

In the initial construction of the ISyGC table (Rosandić and Paar 2014), it seemed that our discovery denied the "frozen accident" theory as we found ideal purine-pyrimidine symmetries between all codons, which is not present in the SGC table. However, from the present investigations between our ISyGC table and more than 30 known other genetic codes, we discovered the fundamental common unique purine-pyrimidine symmetry net (Fig. 2A), consisting of physicochemical properties with symmetries between bases and codons as well. In fact, the ISyGC table with the stable and unchangeable purine-pyrimidine net is invariant of biological evolution and as such can be considered a "frozen accident".

*a)*

**A) ISyGC table**

codon boxes

| | 1st | 2nd | 3th | 4th |
|---|---|---|---|---|
| 1-4 box direct | 010 | 010 | 010 | 010 |
| | 010 | 010 | 010 | 010 |
| | 011 | 011 | 011 | 011 |
| | 011 | 011 | 011 | 011 |
| 5-8 box compl. | 101 | 101 | 101 | 101 |
| | 101 | 101 | 101 | 101 |
| | 100 | 100 | 100 | 100 |
| | 100 | 100 | 100 | 100 |
| 9-12 box direct | 000 | 000 | 000 | 000 |
| | 000 | 000 | 000 | 000 |
| | 001 | 001 | 001 | 001 |
| | 001 | 001 | 001 | 001 |
| 13-16 box compl. | 111 | 111 | 111 | 111 |
| | 111 | 111 | 111 | 111 |
| | 110 | 110 | 110 | 110 |
| | 110 | 110 | 110 | 110 |

**B) SGC table**

codon boxes

| | 1st | 2nd | 3th | 4th | |
|---|---|---|---|---|---|
| 1-4 box no direct | 111 | 111 | 101 | 101 | |
| | 111 | 111 | 101 | 101 | 1-4 box |
| | 110 | 110 | 100 | 100 | no direct |
| | 110 | 110 | 100 | 100 | |
| 5-8 box no compl. | 111 | 111 | 101 | 101 | |
| | 111 | 111 | 101 | 101 | 5-8 box |
| | 110 | 110 | 100 | 100 | no compl. |
| | 110 | 110 | 100 | 100 | |
| 9-12 box no direct | 011 | 011 | 001 | 001 | |
| | 011 | 011 | 001 | 001 | 9-12 box |
| | 010 | 010 | 000 | 000 | no direct |
| | 010 | 010 | 000 | 000 | |
| 13-16 box no compl. | 011 | 011 | 001 | 001 | |
| | 011 | 011 | 001 | 001 | 13-16 box |
| | 010 | 010 | 000 | 000 | no compl. |
| | 010 | 010 | 000 | 000 | |

*b)*

**ISyGC table**

| | | | | |
|---|---|---|---|---|
| ***d*** | AGAG | AGAG | AGAG | AGAG |
| ***c*** | UCUC | UCUC | UCUC | UCUC |
| ***d*** | GAGA | GAGA | GAGA | GAGA |
| ***c*** | CUCU | CUCU | CUCU | CUCU |
| ***d*** | UCCU | UCCU | UCCU | UCCU |
| ***c*** | AGGA | AGGA | AGGA | AGGA |
| ***d*** | AGGA | AGGA | AGGA | AGGA |
| ***c*** | UCCU | UCCU | UCCU | UCCU |
| ***d*** | GAAG | AGGA | CUUC | UCCU |
| ***c*** | CUUC | UCCU | GAAG | AGGA |
| ***d*** | GAAG | AGGA | CUUC | UCCU |
| ***c*** | CUUC | UCCU | GAAG | AGGA |

**SGC table**

| | | | | |
|---|---|---|---|---|
| 1. base | UUUU | UUUU | UUUU | UUUU |
| | CCCC | CCCC | CCCC | CCCC |
| | AAAA | AAAA | AAAA | AAAA |
| | GGGG | GGGG | GGGG | GGGG |
| 2. base | UCAG | UCAG | UCAG | UCAG |
| | UCAG | UCAG | UCAG | UCAG |
| | UCAG | UCAG | UCAG | UCAG |
| | UCAG | UCAG | UCAG | UCAG |
| 3. base | UUUU | CCCC | AAAA | GGGG |
| | UUUU | CCCC | AAAA | GGGG |
| | UUUU | CCCC | AAAA | GGGG |
| | UUUU | CCCC | AAAA | GGGG |

**Fig. 2.** a). The unique fundamental purine-pyrimidine symmetry net of the ISyGC table, which is the common natural rule for all nuclear and mitochondrial genetic codes A), in comparison with the purine-pyrimidine net of the SGC table B). Red or black are denoted bases of codons with the same purine-pyrimidine profile. A) Alternating positions of boxes with direct and complement (compl.) denoted codons; bold: in the ISyGC table, all four columns have the same distribution of purine-pyrimidine profile, and simultaneously the same profile distribution pairs of codon rows within each box. B) In the Standard genetic code table there are also the same pairs of codon rows of purine/pyrimidine profile distribution, but the symmetry profile is not the same in four vertical columns but in the 1st and 2nd as well as in 3rd and 4th. There is also no direct/complement symmetry of codons between boxes as in our ISyGC table. b) An array of individual 1st, 2nd and 3rd base of each codon of the ISyGC table in comparison with the 1st, 2nd and 3rd base of codons in the SGC table. The bases of codons (1st, 2nd, 3rd) are ordered according to the arrangement of four boxes in each row as in the whole ISyGC table. Italic bold d – direct, italic bold c – complement. In our ISyGC table, there is a purine-pyrimidine Watson-Crick pairing symmetry between bases of each pair of rows. In the SGC table, Watson-Crick pairing does not exist except for the first base. The symmetry between all bases in the SGC table is only an aesthetic category.

## 3. Results

### *3.1. Recognition of the first two codon's bases*

Symmetries in our genetic code table and involvement in symmetries in all three bases within codons offer clarification as to how tRNA cognate anticodon can recognize synonymous codons during mRNA decoding for each individual amino acid. We excluded wobbling modifications (Crick, 1966; Agris et al., 2018) and alterations to the ribosomal structure which arise by the disproportion between numbers of mRNA codons and tRNA anticodons and where the third base of the codon plays a minimum role as a specificity determinant.

In our approach, differentiation of codons is possible with the recognition of an H-bonds profile of strong (C, G) and weak (A, U) bases.

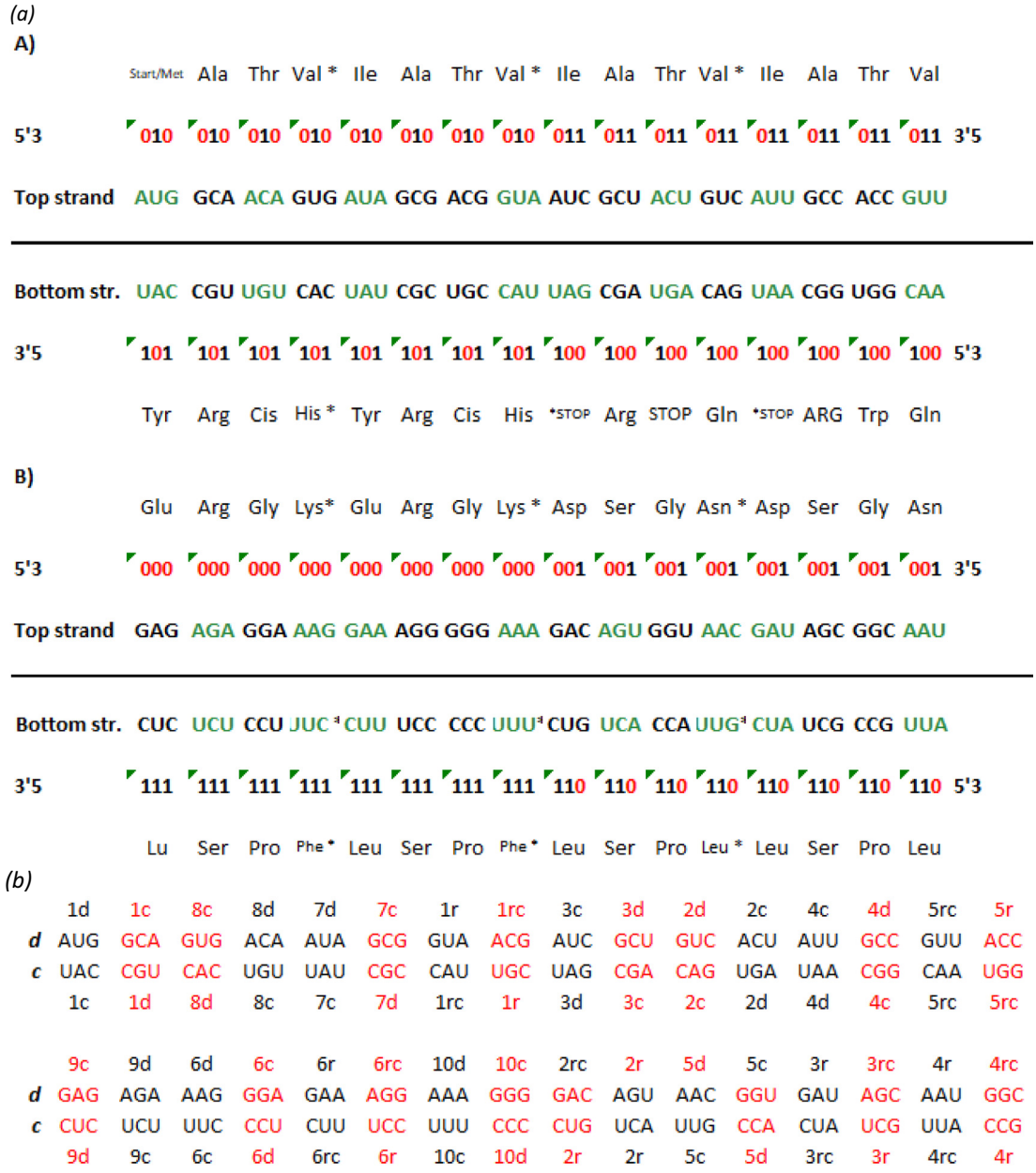


**Fig. 3.** (a) Transformation of the Ideal symmetry genetic code table into DNA form. 0 red – purine, 1 black – pyrimidine; *border between rows; green: A + U reach codons; black bold: C + G reach codons. All codons are ordered according to the arrangement in the ISyGC table with horizontal arraying of codons: A) in a top strand (direct codons from 1st to 4th box) and in a bottom strand (complement codons from 5th to 8th box), and B) in a top strand (direct codons from 9th to 12th box) and in a bottom strand (complement codons from 13th – 16th box). In each DNA strand, 8 by 8 codons in a line have the same purine-pyrimidine profile, which corresponds to pairs of rows in the ISyGC table. Purine ↔ purine and pyrimidine ↔ pyrimidine transformation alternatively arrayed A + U rich and C + G rich codons. In DNA form of the ISyGC table direct (top strand) and complement (bottom strand) and *vice versa* automatically represent a sequence akin to the 5′3 codon and 3′5 anticodon pattern, respectively. (b) Location of 10 A + U rich and 10C + G rich quadruplets in the DNA form of the ISyGC table. Black: A + U rich quadruplets; red: C + G rich quadruplets; italic bold d – direct, italic bold c – complement; d – direct, c – complement, r – reverse, rc – reverse complement. The enumeration of quadruplets is the same as in our classification of codons in Fig. 4. All codons are ordered according to the arrangement in the ISyGC table. There are alternately Watson Crick A + U rich and C + G rich direct ↔ complement pairs and reverse ↔ reverse complement pairs.

The first two bases in codons also reflected on the genetic code no-split boxes of strong bases CG, GC, CC, and GG. Their four codons in the box code the same amino acid. On the other hand, the boxes with weak first two bases AU, UA, AA, and UU are split. Thus, the ISyGC table is built of regularly arranged 8 no-split boxes, each for the same amino acid, and of 8 split boxes, each for 2 amino acids, one start signal, and 3 stop signals (Fig. 1).

In 2x2 no-split vertical boxes, the first two strong bases are C and/or G, mutually in a direct (d)-complement (c) relationship: GC (d) ↔ CG (c), GG (d) ↔ CC (c). The second two 2x2 no-split horizontal boxes have the first two mixed bases: CU and UC, GU and AC. It is evident that the first base contains all four possible bases (C, U, G, A), while the second base is always pyrimidine. It is important to stress that in all these 8 no-split boxes, the second base is not adenine (Figs. 1 and 5).

From 8 AU split vertical boxes, in 2x2 boxes the first two weak bases are also in a direct (d) ↔ complement (c) relationship: AU (d) ↔ UA (c), AA (d) ↔ UU (c)). The second 2x2 split horizontal

| A+U rich group (I) | | | | | C+G rich group (II) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| D | RC(D) | C(D) | R(D) | Sub-group | *D* | *RC(D)* | *C(D)* | *R(D)* | Sub-group |
| AUG 010 | CAU 101 | UAC 101 | GUA 010 | Ia | CGU 101 | ACG 010 | GCA 010 | UGC 101 | IIa |
| UGA 100 | UCA 110 | ACU 011 | AGU 001 | | GUC 011 | GAC 001 | CAG 100 | CUG 110 | |
| UAG 100 | CUA 110 | AUC 011 | GAU 001 | | GCU 011 | AGC 001 | CGA 100 | UCG 110 | |
| UAA 100 | UUA 110 | AUU 011 | AAU 001 | Ib | GCC 011 | GGC 001 | CGG 100 | CCG 110 | IIb |
| AAC 001 | GUU 011 | UUG 110 | CAA 100 | | CCA 110 | UGG 100 | GGU 001 | ACC 011 | |
| AAG 000 | CUU 111 | UUC 111 | GAA 000 | | CCU 111 | AGG 000 | GGA 000 | UCC 111 | |
| AUA 010 | UAU 101 | UAU* 101 | AUA* 010 | Ic | CGC 101 | GCG 010 | GCG* 010 | CGC* 101 | IIc |
| ACA 010 | UGU 101 | UGU* 101 | ACA* 010 | | CAC 101 | GUG 010 | GUG* 010 | CAC* 101 | |
| AGA 000 | UCU 111 | UCU* 111 | AGA* 000 | | CUC 111 | GAG 000 | GAG* 000 | CUC* 111 | |
| AAA 000 | UUU 111 | UUU* 111 | AAA* 000 | | CCC 111 | GGG 000 | GGG* 000 | CCC* 111 | |

**Fig. 4.** Our quadruplet classification of 61 codons and 3 stop signals. Purine 0, pyrimidine 1. Each quadruplet is unique and consists of codons denoted as direct (D) and its reverse complement (RC), complement (C), and reverse (R). Ten A + U rich (group I) and ten C + G rich (group II) quadruplets are organized into three subgroups. Ia consists of nonsymmetrical codons containing four different nucleotides, Ib consisting of nonsymmetrical codons containing two different nucleotides, Ic, symmetrical codons which contain duplicated codons labeled with an asterisk (D = R, C = RC). The first four A + U rich quadruplets we generated with start AUG, and stop UGA, UAG, and UAA signals. The C + G rich codons correspond to the purine-pyrimidine transformation of A + U rich codons within A (purine) in C (pyrimidine) and *vice versa*, and G (purine) in U (pyrimidine) and *vice versa*. Three symmetries are present in our codon classification: 1) purine-pyrimidine symmetries in each quadruplet, 2) purine -pyrimidine symmetries within and between A + U rich and C + G rich quadruplets in the same row; 3) a mirror symmetry between direct – reverse and complement – reverse complement in the same quadruplet. For clarity, the white and grey rows are alternating, in order to emphasize pairs of A + U rich and C + G rich codons.

boxes have two mixed bases: GA and AG, CA and UG. It is evident that the first base contains all four possible bases (G, A, C, U), while the second base is always purine. It is important to stress that in all these 8 split boxes, the second base is not cytosine (Figs. 1 and 5).

### 3.2. The role of the third base in split boxes

All split boxes are divided in the same way: the third base in one pair of codons within a split box are both purines (A, G, or *vice versa*), and in the other pair are both pyrimidines (U, C, or *vice versa*) (Figs. 1 and 6). Simultaneously, in horizontal mixed boxes, one of the codons in a pair is always A + U rich, and the other is C + G rich (Figs. 1 and 6). Such regularly distributed differences among codons are essential for identifying which amino acid belongs to an individual codon in two pairs of codons in the split boxes.

In the case of a pair of codons divided into a $2\times1$ codon for each amino acid, division takes place always between the third A ↔ G base, while the first two bases are identical as in the whole box (Figs. 1, 6, and 7). This rule is valid for coding of the only trinucleotide from start and stop signals and the only codon for Tryptophan.

### 3.3. Recognition of start/stop signals

In our ISyGC table, it can be seen that the locality of all three stop signals (UAG, UAA, UGA) is in the same corridor of the 5th- 8th boxes and takes place in the 3rd and 4th codons in the corresponding box (Fig. 1). Therefore, the relationship between purines (0) and pyrimidines (1), "1–0-0" is the same, and together with the start signal AUG "0–1–0", belong to A + U rich codons. It should be stressed that they do not contain cytosine. This differs from other codons with the same purine-pyrimidine ratio "1–0-0" in the same corridor. Stop signals UAG and UAA divide the same half of the 5th box, while the stop signal UGA shares its half in the 8th box with Tryptophan (UGG), which is C + G rich, determining a difference between them.

The start signal AUG is placed simultaneously as Methionine in the ISyGC table purine-pyrimidine net in the first position in the first box, while the other three codons in the same box code as Isoleucine (Ile). AUG has the same purine-pyrimidine ratio "0–1-0", as the other codons in the same corridor. However, these codons belong to the no-split boxes that distinguish them from AUG.

### 3.4. Differentiation for sextets

Problems in codon recognition arise for the sextets Leucine, Arginine, and Serine amino acids. For each of them, four codons are in a whole no-split box and the other two in a split box (Fig. 1). Due to this, they create the leading group of the ISyGC table and determine symmetrical positions of boxes in the whole non-leading group (Rosandić and Paar 2014, Rosandić et al. 2019).

Leucine (Leu) corresponds to the whole box with codons CUC, CUU, CUG, CUA. Two codons (UUG, UUA) belong to a split box in the nonleading group of the ISyGC table. It is seen that they have the same 2nd and 3rd base as the last two codons in the no-split

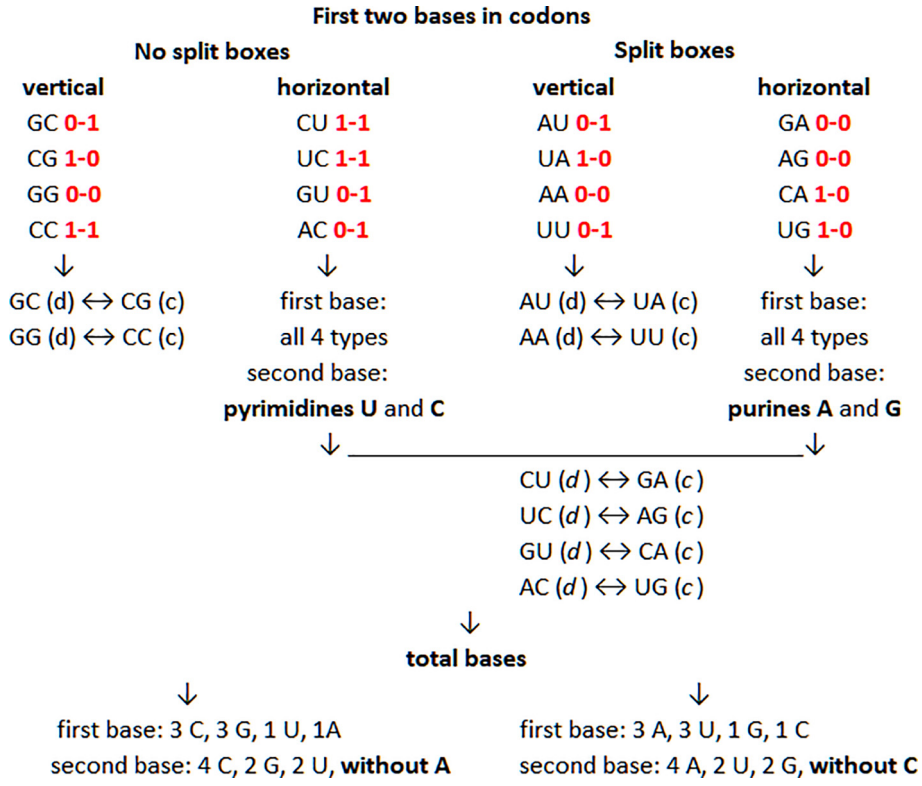


**Fig. 5.** Algorithmic determination of each codon for an individual amino acid with the first two bases of codons within all 16 boxes in the ISyGC table. All bases are ordered according to the arrangement in the ISyGC table with a horizontal arraying of codons. Purine and pyrimidine profiles of the first and second base in codons. 0, red: purine, 1, red: pyrimidine; d: direct; c: complement. Purine-pyrimidine and direct–complement symmetries between the first and second bases enable a distinction between all codons for individual amino acids in no split and split boxes. For sextet codons and start/stop signals, the third base has an important role (see the text and Figs. 6 and 7).

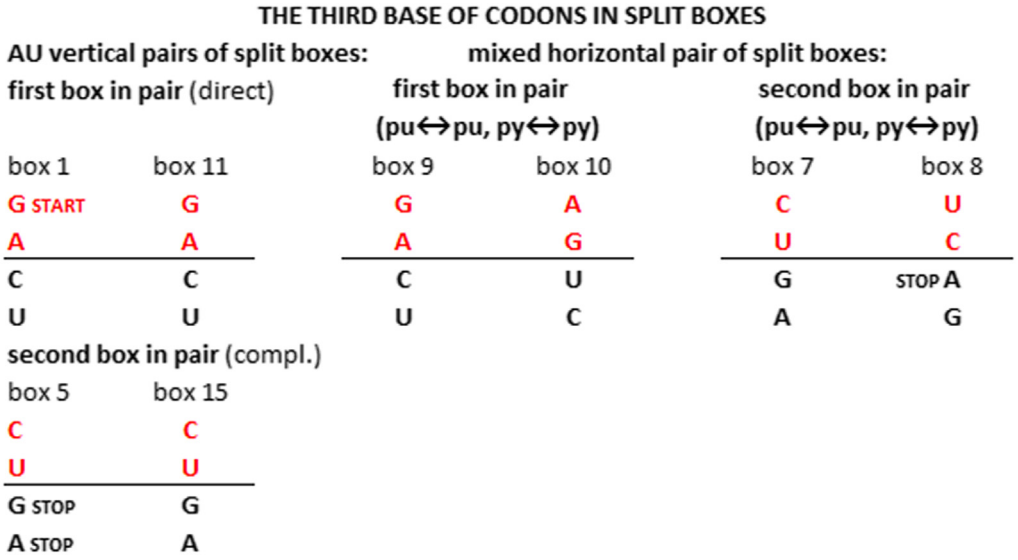


**Fig. 6.** Distribution of the purine and pyrimidine profile of the third base in two pairs of codons of split boxes. Red – first pair of codons in the box, black – second pair in the same box; d – direct; compl. – complement; pu – purine; py - pyrimidine, each pair belonging to a different amino acid. Distribution of the third base between vertical (direct ↔ compl.) (1 and 5, 11 and 15) and horizontal (pu ↔ pu, py ↔ py) (9 and 10, 7 and 8) pairs of split boxes in the ISyGC table; Each box in the ISyGC table is divided in the following way: one pair of codons contains only purines in the third base (G, A), and only pyrimidines in the second pair (C, U). This rule is valid for all boxes and is important for the differentiation of codons for two amino acids in split boxes. In each pair of codons of all mixed horizontal split boxes one codon is A + U rich, and the other C + G rich (Table 1). In all including split boxes of the ISyGC table, there is a direct (1st and 3rd base) ↔ complement (2nd and 4th base) relationship of the third base of codons within the same box. In start/stop signals, the third base is a purine.

box (underlined). In observing the whole ISyGC table, the only alternative could have been the start signal (AUG) and the codon for Isoleucine (AUA) from the first split box in the ISyGC table. Due to the importance of the start signal, it was probably created before the 5th and 6th codons for Leucine were in the process of creation.

The analog rule is valid for Arginine (Arg). Its four codons are included in the whole no-split box: CGU, CGC, CGA, CGG. The 5th

**Bases:** **Differentiation:**

**1st and 2nd -** vertical strong no split two pairs of boxes (d↔c relation)

- vertical weak split two pairs of boxes (d↔c relation)

- horizontal mix no split two pairs of boxes (pu ↔pu, py↔py relation)

- horizontal mix split two pairs of boxes (pu ↔pu, py↔py relation)

**3rd** - in one pair of codons for each amino acid in all split boxes are purines (A or G)

- in the other pair of codons for the other amino acid in the same boxes are pyrimidines (U or C)

- pair of codon splits only between purines (A, G) for start/stop signals and Trp

**2nd and 3rd** - identical in two codons of split box and two codons of four in no split box for sextets Arginine and Leucine

**1st, 2nd, 3rd**- all four codons as direct (in 1st and 3rd row of boxes) and complement (in 2nd and 4th row of boxes) in leading and nonleading group of codons of the ISyGC table

- as direct in two codons of four in no split box and their complement in two codons of split box for sextet Serine

- as direct and complement between pair of boxes of vertical strong no split and vertical weak split boxes

-as A+U rich and C+G rich pair of codons in horizontal mix split boxes

**Fig. 7.** Differentiation between no split and split boxes of codons in relationship to 1st, 2nd and 3rd bases in the ISyGC table. d direct; c complement; pu purine; py pyrimidine, Trp Tryptophan.

and 6th codons (AGA, AGG) have the same 2nd and 3rd bases (CGA, CGG) in the split box. The only alternative could have been the stop signal (UGA) and the only codon for Tryptophan (UGG) from the split box in the ISyGC table. Due to the importance of the stop signal, it was probably created before the 5th and 6th codons for Arginine were in the process of creation.

### 3.5. The unique role of Serine in the creation of the ISyGC table

Serine (Ser) has a different selection of codons. The whole no-split box for Serine includes codons UCU, UCC, UCA, and UCG. Their 2nd and 3rd bases are CU, CC, CA and CG, correspond to the no-split boxes for Alanine (Ala), Threonine (Thr), and Proline (Pro) amino acids (Fig. 1). As a result, Serine could not have obtained the 5th and 6th codons from these boxes. Instead, it received AGU and AGC from the split box, shared with the 5th and 6th codons from neighboring Arginine. Both boxes containing codons for Serine are in a unique mutual relationship as direct and complement (AGU (d) ↔ UCA (c), AGC (d) ↔ UCG (c)) and thus are connected and recognizable. This relationship positions *in continuo* codons for Arginine and Serine in the same column from the leading group in the ISyGC table (Fig. 1). Simultaneously, with a purine-purine and pyrimidine-pyrimidine transformation, Serine positions codons also for Leucine. In this way, Serine determined the positions of all the codons in the leading group of the ISyGC table, which also creates codon distribution in the nonleading group (Rosandić and Paar 2014, Rosandić et al. 2019).

Serine also has important roles in relation to Asparagine and Threonine forming glucoproteins, and it is involved in the regulation of energy metabolism and fuel storage in the body (Taniguchi and Nagai 2014). Therefore, Serine is found in most proteins. As one of ten prebiotics, it has been identified in meteorites (Novozhilov et al. 2007). The Mitochondrial genetic codes for Invertebrate, Trematode, Echinoderm flatworm, and Alternative flatworm even have eight codons for Serine (Fig. 8B).

## 4. Discussion

### 4.1. The ISyGC table and genetic code variations

Up to the present, more than 30 slightly alternative nuclear and mitochondrial genetic codes have been detected. Our novel ISyGC table contains a unique fundamental common purine-pyrimidine symmetry net for all variations of nuclear and mitochondrial genetic codes (Fig. 2A). It is structured on the basis of the natural triplet symmetries among codons: purine-pyrimidine symmetry, A + T rich / C + G rich symmetry, and direct ↔ complement symmetry. Nuclear genetic codes usually have only one change at the expense of usurping stop signals, while mitochondrial genomes have more changes (Shu, 2017; Žihala and Eliaš, 2019) (Fig. 8). Examples of, for example, the mitochondrial Vertebrate code,

**a)**

| | Leading group | | | | | | Nonleading group | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Box | Amino acid | I. Codons | Pu/Py | Pu/Py | II. Codons | Amino acid | Amino acid | III. Codons | Pu/Py | Pu/Py | IV. Codons | Amino acid |
| Direct boxes 1-4 | Start/M | *AUG* | 010 | 010 | GCA | Ala | Thr | *ACA* | 010 | 010 | GUG | Val |
| | Met | *AUA* | 010 | 010 | GCG | | | ACG | 010 | 010 | *GUA* | |
| | Ile | *AUC* | 011 | 011 | GCU | | | *ACU* | 011 | 011 | GUC | |
| | | *AUU* | 011 | 011 | GCC | | | ACC | 011 | 011 | *GUU* | |
| Complement boxes 5-8 | Tyr | *UAC* | 101 | 101 | CGU | Arg | Cys | *UGU* | 101 | 101 | CAC | His |
| | | *UAU* | 101 | 101 | CGC | | | UGC | 101 | 101 | *CAU* | |
| | Stop | *UAG* | 100 | 100 | CGA | | Trp | *UGA* | 100 | 100 | CAG | Gln |
| | Stop | *UAA* | 100 | 100 | CGG | | Trp | *UGG* | 100 | 100 | *CAA* | |
| Direct boxes 9-12 | Glu | GAG | 000 | 000 | *AGA* | STOP | Gly | GGA | 000 | 000 | *A AG* | Lys |
| | | *GAA* | 000 | 000 | AGG | | | GGG | 000 | 000 | *AAA* | |
| | Asp | GAC | 001 | 001 | *AGU* | Ser | | GGU | 001 | 001 | *AAC* | Asn |
| | | *GAU* | 001 | 001 | *AGC* | | | GCC | 001 | 111 | *AAU* | |
| Complement boxes 13-16 | Leu | CUC | 111 | 111 | *UCU* | Ser | Pro | CCU | 111 | 111 | *UUC* | Phe |
| | | *CUU* | 111 | 111 | UCC | | | CCC | 111 | 111 | *UUU* | |
| | | CUG | 110 | 110 | *UCA* | | | CCA | 110 | 110 | *UUG* | Leu |
| | | *CUA* | 110 | 110 | UCG | | | CCG | 110 | 110 | *UUA* | |

**b)**

| | Leading group | | | | | | Nonleading group | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Box | Ile | I. Codons | Pu/Py | Pu/Py | II. Codons | Amino acid | Amino acid | III. Codons | Pu/Py | Pu/Py | IV. Codons | Amino acid |
| Direct boxes 1-4 | Start/M | *AUG* | 010 | 010 | GCA | Ala | Thr | *ACA* | 010 | 010 | GUG | Val |
| | Met | *AUA* | 010 | 010 | GCG | | | ACG | 010 | 010 | *GUA* | |
| | Ile | *AUC* | 011 | 011 | GCU | | | *ACU* | 011 | 011 | GUC | |
| | | *AUU* | 011 | 011 | GCC | | | ACC | 011 | 011 | *GUU* | |
| Complement boxes 5-8 | Tyr | *UAC* | 101 | 101 | CGU | Arg | Cys | *UGU* | 101 | 101 | CAC | His |
| | | *UAU* | 101 | 101 | CGC | | | UGC | 101 | 101 | *CAU* | |
| | Stop | *UAG* | 100 | 100 | CGA | | Stop | *UGA* | 100 | 100 | CAG | Gln |
| | Stop | *UAA* | 100 | 100 | CGG | | Trp | *UGG* | 100 | 100 | *CAA* | |
| Direct boxes 9-12 | Glu | GAG | 000 | 000 | *AGA* | Ser | Gly | GGA | 000 | 000 | *A AG* | Lys |
| | | *GAA* | 000 | 000 | AGG | | | GGG | 000 | 000 | *AAA* | |
| | Asp | GAC | 001 | 001 | *AGU* | Ser | | GGU | 001 | 001 | *AAC* | Asn |
| | | *GAU* | 001 | 001 | *AGC* | | | GGC | 001 | 001 | *AAU* | |
| Complement boxes 13-16 | Leu | CUC | 111 | 111 | *UCU* | Ser | Pro | CCU | 111 | 111 | *UUC* | Phe |
| | | *CUU* | 111 | 111 | UCC | | | CCC | 111 | 111 | *UUU* | |
| | | CUG | 110 | 110 | *UCA* | | | CCA | 110 | 110 | *UUG* | Leu |
| | | *CUA* | 110 | 110 | UCG | | | CCG | 110 | 110 | *UUA* | |

**c)**

| | Leading group | | | | | | Nonleading group | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Box | Amino acid | I. Codons | Pu/Py | Pu/Py | II. Codons | Amino acid | Amino acid | III. Codons | Pu/Py | Pu/Py | IV. Codons | Amino acid |
| Direct boxes 1-4 | Start/M | *AUG* | 010 | 010 | GCA | Ala | Thr | *ACA* | 010 | 010 | GUG | Val |
| | Met | *AUA* | 010 | 010 | GCG | | | ACG | 010 | 010 | *GUA* | |
| | Ile | *AUC* | 011 | 011 | GCU | | | *ACU* | 011 | 011 | GUC | |
| | | *AUU* | 011 | 011 | GCC | | | ACC | 011 | 011 | *GUU* | |
| Complement boxes 5-8 | Tyr | *UAC* | 101 | 101 | CGU | Arg | Cys | *UGU* | 101 | 101 | CAC | His |
| | | *UAU* | 101 | 101 | CGC | | | UGC | 101 | 101 | *CAU* | |
| | Stop | *UAG* | 100 | 100 | CGA | | Trp | *UGA* | 100 | 100 | CAG | Gln |
| | Stop | *UAA* | 100 | 100 | CGG | | Trp | *UGG* | 100 | 100 | *CAA* | |
| Direct boxes 9-12 | Glu | GAG | 000 | 000 | *AGA* | Arg | Gly | GGA | 000 | 000 | *A AG* | Lys |
| | | *GAA* | 000 | 000 | AGG | | | GGG | 000 | 000 | *AAA* | |
| | Asp | GAC | 001 | 001 | *AGU* | Ser | | GGU | 111 | 111 | *AAC* | Asn |
| | | *GAU* | 001 | 001 | *AGC* | | | GGC | 111 | 111 | *AAU* | |
| Complement boxes 13-16 | Thr | *CUC* | 111 | 111 | *UCU* | Ser | Pro | CCU | 111 | 111 | *UUC* | Phe |
| | | *CUU* | 111 | 111 | UCC | | | CCC | 111 | 111 | *UUU* | |
| | | *CUG* | 110 | 110 | *UCA* | | | CCA | 110 | 110 | *UUG* | Leu |
| | | *CUA* | 110 | 110 | UCG | | | CCG | 110 | 110 | *UUA* | |

**Fig. 8.** (a) The mitochondrial vertebrate code incorporated in the ISyGC table. Methionine (M, Met) expands to the neighboring Isoleucine (Ile) codon AUA; Tryptophan (Trp) expands to the neighboring stop UGA codon; Arginine (Arg) AGA and AGG codons become Stop signals. (b) The mitochondrial invertebrate code incorporated in the ISyGC table. Methionine (Met) expands to the neighboring Isoleucine (Ile) codon AUA; Serine (Ser) as octet expands to neighboring Arginine (Arg) codons AGA and AGG in the whole box 10. (c) The mitochondrial yeast code incorporated in the ISyGC table. Methionine (M, Met) expands to neighboring Isoleucine (Ile), Tryptophan (Trp) expands to neighboring stop signal UGA. Threonine (Thr) with four codons expands to the whole box from Leucine (Leu) which is in the neighboring position if we put the Nonleading group below the Leading group. In different genetic codes, individual amino acids usually capture a codon from a neighboring amino acid. All three examples of mitochondrial genetic codes show that our novel ISyGC table has the common unchangeable fundamental purine ↔ pyrimidine symmetry net structure. In addition, in the ISyGC table all codons with their direct ↔ complement, purine ↔ pyrimidine and A + U rich and C + G rich symmetries are incorporated in the symmetry purine-pyrimidine net. The usurpation of some codons by some amino acid points to a larger metabolic requirement for individual amino acids, regardless of whether it is for nuclear or mitochondrial genetic codes.

Invertebrate code, and Yeast code demonstrate that variations of the number of codons for individual amino acids, inserted into the ISyGC table, arise most often by a capture from a neighboring codon from a weak split box or a whole box (Fig. 8). A change can arise from a direct to reverse complement of codons: one of six codons of Leucine CUA (direct) usurped the UAG (reverse complement) stop signal (Chlorophycean code and Scenedesmus obliquus code), the stop signal UAA (direct) usurped UUA (reverse complement) from Leucine (Thraustochytrium code). Thereby, the ISyGC table and the fundamental common purine-pyrimidine symmetry net remain preserved. The most unstable A + U rich codons are the start AUG and stop UGA, UAG and UAA signals and two codons AGA and AGG from the most harmful Arginine (Novozhilov et al. 2007) in weak split boxes and there are most often usurped. Namely, Arginine has the minimum attainable robustness of all amino acids. Such usurpation points to a larger metabolic requirement for individual amino acids, regardless of whether it is for nuclear or mitochondrial genetic codes.

Translation is not limited to twenty amino acids. Additional Selenocysteine (Wong et al. 2016) as the 21st amino acid takes the same possession of UGA stop signal and Pyrolysine (Wong et al. 2016) as does the 22nd amino acid of UAG stop signal in the ISyGC table, but the basic purine – pyrimidine symmetry net also remains untouched.

### *4.2. Unique symmetry net for RNA and DNA viruses and other DNA species*

DNA viruses have also the ISyGC table with the same common fundamental physicochemical purine-pyrimidine symmetry net (Fig. 2A) as all DNA species (prokaryotes and eukaryotes).

RNA viruses contain only one strand of ribonucleic acid. Trinucleotides in the RNA molecule have the structure of triplet bases as in the DNA molecule. However, it is impossible for one strand of an RNA molecule to have Watson-Crick pairing (A ↔ T, C ↔ G) and strand symmetry (Chargaff's second parity rule) (Rosandić et al., 2016, 2019) for which both strands as in the DNA molecule are necessary. Despite this, codons of RNA viruses create an identical ISyGC table with a common purine-pyrimidine symmetry net as codons in the DNA species. It means that the unique purine-pyrimidine symmetry net in the ISyGC table of RNA viruses has a direction for DNA transformation (Fig. 3(a)). In such a way, the purine-pyrimidine symmetry net is a key for the common basic ISyGC table of RNA and DNA molecular physicochemical symmetry structure with purine-pyrimidine A ↔ U, C ↔ G pairing which enabled finger-post for evolution from an RNA to a DNA molecule. During the whole evolution the purine-pyrimidine symmetry net of ISyGC table for all species remains unchanged!

One important criterion is whether the model is able to accommodate, in a natural way not only the SGC table but also nonstandard codes like some nuclear and mitochondrial. It is the case of the ISyGC table and its purine-pyrimidine symmetry net (Fig. 8).

We determine that the ISyGC table net is the basic unchangeable purine-pyrimidine symmetry structure of genetic code regardless of which codon belongs to which amino acids in different genetic codes. In this way namely, the purine-pyrimidine symmetry net is a natural law according to physicochemical properties of purines and pyrimidines, incorporated in the ISyGC table, as well as in the other genetic codes. Amino acids capture the corresponding codons according to their metabolic requirements for proteins of species and organelles such as mitochondria (Fig. 8). In this framework, all genetic codes that will be discovered in the future will also contain the same purine-pyrimidine symmetry net.

We show with the algorithms of nitrogenous bases in codons in the ISyGC table that the purine-pyrimidine symmetry net with purine-pyrimidine A ↔ U, C ↔ G pairing at the same time is the key for clarification as to how tRNA cognate anticodon can recognize synonymous codons during mRNA decoding for each individual amino acid.

Localization of each codon in the ISyGC table is strictly determined because of the universal purine-pyrimidine symmetry net. Since all variations of genetic codes have basically the same purine-pyrimidine symmetry net as in the ISyGC table, differentiations of codons for amino acids are also common: the same boxes are recognized according to the first two bases in codons, and, according to the same rules, are distributed individual boxes and their halves.

In general, the role of symmetries as a dominant concept in the fundamental laws of physics was reviewed by Gross (1996) in a paper presented at a colloquium entitled "Symmetries throughout the sciences" organized at the National Academy of Sciences USA. Some of the emphasis was as follows: Newton's laws embodied symmetry principles, notably the principle of equivalence of inertial frames (Galilean invariance). These symmetries implied the conservation laws. In the past the conservation laws, especially those of momentum and energy which are of fundamental importance, were first regarded as consequences of the dynamical laws of nature rather than as consequences of the symmetries that underlay these laws. Einstein's great advance was to put symmetry first, to regard the symmetry principle as the primary feature of nature that constrains the dynamical laws. At the beginning of the 20th century, Emmy Noether proved her famous theorem relating to symmetry and conservation laws (Noether 1918). In the latter half of the 20th century, symmetry has been the most dominant concept in the exploration and formulation of the fundamental laws of physics. Today it serves as a guiding principle in the search for further unification and progress. We realize that the symmetry principles dictate the form of laws of nature (Gross, 1996).

The secret of nature is symmetry, but there are a variety of mechanisms wherein the symmetry of nature can be hidden or broken. Heisenberg's isotopic symmetry of the nuclear force is an example of approximate symmetry (Heisenberg 1932). At a fundamental level nature, for whatever reasons, prefers beauty and is marvelously inventive in inventing new forms of beauty. When searching for new and more fundamental laws of nature we should search for new symmetries (Gross, 1996).

The concept of symmetry group and symmetry breaking was introduced by Heisenberg in 1932 for proton and neutron (Thiessen and Ceulemans, 2020). Heisenberg considered that symmetries represent the fundamental level of reality, whereas particles constitute a secondary level of reality.

In the symmetry-breaking study of the genetic code (Hornos and Hornos, 1993; Hornos et al., 1999), built on the basis of the SGC, it was assumed that the symmetry breaking along the sp(6) chain of subgroups coincides with the evolutionary tree. This "Progressive symmetry breaking theory" explained the observed degeneracies of the genetic code (Antoneli et al., 2010; Antoneli and Forger, 2011; Bashford et al., 1998; Forger et al., 1997; Hornos et al., 1999, 2004; Hornos and Hornos, 1993; Kent et al., 1998). It was pointed out that symmetry breaking is a mathematical technique for organizing the group-theoretical structure.

Other approaches in studying symmetries in standard genetic code involve the search for symmetry distributions of codons and amino acids within genetic code and their presentations (Ahmed et al., 2010; Jose et al., 2017; Michel and Pirillo, 2010; Nemzer, 2017; Seligmann and Warthi, 2017; Shu, 2017; Štambuk and Konjevoda, 2020).

On the other hand, the "Symmetry theory" (Rosandić et al. 2013, Rosandić et al., 2016; Rosandić et al., 2019; Rosandić and Paar, 2014) is built on the basis of a unique physicochemical purine-pyrimidine symmetry net between codons of the ISyGC

table which is identical for all genetic codes. This unique symmetry net does not depend on the fact which codons are assigned to an individual amino acid. Therefore, the ISyGC table exhibits robustness of purine-pyrimidine symmetry for all species' and opens a possibility of a new look at the evolution. Symmetries are the main road to restrict disorder (entropy increase) which enables evolution and existence of living organisms.

Namely, our investigation shows that the evolution of species proceeds according to the physicochemical laws of DNA quadruplet purine-pyrimidine symmetries (Rosandić et al. 2019) and the common purine-pyrimidine symmetry net of all genetic codes. In this way symmetries are essential for the creation of life. We can conclude that evolution is a road paved with symmetries.

## CRediT authorship contribution statement

**Marija Rosandić:** Conceptualization, Methodology, Investigation, Visualization, Writing - original draft. **Vladimir Paar:** Investigation, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We thank the reviewers for very helpful comments and stimulating suggestions. This work has been fully supported by the Croatian Science Foundation under the project IP-2019-04-2757.

## References

Agris, P.F., Eruysa, E.R., Narendran, A., Vare, V.Y.P., 2018. Celebrating wobble decoding: Half a century and still much is new. RNA Biol. https://doi.org/10.1080/15476286.2017.1356562.

Ahmed, A., Frey, G., Michel, C.J., 2010. Essential molecular functions associated with circular code evolution. J. Theor. Biol. https://doi.org/10.1016/j.jtbi.2010.02.006.

Antoneli, F., Gaviria, P.A., Forger, M., Hornos, J.E.M., 2010. On amino acid and codon assignment in algebraic models for the genetic code. J. Mod. Phys Int. https://doi.org/10.1142/S0217979210054944.

Antoneli, F., Forger, M., 2011. Symmetry breaking in the genetic code: Finite groups. Comput. Modelling Math. https://doi.org/10.1016/j.mcm.2010.03.050.

Bashford, J.D., Tsohantjis, I., Jarvis, P.D., 1998. A supersymmetric model for the evolution of the genetic code. Natl. Acad. Sci. USA Proc. https://doi.org/10.1073/pnas.95.3.987.

Berleant, D., White, M., Pierce, E., Tudoreanu, E., Boeszoermenyi, A., Shtridelman, J., Macosko, J.C., 2009. The genetic code – more than just a table. Biophys. Cell Biochem. https://doi.org/10.1007/s12013-009-9060-9.

Castro-Chavez, F.A., 2012. A tetrahedral representation of the genetic code emphasizing aspects of symmetry, BioComplexity, 10.5048/BIO-C.2012.2.

Crick, F.H., 1966. Codon-anticodon pairing: wobble hypothesis. J. Mol. Biol. https://doi.org/10.1016/0022-2836(66)80022-0.

Crick, F.H., 1968. Evolution of the genetic code. J. Mol. Biol. https://doi.org/10.1016/0022-2836(68)90392-6.

Forger, M., Hornos, Y.M.M., Hornos, J.E.M., 1997. Global aspects in the algebraic approach to the genetic code. Rev. E Phys. https://doi.org/10.1103/Phys.Rev.E.56.7078.

Freeland, S.J., Hurst, I.D., 1998. The genetic code is one in a million. J. Mol. Evol. https://doi.org/10.1007/pl00006381.

Freeland, S.J., Knight, R.D., Landwebwr, L.F., Hurst, L.D., 2000. Early fixation of an optimal genetic code. Mol. Biol. Evol. https://doi.org/10.1093/oxfordjournals.molbev.a026331.

Gamow, G., 1954. Possible relation between deoxyribonucleic acid and protein structure. Nature. https://doi.org/10.1038/173318a0.

Grosjean, H., Westhof, E., 2016a. An integrated structure and energy-based view on the 93.25genetic code. Nucleic Acid Res.

Grosjean, H., Westhof, E., 2016b. An integrated structure and energy-based view on the genetic code. Nucleic Acids Res. https://doi.org/10.1093/nar/gkw608.

Heisenberg, W., 1932. Über den Bau der Atomkerne. 2.Physik 77, 1–11.

Hornos, J.E.M., Hornos, Y.M.M., 1993. Algebraic model for the evolution of the genetic code. Rev. Lett. Phys. https://doi.org/10.1103/Phys. Rev.Lett.71.4401.

Hornos, J.E.M., Hornos, Y.M.M., Forger, M., 1999. Symmetry and symmetry breaking: an algebraic approach to the genetic code. J. Mod. Phys Int. https://doi.org/10.1142/S021797929900268X.

Hornos, J.E.M., Braggion, L., Magini, M., Forger, M., 2004. Symmetry preservation in the evolution of the genetic code. IUBMB Life. https://doi.org/10.1080/15216540410001687838.

Jose, M.V., Zamudio, G.S., Morgado, E.R., 2017. A unified model of the standard genetic code. R. Soc. Open Sci. https://doi.org/10.1098/sos.160908.

Kent, R.D., Schlesinger, M., Wybourne, B.G., 1998. On algebraic approachjes to the genetic code. Canad. J. Phys. 1998, 236234510.

Koonin, E.W., Novozhilov, A.S., 2009. Origin and evolution of the genetic code: the universal enigma. IUBMB Life. https://doi.org/10.1002/iub.146.

Koonin, E.W., 2017. Frozen accident – pushing 50: stereochemistry, expansion and chance in the evolution of the genetic code. Life. https://doi.org/10.103390/life/7020022.

Kubyshkin, V., Acevedo-Rocha, C.G., Budisa, N., 2018. On universal coding events in protein biogenesis. BioSystems. https://doi.org/10.1016/j.biosystem.2017.10.004.

Lenstra, R., 2014. Evolution of the genetic code through progressive symmetry breaking. J. Theor. Biol. https://doi.org/10.1016/j.jtbc.2014.01.002.

Michel, C.J., 2013. Circular code motifs in transfer RNAs. Comput. Biol. Chem. https://doi.org/10.1016/j.compbiolchem.2010.03.004.

Michel, C.J., Pirillo, G., 2010. Identification of all trinucleotide circular codes. Biol. Chem. Comput. https://doi.org/10.1016/j.compbiolchem.2010.03.004.

Nemzer, L.R., 2017. A binary representation of the genetic code. BioSystems. https://doi.org/10.1016/j.biosystems.2017.03.

Noether, E., 1918. Invariante Variationsprobleme. Nach Kgl.Ges.Wiss. Göthingen, Math.Phys.Kl. 1918, 235–257

Nirenberg, M.W., Matthaei, J.H., 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally polyribonucleotides. Proceed Nat. Acad. Sci. USA. https://doi.org/10.1073/pnas.47.10.1588.

Novozhilov, A.S., Wolf, Y.I., Koonin, E.V., 2007. Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. Biology Direct. https://doi.org/10.1186/1745-6150-2-24.

Rosandić, M., Paar, V., Glunčić, M., 2013. Fundamental role of start/stop regulators in whole DNA and new trinucleotide classification. Gene. https://doi.org/10.1016/j.gene.2013.09.021.

Rosandić, M., Paar, V., 2014. Codon sextet with leading role of serine create "ideal" symmetry classification scheme of the genetic code. Gene. https://doi.org/10.1016/j.gene.2014.04.009.

Rosandić, M., Vlahović, I., Glunčić, M., Paar, V., 2016. Trinuclotide's quadruplet symmetries and natural symmetry law of DNA creation ensuing Chargaff's second parity rule. J. Biomol. Struct. Dyn. https://doi.org/10.1080/07391102.2015.1080628.

Rosandić, M., Vlahović, I., Paar, V., 2019. Novel look at DNA and life – symmetry as evolutionary forcing. J. Theor. Biol. 483, (2019). https://doi.org/10.1016/j.tbi.2019.08.016 109985.

Seligmann, H., Warthi, G., 2017. Genetic code optimization for cotranslational protein folding: codon directional asymmetry corelate with antiparallel betasheets tRNA synthetase classes. Comput. Struct. Biotechnol. J. https://doi.org/10.1016/j.csbj.2017.08.001.

Shu, J.-J., 2017. A new integrated symmetrical table for genetic codes. BioSystems. https://doi.org/10.1016/biosystems.2016.11.004.

Stewart, I., 1994. Broken symmetry in the genetic code?. New Sci. 141, 16.

Štambuk, N., 2000. Universal metric properties of the genetic code Croat. Chem. Acta. hrčak.srce.hr/131993.

Štambuk, N., Konjevoda, N., 2020. Determining amino acid scores of the genetic code table: complementarity, structure function and evolution. BioSystems. https://doi.org/10.1016/biosystems.2019.104026.

Taniguchi, N., Nagai, R., 2014. Amino acids and proteins. In: Baynes, J.W., Dominiczak, M.H. (Eds.), Medical biochemistry. fourth edition. Saunders Elsevier, Edinburg.

Thiessen, P., Ceulemans, A., 2020. Particular symmetries: group theory of the periodic system. Substan. Int. J. History Chem. 4, 7–22.

Wong, J.T., Ng, S.K., Mat, W.K., Hu, T., Xue, H., 2016. Coevolution theory of the genetic code at age forty: pathway to translation and synthetic life. Life Basel (Basel). https://doi.org/10.3390/life 6010012.

Žihala, D., Eliaš, M., 2019. Evolution and unprecedented variants of the mitochondrial genetic code in lineage of green algae. Genome. Biol. Evol. https://doi.org/10.1093/gbe/evz210.