ROOTS OF COMPLEXITY IN THE SELF-REFERENTIAL GENETIC CODE

Romeu Cardoso Guimarães

Laboratório de Biodiversidade e Evolução Molecular, Departamento de Genética,
Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Av Pres Antonio Carlos 6627,
31270-901 Belo Horizonte MG Brazil. TelFax +55-31-32744988, Mobile +55-31-988976439
romeucardosoguimaraes@gmail.com , romeucg@icb.ufmg.br

Abstract The genetic code is the correspondence between 'letter'-units that cells utilize for translation: triplets of bases in the producers (genes), amino acids in the products (proteins). The self-referential model indicates that the codes resulted from Proto-tRNA Dimer-Directed Protein Synthesis. The dimerized proto-tRNAs became codes when the peptides they produced bound back to them and stabilized the correspondence between the units and the protein production system. Anticodons are representative sites of the initial binding oligomers, that guided the complementariness at dimerization. The process of producing stabilized associations is a 'dynamic, epigenetic kind of memory'. The associated system is a module for the construction of polymers – genes, in the realm of 'memories in strings'. Memories guarantee stability while plasticity refers to the dynamics, which are the two main and interdependent characters of the living. Further stabilization and partial autonomy come from diversity in proteins at construction of structures and functions for the metabolic flow network. The metabolic system remains dependent on the environment, in a tense relationship with the degradation it provokes. A necessary component of biological complexity is the plasticity in behaviors, which mediates the diversity, adaptations and open-ended evolution. It is constitutive to protein structures and functions. Plastic behaviors are enhanced through the network organization of the system. Interactions that build networks are dependent on the wide range adhesiveness and binding sites of proteins. The model indicates that networks of nucleoprotein interactions are superposed on those of anticodon dimers, while all components are polymers with variable sequences. The complex behaviors of the resulting Multi-Synthetase Complexes are now minimally rationalized.

Key-words Genetic code, Self-reference, Coherence-decoherence, Memory, Metabolic flow, Plasticity, Networks, tRNA dimers, Multi-Aminoacyl-tRNA Synthetase Complex, Cohesiveness.

Graphical abstract (Left, bottom) **The encoding process** A module is a simple network of tRNA pairs. The first module encodes the anticodon pairs (1) 5´GGG:YCC3´and (3) GGA:YCU. Complex are the many cycles of evolutionary adjustments between the sequences of the synthetases and of the tRNAs. (Right, top) **The ribonucleoprotein network in the Multi-Synthetase Complex of mammals** There are nine enzymes [highlighted pink] and three auxiliary proteins [numbered]. Interactions via anticodon pairs are in straight lines, via protein binding in hand-drawn curved lines. The latter are more abundant upon the subnetworks that are less connected through the tRNA pairs and that join central A:U triplets.

The encoding process A module is a simple network of tRNA pairs. The first module encodes the anticodon pairs (1) 5´GGG:YCC3´and (3) GGA:YCU. Complex are the many cycles of evolutionary adjustments between the sequences of the synthetases and of the tRNAs.

The ribonucleoprotein network in the Multi-Synthetase Complex of mammals There are nine enzymes [highlighted pink] and three auxiliary proteins [numbered]. Interactions via anticodon pairs are in straight lines, via protein binding in hand-drawn curved lines. The latter are more abundant upon the subnetworks that are less connected through the tRNA pairs and that join central A:U triplets.
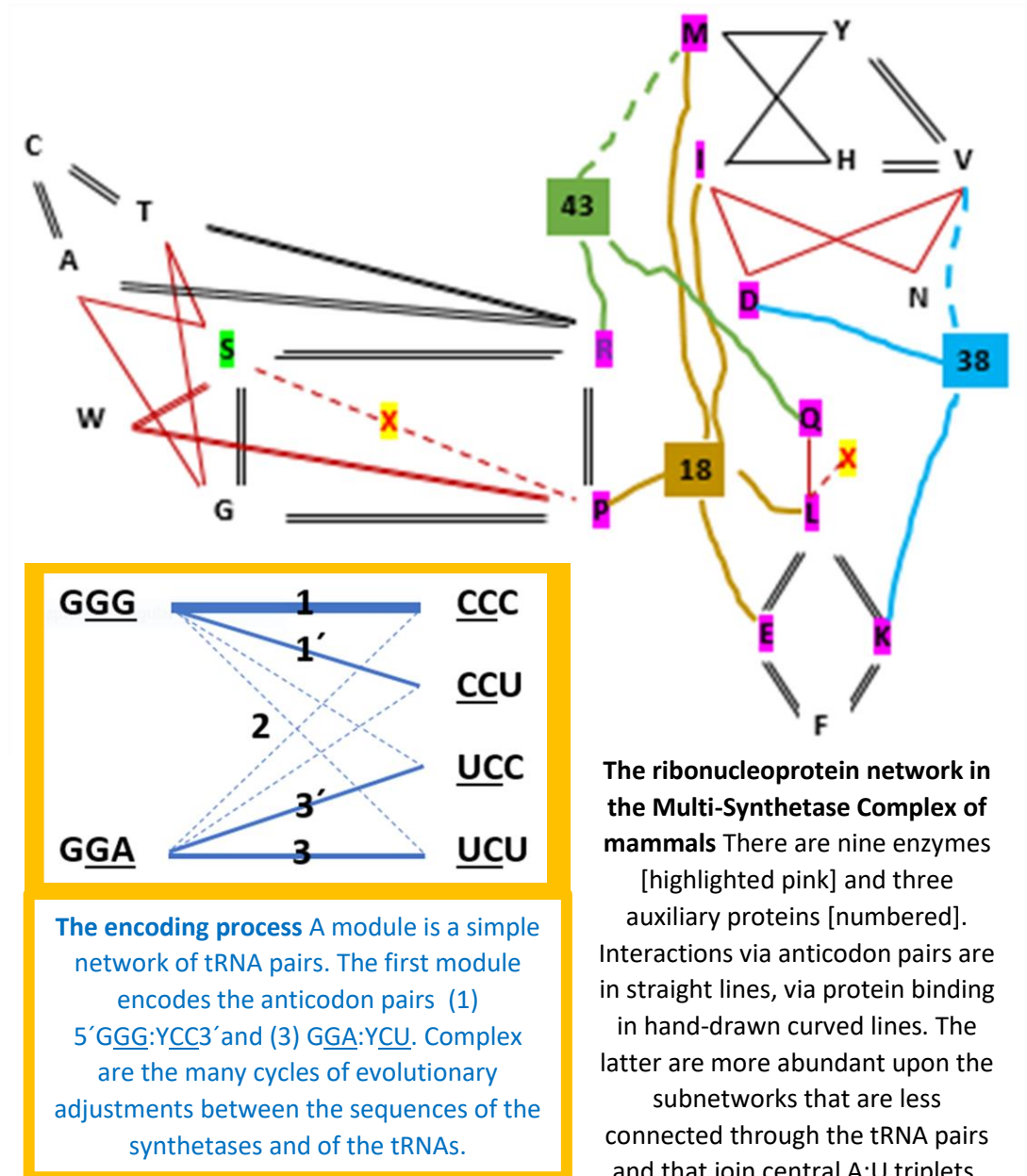
Graphical abstract

Introduction The genetic code is the set of correspondences between the aminoacyl-tRNA synthetases (synthetases) and their substrates tRNAs and amino acids. The working products of the correspondences, the aminoacyl-tRNAs, are represented by the tRNA anticodon triplet codes and the amino acids – the meanings – that are carried by the tRNAs. The correspondences have been settled following functional necessities of the cellular system but from the strictly biochemical point of view they may be considered nearly symbolic or arbitrary because there are no strong evidences of chemical relatedness between codes and meanings (Figure 1).
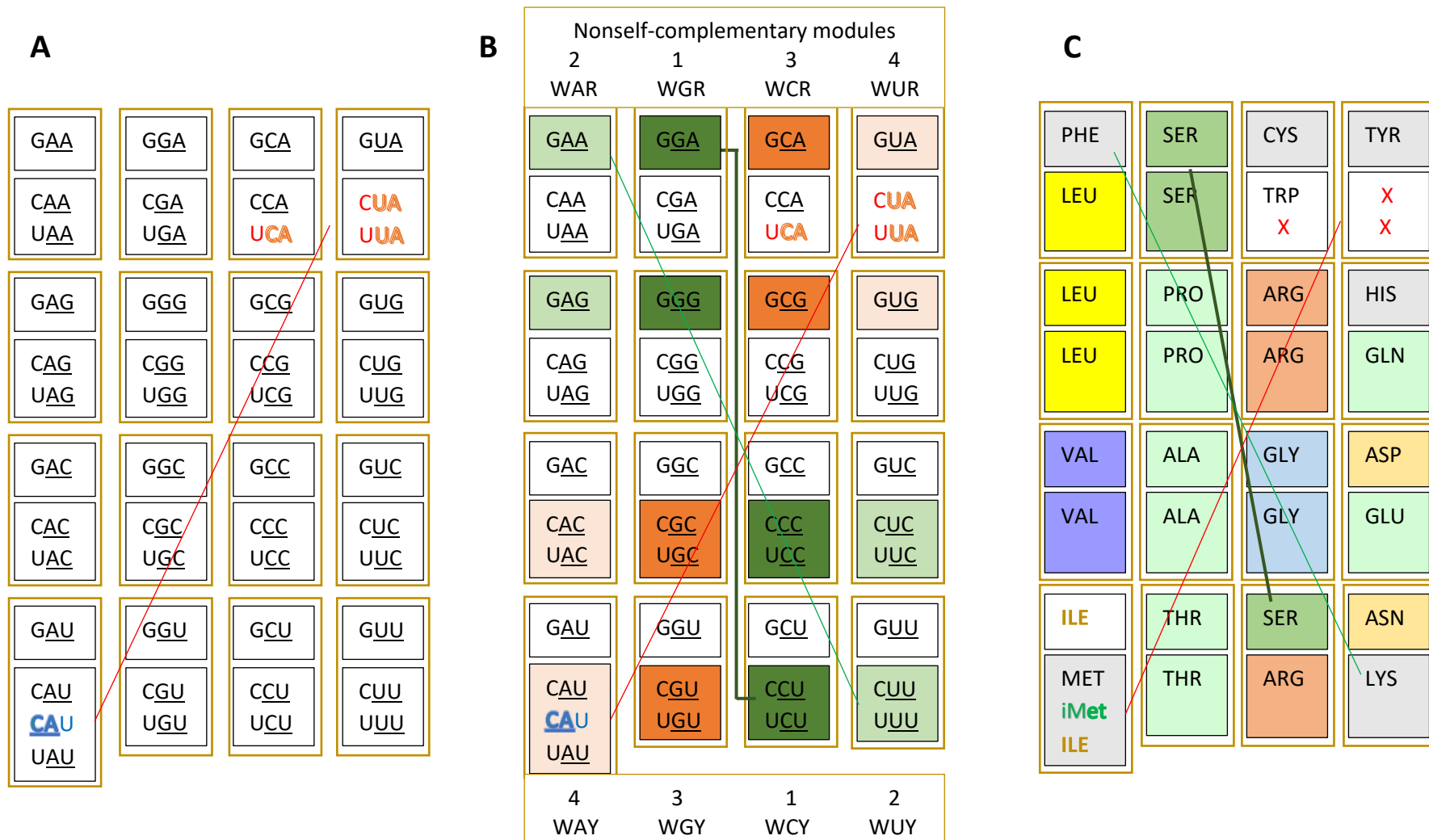
Figure 1 Genetic anticode triplets and meanings in the matrix format

Figure 1. Genetic anticode triplets and meanings in the matrix format. **(A)** The matrices differ from the traditional in exchanging the positions of the two right columns and the two bottom rows, therewith conducing to easy visualization of the symmetries produced by the triplet pairs. The principal dinucleotides, underlined, define the 16 boxes. They are constituted by the columns or the central bases, and the rows or the 3´ or last bases of the triplets; the first base, 5´, is the variable or wobble position, W (G, C, U). The direction is given by the biosynthesis of the polymers: start with the leftmost monomer and keep adding new ones at the right side. Codons are 64 but anticodons are 46: three per box plus the initiator less the three terminators. One of the symmetries is highlighted in colors and by the diagonal line uniting the initiator and the terminator codes. This relationship builds a punctuation system that is examined in details in [16]. Here it is only pointed out the complex relationships: both the initiator and the elongator Met triplets have the same composition but different functionalities of the principal dinucleotides, CAU for elongation and **CA**U for initiation; the termination codons correspond to the anticodons (red) that were eliminated, with a Y**YA** constitution, except the C**CA** that is maintained for Trp. **(B)** The four modules of nonself-complementary triplet pairs, which are the primary encoding modules, are numbered and highlighted in colors; the networks of pairs are drawn in Figure 6. Two other diagonals are indicated, which will be commented upon later: the complementary principal dinucleotides $\frac{\rightarrow NGA}{\leftarrow UCN}$ for the same amino acid Ser; the two atypical synthetases that occupy complementary triplets $\frac{GAA\ Phe}{UUY\ Lys}$. **(C)** The matrix of the meanings. There are three hexacodonic attributions: LeuRS NAG plus YAA runs along one same column inside the homogeneous principal dinucleotide sector; ArgRS NCG plus YCU runs along one same column and traverses from the homogeneous to the mixed principal dinucleotide sector; SerRS utilizes complementary principal dinucleotides in module 1. One half of the boxes is single-meaning, the other half is multi-meaning, and these are all paired symmetrically: single meanings in the core boxes (NGG Pro : NCC Gly, NCG Arg : NGC Ala); multi-meaning in the boxes at the tips (NAA Phe, Leu : NUU Asn, Lys; NAU Ile, Met, iMet : NUA Tyr, X); single-meaning pairing with multi-meaning, respectively: NGA Ser : NCU Ser, Arg; NAG Leu : NUC Asp, Glu; NAC Val : NUG His, Gln; NGU Thr : NCA Cys, Trp, X.

The anticodons find their complements in the codons of mRNA sequences at translation. The translation process would be more strictly called transliteration in view of its punctual or 'letter-by-letter' (triplet-to-amino acid, 'digital') nature, without any hints at interpretation of messages. This set of 'letter' codes is at the origins of cellular organization, from which precision in the structures and functions can be obtained through the construction of sequences – genes and proteins. These acquire functional conformations that build the metabolic flow, which sustains the organisms, through the adequate body structures (Figure 2).
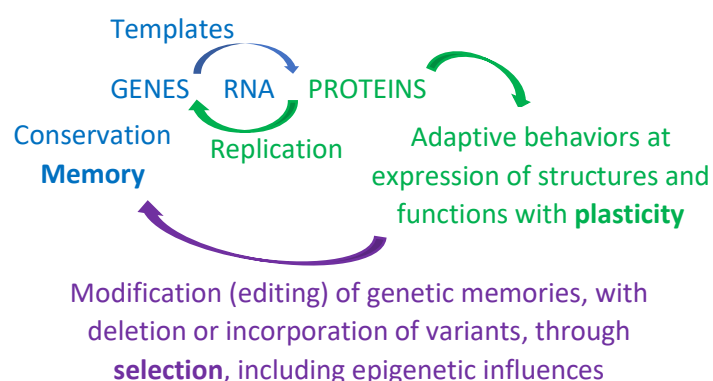

Figure 2 General structure of the nucleoprotein system of cells

Figure 2. General structure of the nucleoprotein system of cells. Aspects highlighted are the conservative (blue) and the evolutionary (green), which are contrasting functions in mutuality and interdependence. Conservation of memories is necessary for identity and regeneration of the proteins. Proteins execute replication of the memories and the vast majority of other functions, constructing the body and relating it to the environments, through their own activities and through regulation of the expression of the memories. These activities are mediated by diverse RNA types together with the proteins, which are the epigenetic mechanisms. Some of the epigenetic signals produce hotspots or facilitate genetic activities that are related to the generation of variability. Selection acts upon the variant sets – the 'editing function' (violet) – evaluating the system's fitness in relation to the environments, which results in changes in the populations of individuals with their genomes.

Model Understanding the structure and the process that formed the code is still non-consensual, in spite of the half century span from the deciphering of the meaning of the triplets. The Self-Referential Model [17] indicates that the formation of the code was based on 'Protein Synthesis Directed by Dimers of proto-tRNAs'. The dimers are considered mimics of the ribosomes – structures that hold two tRNAs together and facilitate the transferase reaction – but may be also considered among other instances of Non-Ribosomal Protein Synthesis [8, 9, 26, 28]. While the couple of tRNAs is laterally associated in ribosomes, the dimers associate proto-tRNAs through pairing of the anticodon loops (Figure 3). Information on the original data, going back to 1996 [12], and further references are compiled and reviewed in [14, 16].

Figure 3. Ribosomal, mRNA-directed protein synthesis and proto-ribosomal Dimer-Directed Protein Synthesis. The first (left) is uni-directional and the transferase reaction (green arrow) occurs between a couple of laterally-associated tRNAs. The peptide in the peptidyl-tRNA (n) is added to the entering aminoacyl-tRNA; the peptidyl becomes n+1 and this peptidyl-tRNA is translocated (blue arrow) to the peptidyl site. The second (right) is bi-directional and the tRNAs are paired through the anticodon loops. The paired proto-tRNAs have an undecided identity, the anticodons being at the same time codons one for the other. The product protein may be semi-repetitive in constitution, combining some preferential amino acids in consequence of some chemical affinities, but also openness to external availability of amino acids. The dashed curved line elongates the chain of amino acids in the peptide and wraps around the producer dimer in the process of stabilization and coating/protection. Variations in the product peptides that bind to the producers may lead to continuation of the 'superposed or coherent' qualities (in analogy with quantum systems properties) or to differentiation ('decoherence') into singular identities after preferential binding of the products to one or the other proto-tRNAs. It is possible that the Ser codes are reminiscent (chemical ´fossils´) of the coherent state, conserving the complementary principal dinucleotides (NGA:NCU) and an auto-aggregated synthetase.
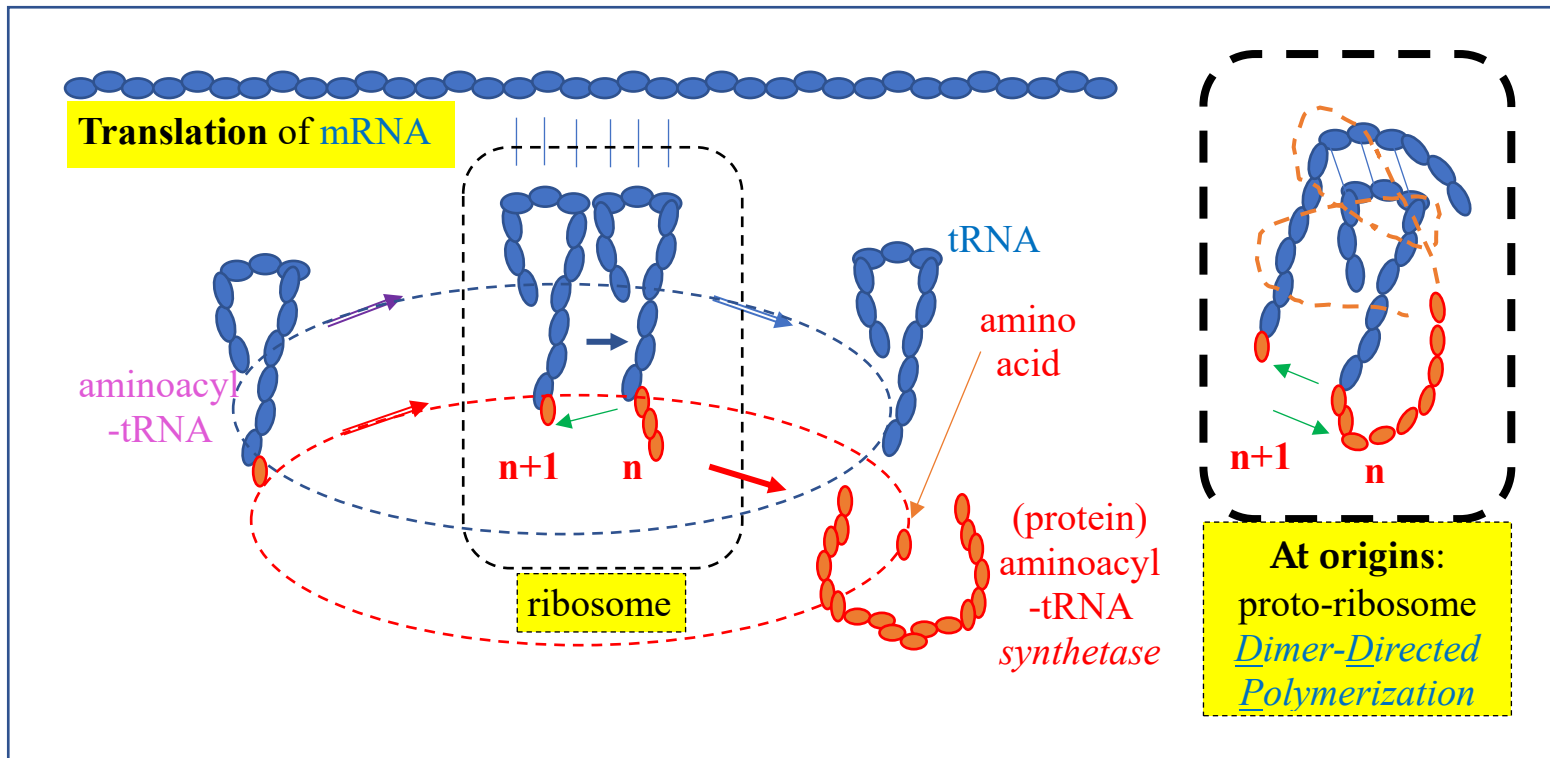
Figure 3 Ribosomal, mRNA-directed protein synthesis and proto-ribosomal Dimer-Directed Protein Synthesis

We concentrate here on mechanisms involved with the generation of complexity in biosystems, that are centered on the evolutionary construction and diversification of sequences of the biopolymers – proteins and nucleic acids. These combine in the construction of cellular bodies through the criterion of functionality of the metabolic flow that is directed to serve the regeneration of the system of biopolymers. The fundamental structure is of networks that are rich in self-referential loops and provide for partial sustainment of the system – the environmental dependency is irrevocable. The small networks studied here involve, in the order of increasing complexity and flexibility, dimerization of tRNAs, degeneracy of the aminoacyl-tRNA synthetases and association of these in the ribonucleoprotein aggregates Multi-Synthetase Complexes (MSC).

RNA world The dimer-directed encoding process overcomes and by-passes the problems introduced by the RNA World hypothesis, of having RNA-only protocells that, besides not having a solid foundation on abiotic availability of nucleotides, requires long chains of these, which are known to be fragile and unstable. Instead of starting with long RNA molecules to be translated, only thereafter acquiring a meaning other than the self-directed (RNA makes RNA that makes RNA- - - - -), the dimers may be composed of oligomers; sizes below or around 20mer are compatible with abiotic synthesis. The constitution of the oligomers is left open in view of the many possibilities that are offered by prebiotic chemistry [e. g. 7]; they might have been polymerized on crystal or clay surfaces. It is only required that they should function similarly to the tRNAs, being able to carry attached molecules, including amino acids, to dimerize (Figure 4) and to transfer the load from one to the other.
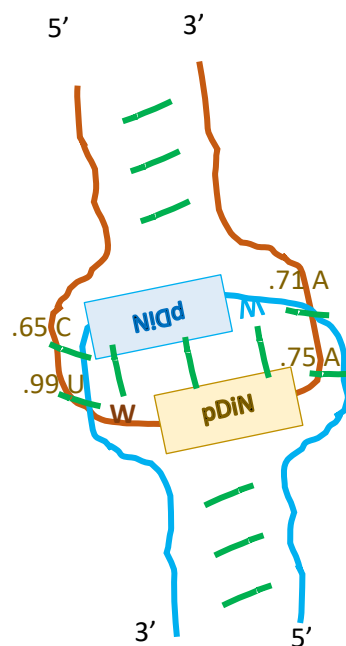

Figure 4 Sketch of a general structure for a pair of anticodon loops

Figure 4. Sketch of a general structure for a pair of anticodon loops. The central base pair joining the principal dinucleotides is of the standard G:C or A:U kind. The 3' base of a principal dinucleotide pairs with one of the choices offered by the wobble (W) position in the other strand, according to the generic R:Y rule. In pairs of present day tRNAs the two bases lateral to the anticodon, in both sides, are indicated to extend the pairing, since they are frequently purines in one side and pyrimidines in the other side [37], namely (in the brown strand) base 32 is 65% C : base 37 is 75% A, base 33 is 99% U : base 38 is 71% A. The structure in [27] is not entirely adequate for comparison due to having been obtained from pairs that went through harsh purification procedures and show only the triplet pair. The thermodynamic data [10] indicated stability strength compatible with about seven base pairs. Possible involvement of curvatures in the anticodon loops, such as the U-loop involving U33 and the central purine in the same anticodon loop [24], are not drawn.

Precursors of anticodon triplets would have been oligomeric complementary sites, like others that participate in intermolecular associations. Present day triplet structures would have been derived from a 'compression' process imposed upon the mRNA chain and the tRNA L-shape inside the ribosome. In order to accommodate the interacting segment of the mRNA plus the two tRNAs inside the organelle, the anticodon loop and the mRNA developed torsions and curvatures that should reflect the 1 + 2 functional differentiation of the wobble + principal dinucleotide positions. The ribosomal decoding site structure became physically separated into a 1 + 2 non-contiguous construction. The synthetases conserve the preferential interaction with the principal dinucleotide of anticodons at most of the single-meaning boxes. While it is not possible to have the relevant prebiotic samples to work with, biochemical tests may utilize the known tRNAs or some mini-versions of them, as proxies, at the same time attributing known functions to the codes and adding biological qualities to the model components.

Network origins A sketch of the organization of the code based on pairs of anticodons is presented [detailed in 16], on which basis the formation of biological networks can be visualized. These go from the more rigid and regular kinds of interactions between the triplets of bases in RNAs, which are reminiscent of nearly crystalline structures (Figure 5), to the more plastic and pleomorphic that are formed by protein interactions, sometimes described through the similarity with sticky gels.

Triplets and dimers The pairs of anticodons have a structure $\frac{\to 5'\ wobble-central-3'}{\leftarrow 3'-central-wobble\ 5'}$ that generates small networks due to the choices allowed by the composition of bases in the wobble position of the triplets. The central base pair is of the strict G:C and A:U kind. The lateral pairs are dictated by the 3' base of the principal dinucleotide, choosing the complement among the possibilities offered by the wobble position and accepting the generic R:Y pairs. This is necessary in view of the elimination of A at the 5' position. A basis for the Self-Referential Model is the full credit given to this 1 + 2 structure of the anticode triplets, which is only now being introduced into studies of codons [35].

Figure 5. The common base pairs in RNA. The G:C pair is strong, with three hydrogen bonds. All others are weak, with two hydrogen bonds. The G:C and A:U are standard, conserving the precise angles and distances, which are important for the double-helical strict regularity. The G:U pair is somewhat weaker than the A:U due to some distortion in distances and angles. The A:C pair, not shown, is topologically similar to the G:U but even weaker and rarer. The general rule is purine:pyrimidine, R:Y, (G or A):(C or U), allowing for shuffling of kinds along the sequences. Picture obtained from Google Images – EteRNA WiKi, August 2017.
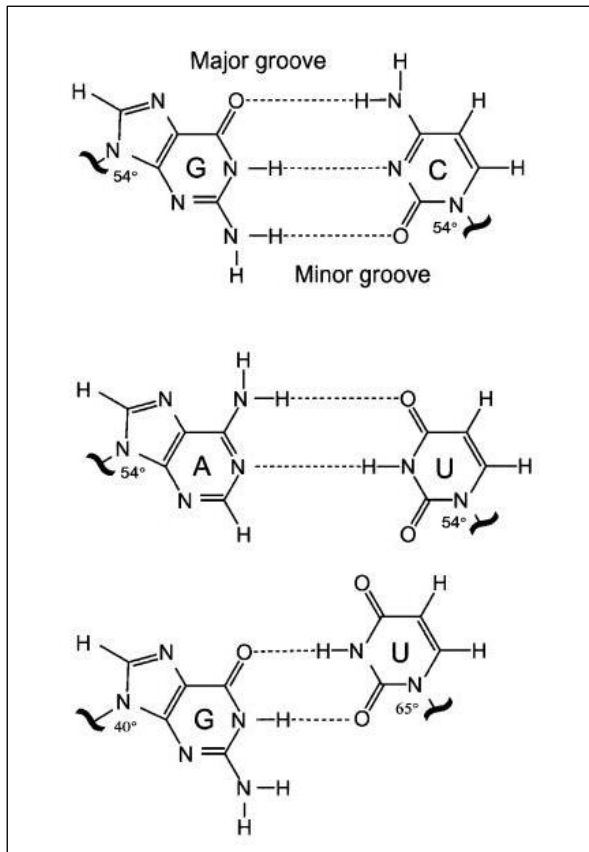
Figure 5 The common base pairs in RNA

Figure 6. Networks of anticodon pairs. The pairing rules are: central standard G:C and A:U. Laterals generic R:Y dictated by the 3' base, choosing the base to pair among the variety offered by the wobble position $\frac{\rightarrow 5'WNN3'}{\leftarrow 3'NNW5'}$. The anticode matrix is separated in the nonself-complementary (left) and self-complementary (right) types of triplets. The first kind find complements among themselves diagonally and form four modules (1-4) with identical topology, asymmetric due to combining two triplets from the upper two rows (5'G-central-3'R) and four from the lower two rows (5'Y-central-3'Y), totaling 8 pairs per module. The four modules compose two sectors, with homogeneous (NRR and NYY) or mixed principal dinucleotides (NRY and NYR). The self-complementary triplets from the upper two rows (5'Y-central-3'R) are untouched by the 5'A elimination and conserve the 4 x 4 pairs symmetrical topology. The topology is identical for the central G:C and central A:U modules, that find pairs combining the two sectors, horizontally along the upper two rows. In the lower two rows, the self-complementary triplets are all 5'G-central-3'Y, there being only four triplets in each group of central base kind; the two networks have symmetric 2 x 2 pairs, again combining the two sectors. Module numbers indicate the order of encoding: homogeneous principal dinucleotides before the mixed and inside a sector, the central G:C before the central A:U.
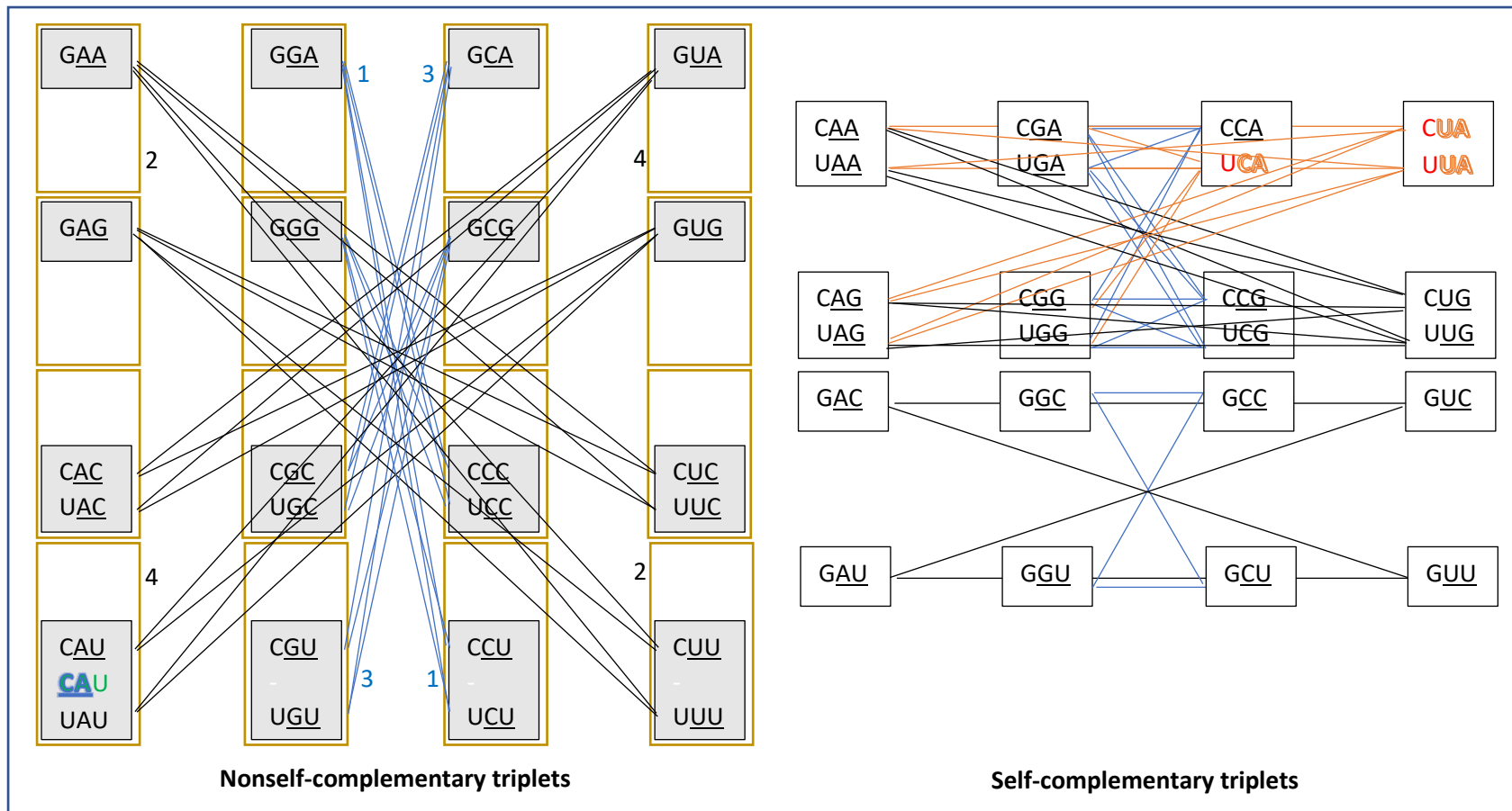
Figure 6 Networks of anticodon pairs

Nonself-complementary triplets and modules The networks are of two types, distinguished by the kinds of triplets that suffered the consequences of the 5'A elimination in different ways (Figure 6). In the matrix of triplets, note the hemi-boxes called nonself-complementary: both lateral bases are of the same kind, both R or both Y. These triplets pair only with others of the same nonself-complementary set. This set of pairs builds two sectors in the matrix that run along the diagonals. In one sector, from the upper left corner to the lower right, the principal dinucleotides are called homogeneous, composed of either two R or two Y bases. One nice consequence of being nonself-complementary is that the triplets display for interactions fully planar surfaces, where each kind of radicals reach the same height for either run of three R or three Y. The repetitiveness also means simplicity in the set of radicals along the triplet, and symmetry from the center to the sides. It is indicated nevertheless that it is not the symmetric character per se that constitutes a qualitative requirement of the encoding process; it is the non-complementary character of the lateral bases that adds the meaningful quality. In other words, it might be possible to have another molecule in the place of one of the bases that would create asymmetry but if it maintained the avoidance of the complementary pairing it could still be accepted by the encoding system.

The sector that runs from the lower left to the upper right in called mixed, more complex due to the composition of the principal dinucleotide having one R and one Y, therefore with a rugged surface while maintaining the nonself-complementary property. The networks of dimers formed by these triplets become, after the 5'A elimination, asymmetric due to having two 5'G triplets pairing with four 5'Y triplets, and this structure is fully repetitive the whole matrix traverse. Four modules are generated, introducing a quality of importance for the encoding process: what has been learned and developed in the first module may be applied with expediency to the others, through duplications. The process of evolution that is facilitated by duplication followed by diversification is common and may be applied to various kinds of structures [4, 5, 21], possibly made easier in the case of modular repeats such as indicated by the self-referential model.

Encoding letters The process envisaged for the encoding is the generation of a circularly structured association system that keeps practicing cycles of the transferase reaction. It is composed by **(1)** producers, which are the dimers of oligomers, and the **(2)** products – proteins (oligomers to polymers), among other possibilities, depending on the kinds of substrates utilized for the synthesis. **(3)** When the products acquire the adequate composition for not being lost to the environment and for binding back to their producers, therewith coating and protecting them from degradation, stabilization of the ensemble is reached. **(4)** Stability means habituation, where the system develops a longer duration and, provided that its original function is not impaired, it will keep producing more of what it got used to do, according to the natural availability of reactants. The end result of the cycles of synthesis and of the association is **(5)** a stable producer-product correspondence, which is a letter code. The evolved producer acquires the property of memory for the product; the cycle is identical in structure to the epigenetic processes [32]. Composition of the products of the dimer-directed syntheses might be originally biased with respect to the constitution of the dimers and of the available monomers that they carry, due to chemical affinities and abundances, therefore not homogeneous but also not entirely dictated by external availability. The cycling process would contribute to the enforcement of some aspects of the interactants and lead to mutual adjustments.

Encoding practical The nonself-complementary modules fit convincingly the requirements for encoding structures. Their simplified and asymmetric character facilitates the process. A high stability dimer is encoded first (1). This is composed by triplets belonging to the two middle rows of the matrix. Data on the estimated stability of triplets are in [13]. (2) All other dimers the first triplets would also be involved with become of low concentration, while (3) the second set of dimers remain at high concentrations and develop the second encodings. This set is composed by the triplets in the upper and lower rows. At the encoding, the synthetase joins the degeneracies in the wobble position together in the same principal dinucleotide (1', 3') (Figure 7).
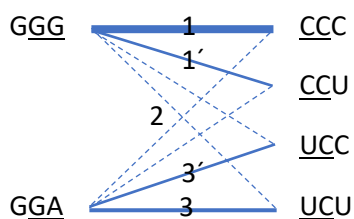
Figure 7 Encoding the two pairs of boxes in the asymmetric nonself-complementary triplet modules through cycles of Dimer-Directed Protein Synthesis

Figure 7. Encoding the two pairs of boxes in the asymmetric nonself-complementary triplet modules through cycles of Dimer-Directed Protein Synthesis. Among the products of each dimer there are the aminoacyl-tRNA synthetases that materialize the encoding. Module 1 is an example. The high ΔG pair (1) is stable enough to facilitate protein synthesis from which a stabilized precursor-product correspondence is generated. This pair is composed by triplets belonging to the two middle rows of the matrix. All other pairs the triplets of pair 1 would be involved with become of low concentration (2), leaving the other pair (3) at high concentration, from which another correspondence is fixed. This pair is composed by the triplets in the upper and lower rows. Synthetase recognition of tRNA evolves from – the initial state (a) – a collection of distributed sites along the protein and the tRNA sequences, which may or may not involve the anticodon, to (b) involve specificity towards the principal dinucleotide of the anticodon. At these stages, high degeneracy is the rule so that the other triplets in the module (1', 3') follow their cognate principal dinucleotides.

Across the modules, the rule is to encode the homogeneous principal dinucleotide sector before the mixed, following the synthetase interaction sites from simple (repetitive, planar) to complex (mixed, rugged) and, inside the sectors, encode first the central G:C before the central A:U, following the triplet dimer stability. This was found entirely in accordance with the starting metabolic pathway, which is the Glycine-Serine Cycle, and also with the late installation of the specific punctuation system.

Diversity and combinatorics The availability of a letter code is a main and essential attribute of living beings, which allowed the construction of specific structures and functions through the enchaining of the letters into linear sequences of polymers; these fold in the 3D space into precise functional conformations. An apparently endless array of sequences is possible to be generated, giving support to the enormous diversity of the biosphere, all possibly based on combinatorics, among other processes, with some intriguing similarity to human languages, at least metaphorically or as an appealing analogy. The evolutionary panorama of large diversity of living beings that form highly complex bodies and ecosystems would be adequately described by the already well settled assertion of an endless or open process – open-ended evolution, which is the biological counterpart to the infinites of logics and mathematics.

Encoding sequences The encoded letters would be able to generate sequences or chains of codes through ligation and other choices among the variety of molecular evolution processes, therewith forming genetic sequences – chains of the tRNAs or of segments of their sequences. These would have the quality of being meaningful from the beginning, therewith reducing the problem of having nonsense or stop segments inside the coding sequences. The first kinds of selection criteria (Figure 2) would be plain and simple stability of the products, thereafter their ability to bind the producers, their protection from degradation, and the capacity of not harming their activities. In this way, the small precursor-product system may keep producing more of itself, which is akin to reproduction. The process has similarities to other models called auto-catalytic sets or systems [20].

Early sequences It is envisaged that at some early stage in evolution of the system its composition would be biased according to the affinities presented by the monomers that were obtained from external sources. If, e. g., they would be similar to the products of Miller type experiments, they would contain mostly organic acids, including keto-acids, which may be aminated to become less reactive and more stable in the form of amino acids, among other compounds. Organic materials from meteorites indicate the presence of amino acids in the parts-per-million range, of nucleobases in the parts-per-billion range, which could give estimates on what could have been the composition of some early oligomers. At some later stages, this could have given rise to the variety of Non-Ribosomal Protein Synthesis, including the proto-tRNA Dimer-Directed Protein Synthesis and the formation of RNP systems, such as the synthetase-tRNA and the ribosomal. It is appealing, e. g., the known high participation of Glycine in the RNA-binding sites of today's proteins, whose affinity might have been relevant for selection in favor of RNA along the process.

Evolution It is adequate to highlight three key-words in the context of early biological evolution. **(1)** Stability is the crux in a process that should ideally acquire the **(2)** self-stimulation capacity (positive feedback, self-feeding). Stability should be partial so that the structures do not 'freeze' in one form but keep open to change and evolution, which is the attribute of **(3)** plasticity. In later steps the process, we might identify two characters that summarize and identify biosystems, in interdependent circularity and in spite of the danger of reductionism but with the quality of simplicity: **(a)** stability and conservativeness, which are main attributes of the genes, nucleic acids and replication; **(b)** evolutionary potential and realization, openness to change and adaptation, which might be pinpointed to protein plasticity [3, 23, 29].

The process of formation of the code is pre-Darwinian, in the Self-Organization realm. It may be estimated that it would have taken hundreds of million years, in the interval from the origin of the Earth to the paleontological dating of cellular microfossils, this at ≤3.5 Gya. This is the period of maturation of the LUCA lineages and their confluence into the populations that share the nearly universal code. We stress, inside the Self-Organization paradigm, the Self-Referential mechanism, which is intended to be at the same time wider and softer than the auto-catalytic; it would be closer to the more systemic ´auto-catalytic sets´ [20]. The specificity in the case of the encoding indicates the supposed dominance of characters of the products, which are the proteins made of simple monomers, in shaping the outcomes in the system under construction. It is the quality in the product (such as stability and the binding ability) that chooses among variants of the producers which will be adequate for the ensemble to fit together in a system. In the case of encoding, it is suggested that peptides that were stable in themselves and adequate for binding to the proto-tRNAs, directed the development of the RNA structure, that is considered of biotic origin.

A result of the chronology of amino acid fixation in the code that highlights the precedence of the plastic character over the internally organized protein structures is the composition of initial set of amino acids, which is more adequate to build Intrinsically Disordered Regions of proteins. The evolutionary path indicated is from disorder to order, meaning that disordered regions are original and open to develop order and informational patterns at the interaction events, in mutuality and in accordance with the kinds of interactants [15].

Metabolism Besides delineating modes of experimentally testing the Dimer-Directed Protein Synthesis activity, the significance of the self-referential encoding process for cellular systems was immediately evaluated through overlaying the structure of the sets of dimers upon various sets of properties of amino acid residues in proteins and of properties of proteins. The chronology of amino acid encoding that was generated found support in a biosynthesis pathway that makes sense as candidate among the first in the metabolic network, the Glycine-Serine Cycle: (1) It is the simplest among the central metabolic pathways – it starts with C1 and its most complex components reach only the C4 level. (2) It is the only one containing amino acids, while the others contain precursors to amino acids. (3) It contains the two direct precursors of gluconeogenesis. The chronology of amino acid incorporation into the code is presented in Table 1.

Table 1 Overview of the chronology of encodings

Homogeneous sector

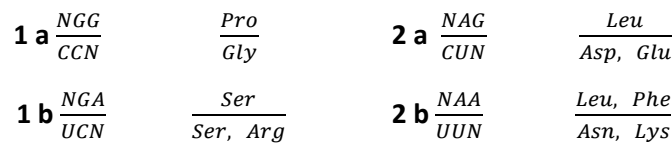|	
1a GGG → CCC	**2a** GAG → CUC
↘ CCU	↘ CUU
↗ UCC	↗ UUC
1b GGA → UCU	**2b** GAA → UUU

First encodings

1 a $\dfrac{GGG}{CCY} \rightarrow \dfrac{NGG}{CCN} \rightarrow \dfrac{Gly}{Gly}$ **2 a** $\dfrac{GAG}{CUY} \rightarrow \dfrac{NAG}{CUN} \rightarrow \dfrac{Leu}{Asp}$

1 b $\dfrac{GGA}{UCY} \rightarrow \dfrac{NGA}{UCN} \rightarrow \dfrac{Ser}{Ser}$ **2 b** $\dfrac{GAA}{UUY} \rightarrow \dfrac{NAA}{UUN} \rightarrow \dfrac{Leu}{Asn}$

Present encodings

1 a $\dfrac{NGG}{CCN} \qquad \dfrac{Pro}{Gly}$ **2 a** $\dfrac{NAG}{CUN} \qquad \dfrac{Leu}{Asp, \; Glu}$

1 b $\dfrac{NGA}{UCN} \qquad \dfrac{Ser}{Ser, \; Arg}$ **2 b** $\dfrac{NAA}{UUN} \qquad \dfrac{Leu, \; Phe}{Asn, \; Lys}$

Mixed sector

3a CGC ← GCG	**4a** CAC ← GUG
UGC ↙	UAC ↙
CGU ↖	CAU ↖
3b UGU ← ACG	**4b** UAU ← AUG

3 a $\dfrac{GCN}{NGC} \qquad \dfrac{Arg}{Ala}$ **4 a** $\dfrac{NUG}{CAN} \qquad \dfrac{His, \; Gln}{Val}$

3 b $\dfrac{NCA}{UGN} \qquad \dfrac{Cys, \; Trp, \; X}{Thr}$ **4 b** $\dfrac{NUA}{UAN} \qquad \dfrac{Tyr, \; X}{Ile, \; Met, \; iMet}$
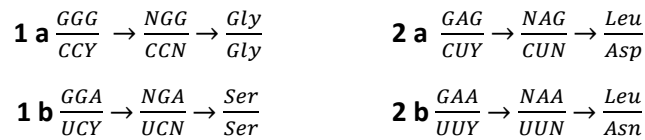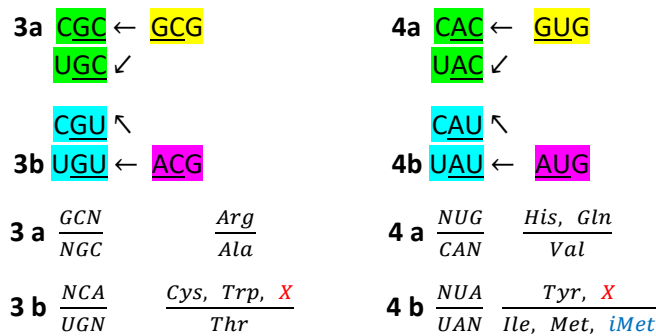

Table 1. Overview of the chronology of encodings. Annotations can be detailed in the homogeneous principal dinucleotide sector, where there are constraints dictated by the constitution of the Gly-Ser Cycle of assimilatory metabolism. This supports the five first encodings, all in single-meaning boxes; Leu is the only synthetase class I. Additional encodings in this sector (total ten) are proposed to have been added at maturation of the central metabolic pathways, starting with gluconeogenesis, then glycolysis and the pentose-phosphate shunt (necessary for the biosynthesis of Phe) plus the Krebs Cycle. The latter includes the precursor to the Glu family of amino acids (Pro, Arg), while Lys may be obtained from both Asp or Glu sources. The richer amino acid repertoire allows synthetase specificity to develop, including the generation of multi-meaning boxes. Encodings in the mixed sector are not metabolically constrained. The sector was initiated by the ArgRS expansion from the YCU hemi-box to the NCG box. Note the complete substitution of Gly by Pro in the NGG box. Other concessions are partial, from the first amino acid (in the left) to the new attributions, separated by comma(s).

The flow It is not adequate to ask which came first, the code or the metabolic pathways. It is not possible to have one without the other. It is considered that generation of triplets would be simple, from replication, but only after the monomers can be available from biosynthesis, and amino acids are precursors to the nucleobases. Amino acid biosynthesis is also difficult, besides having to adjust to – therefore being directed

by – physiologic necessities, so being the main constraint on the process. On the easier side, it is considered that the most relevant amino acids (Gly and Ser) come directly from C1 sources. The solution has to reside in coevolutionary processes among various members of the system, which should include these two components plus the upstream substrates or sources for metabolism and the downstream destinations for its products. This metabolic ensemble composes a flow system where the flow itself is the organizer since it is immersed in the pre-existent general universal flow of masses and energy. In nested circular structures, (a) the flow that is made by the metabolic system is the measure through which the quality of the system is evaluated; (b) the flow produced is checked in relation to and in accordance with the environmental flow. Metabolism, which is the biological dynamics, only adds a new segment with its own preferential sources – C1 organics, and sinks – starting with the uptake and synthesis of amino acids, thereafter with their sequestration in proteins.

Superposition and the pair of Serine codes A problem with the dimer-directed encoding, as a model for the proto-ribosome, would arise from the equal value of the oligomers that dimerize complementarily, while the rule in the set of correspondences for translation is the individual encoding of distinct tRNAs. This indicates a biochemical situation, in the dimer, that is analogous to the phenomenon of superposition of states in quantum objects [31, 34, 39, 40]. It is said that the superposition corresponds to the undecidedness between the states or the coherence of one with the other. The quantum objects will adopt the classical behavior, or give rise and transform into them, after a process of decoherence that is triggered by interactions with some kinds of perturbations that work as if separating the components of the quantum object into classical singular states or classical objects.

There are choices to be investigated for identification of the interacting perturbation in the case of the dimers. They could be the product peptides that bind preferentially to one of the proto-tRNAs of the dimers, or the possible different products from one dimer that could bind differently to the two members and, among other possibilities, they could be different proto-RNAs that would interfere with the pairing of the dimers at the proto-anticodon sites. The latter could be assuming the role of present day mRNAs. While in the dimer the anticodons are at the same time codons for each other – states superposed; the singular anticodon identity is defined when an external RNA substitutes the codon function of the other member of the dimer.

The self-referential model finds in the case of the Serine codes, to my knowledge an enigma that has never before received any minimally suggestive interpretation, a remnant of the ancient and original situation of Dimer-Directed Protein Synthesis where both members of the dimer were adopted by one same synthetase enzyme. SerRS is one of the few synthetases that maintain the original situation of not requiring to interact with the anticodon bases. The anticodon principal dinucleotides are complementary $\frac{WGA}{UCW}$ and both the tRNAs and the enzyme are exceptional. The tRNAs have a very long variable arm and the synthetase is auto-aggregated, with two tRNA binding sites [11], which seems to be a unique situation. The usual interpretation for homodimers is the allosteric where the binding of a substrate to one of the sites activates the other in a synergistic mode.

Meanings in hemi-boxes and the atypical pair $\frac{GAA}{UUY}\frac{PheRS}{LysRS}$ A long known regularity in the code attributions, namely the distribution of meanings according to hemi-boxes (of course, in the cases of multi-meaning boxes), now receives a functional explanation, not just the R versus Y description, that is the partition of the set of codes in a box into the self-complementary and nonself-complementary types. The encoding process starts upon the nonself-complementary high ΔG pair of triplets but the synthetase specificity follows a succession of degrees of precision in discrimination with respect to the participation of the wobble position of anticodons.

The anticodon may not participate in recognition, that is, it does not interact directly with the synthetases (SerRS, LeuRS, AlaRS). When it participates in recognition it may do so only through interactions with the

principal dinucleotide, generating a single-meaning box (the three above plus ArgRS, GlyRS, ThrRS, ProRS, ValRS); all kinds of wobble bases have the same value and meaning, whether generating self- or nonself-complementary triplets. Multi-meaning boxes require that the synthetases interact specifically with the base in the wobble position. A rule that describes the occupation of the wobble bases would be: **(a)** the nonself-complementary is the original or first triplet occupied but it may be 5'G or 5'Y; **(b)** this encoding passes through the single meaning stage where the trace of the triplet of origin is erased (the wobble is any base); **(c)** when a new meaning is to be encoded in a box, the original or initial meaning of the box retains the specificity for the 5'G triplet and **(d)** concedes the 5'Y triplets to the new synthetase; this may or may not coincide with the primary encoder in the box. The reason for step (c) would be the strong and specific pair the 5'G will form at pairing with a 3'C.

The rule is followed by another character, that the second meaning in a multi-meaning box corresponds to class I or punctuation. Such rule of concession of YNN hemi-boxes to the second meanings is followed by six of the eight multi-meaning boxes: WCA Cys/Trp and X, WUA Tyr/X, WUG His/Gln, WUC Asp/Glu, WCU Ser/Arg and WAU Ile/Met, iMet. The other two follow a pair of atypical enzymes, which are in accordance with the Self-Referential Model. The LysRS class II of some organisms is atypical with respect to the rule above, besides the exceptionality of LysRS being of different classes in different organisms. The case of the PheRS / LeuRS split box does not contribute to the accounting above since the model says that LeuRS was originally octacodonic WAR (Table 1); the concession to PheRS was followed by recession of the LeuRS, maintaining the W$\underline{AG}$ + Y$\underline{AA}$ contiguity, while it is the G$\underline{AA}$ PheRS that developed the atypical character of being a class II enzyme acylating in the class I mode.

This explanation is also partly historical and contingent: the amino acids Phe and Lys are the only large of class II enzymes and of extreme hydropathies; they should have been taken up by class I enzymes but these were lacking at the time of fixation of the codes, generating the atypical behaviors. The coincidence of the couple being settled precisely upon a pair of triplets adds confidence to the proposition. The moment of these encodings, at the transition between the sectors, should have been a critical period in the system, possibly of enrichments in the nucleic acid subsystems, as indicated by the rise of sugars, which are necessary for the biosynthesis of Phe, and the rise of the basic amino acids Lys and Arg. This is the last in the homogeneous sector and the first in the mixed sector.

Self-complementary triplets and modules are not good for encoding The triplets containing bases of different kinds in the lateral positions, one purine and one pyrimidine, would not be the choice for the encodings in view of their experimentally observed formation of auto-dimers, especially when they contain the small pyrimidines at the central position; the prime example is the very stable auto-dimer of the tRNA$^{\text{ASP-G}\underline{\text{UC}}}$, in spite of the central mismatch $\frac{GUC}{CUG}$. There follows lower concentration of the bona fide hetero-dimers $\frac{GUC\ Asp}{CAG\ Val}$. This rationale should explain why these self-complementary modules would not be chosen for encoding, especially in the presence of the competing nonself-complementary [38].

The modules formed by the self-complementary triplets are of two kinds (Figure 6). Triplets initiated by a Y base are not affected by the 5'A elimination so that the topology is the original symmetrical 4 x 4. Those initiated by an R base are more reduced than the nonself-complementary and the topology of the network is symmetrical 2 x 2. Each topology is repeated twice, according to the G:C or A:U central pair. It is estimated that a decision process based on symmetrical networks, relying only upon differential abundances of elements and differential thermal stability could eventually happen but would take too long in face of the expediency expected for the nonself-complementary networks.

While in the nonself-complementary modules the pairs unite triplets along diagonals of the matrices, following the axes of the sectors, the triplets in the self-complementary modules belong to boxes in the same rows of the matrix, with horizontal connections. Therewith comes the main attribute of the latter modules: they are integrators, connecting the sectors, but this job is accomplished by the special mode of

evolution of the synthetase specificity for the triplets, that starts with full box degeneracy. Were it not for this stage of the synthetase degeneracy (encoding directed to the principal dinucleotide alone) where the distinction between the self- and nonself-complementary triplets is erased, these two kinds of triplets would belong to disjoint modules from the beginning. We are now proceeding the algebraic treatment of the self- and nonself-complementary submatrices (Figure 6), plus their summed set (Figure 1A), in order to check for formal properties (collaboration with Carlos Henrique Costa Moreira) and for their possibly being at the roots of the thermodynamical distinction that was observed by [38].

Integration of the system via proteins – the code degeneracy The self-complementary modules are in charge of integrating the other modules into the RNP system, that is, they participate together with proteins in the integrative process. Here enters a new instance of self-reference where (a) the encoding utilizing a fraction of the triplets produces proteins, the synthetases. (b) Each synthetase, at the development of specificity directed to the principal dinucleotide, will incorporate into its meaning the other triplets with the same principal dinucleotide, which is the development of a full box degeneracy, each box containing the self- and nonself-complementary triplets joined into a coherent set (Figure 8). Expansion of this process will progressively dissipate the whole space of triplets, all of them being recruited for participation in the code system.

Further consequences of the presence of proteins derive from their ability to bind a large variety of other molecules, since they are the sticky or cohesive components of biosystems, in charge of holding all others together into a whole, usually acquiring the shape of a globule when immersed in a watery solution. In the present case, we follow the formation of a network of proteins, the synthetases, superposed upon the tRNA network. This combined RNP network would be near the roots of cellular organization, where RNAs and proteins functionally meet, and also near the roots of biological complexity. It is convenient and didactic to study this case, in view of its small size and apparent simplicity.

Protein plasticity in complexity One component of the complexity comes from the very wide range of diversity in the modes that protein sites – oligomeric motifs – can accomplish the binding function. The combinatorial possibilities at composing the sites allow for fine tuning so that it is possible to utilize the almost digital properties of single amino acids to reach the analogic properties in the site sequences. Definitions of complexity, especially when referring to biological objects, are problematic in themselves because they would have to take into account the great variety of components and of behaviors in the systems the definition refers to, generally ending with non-satisfactory assertions and lack of consensus.

Another aspect to be considered is that complex objects or entities are systems presenting behaviors that may change along time or may accept some degree of non-constancy in composition, therewith increasing the difficulty of capturing in a definition these ranges of variations. Such considerations may suggest that it should not be expected to reach one consensual proposition but it should be accepted that approaches to the examination of complex systems would inevitably be many and diverse, each of them adequate under its own limits and purposes, and that a composite picture [e. g. 36] would be always under construction and revision.
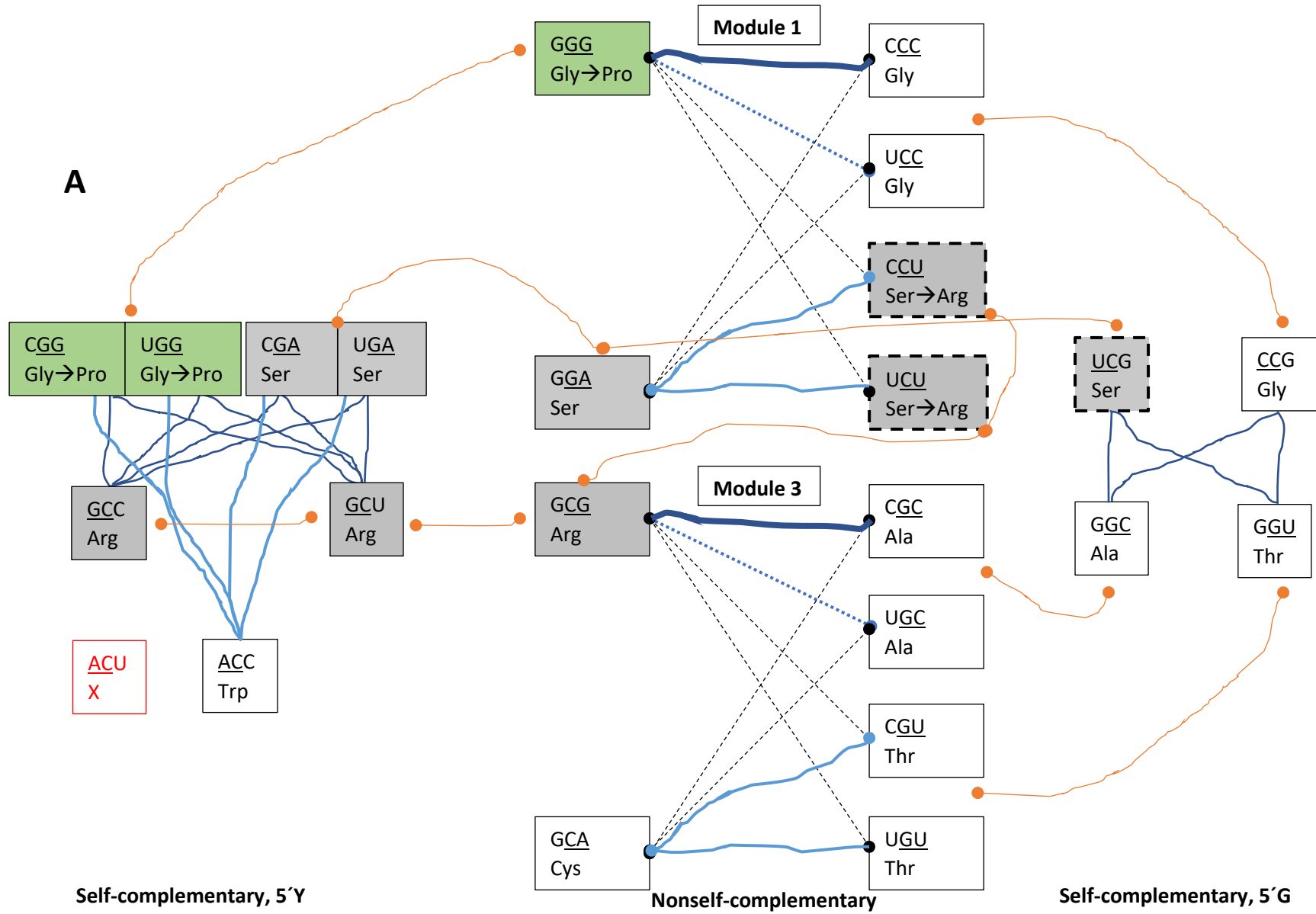
For the biological setting, the proposition I could reach says: (1) Living beings are metabolic flow systems that self-construct on the basis of memories and adapt/evolve on the basis of constitutive plasticity. (2) Life is the ontogenetic and evolutionary process instantiated by living beings [16]. These attempts at definition place complexity, as a quality, in the realm of behaviors, and its material basis would be constitutive to biological entities, here adopting the near synonym plasticity, which is of more general use in biochemistry and not so ´heavy´ with the load of associated connotations as the first term. Plasticity is less intense in biomolecules other than proteins, such as DNA, intermediate in RNA.

Network plasticity – the case of the Multi-Synthetase Complex The most interesting aspects in protein plasticity come from the wide range in the diversity of interactions. Besides the twenty-few encoded amino acids there are the post-translational modifications that enlarge widely the repertoire of elements. This should be reflected strongly in the other material substrate for plasticity that are the networks. These are originated mostly in consequence of the presence of components – nodes – with three or more interactive sites or functions, which is a common feature of biopolymers.

Figure 8. Network graphs of connections between synthetases that are facilitated by the pairwise interactions between their correspondent tRNAs. **(A)** The central G:C and **(B)** the central A:U subnetworks. The structure of the subnetworks of tRNA pairs / dimers is sketched; details are in Table 3. Note that the dimer connections are self-contained and separated into the nonself- and self-complementary kinds, while the integration is developed gradually via properties of the synthetases. In these graphs, the integration comes from the degeneracy properties. Specificity of a synthetase towards tRNAs may be strict single (Σ 9 cases: Phe, Cys, Trp, Tyr, His, Asp, Met, iMet, Asn), or minimally degenerate, to the couple of pyrimidines, which does not add integration beyond the triplet kind of the pyrimidines (Σ 3 cases: Gln, Glu, Lys). The synthetase specificity becomes integrative of different kinds of triplets in the other cases: the tetracodonic or full box Pro, Gly, Ala, Val, Thr (Σ 5), the Ile GAU and UAU, and the three hexacodonics – the simpler case of Leu that is NAG plus YAA, with limited 3´R ambiguity; the peculiar case of Ser that conserves the original complementary principal dinucleotides while concedes YCU to Arg and recedes to GCU; ArgRS is also most peculiar in bridging the two sectors through a wider 3´ ambiguity, going from a homogeneous principal dinucleotide YCU to the mixed NCG. The NUC box is shared by these two hexacodonic synthetases.

Figure 9. Network graph of the connections between synthetases mediated by tRNA pairs and in the Multi-Synthetase Complex. Two sources of the interactions are superposed: those facilitated by the pairing / dimerization of their correspondent tRNAs (straight lines) and those observed experimentally in the isolated Multi-Synthetase Complexes (MSC; curved hand-drawn thick lines). These are colored green for the subcomplexes mediated by protein 43, brown for those mediated by p18 and blue for p38. The straight black lines indicate connections through nonself-complementary triplet pairs, in red the connections through self-complementary pairs. Data on the MSC were taken from the two most complex, that are from Bilaterian animals (see Table 3): the mammalian-type includes four species of mammals, plus *Drosophila melanogaster* and *Artemia salina*, while *Caenorhabditis elegans* is a variant: Mamm 43RQ, 38KD, 18LIEPM; Cele 43RQM, 38KV, LIE. The LIE group of Cele is an example of synthetases that bind to the complex without the intervention of auxiliary proteins. Note the establishment of connection between the two central A:U subnetworks by p38, in different ways in the two MSCs. Protein 43 connects the central G:C with one or both of the central A:U subnetworks in the different kinds of MSC. Protein 18 connects the three subnetworks but it is not present in the MSC of the worm, where connections are only between the two central A:U and obtained by properties of the synthetases.

Figure 8 Network graphs of connections between synthetases that are facilitated by the pairwise interactions between their correspondent tRNAs

A

Module 1

GGG
Gly→Pro

CCC
Gly

UCC
Gly

CCU
Ser→Arg

UCU
Ser→Arg

UCG
Ser

CCG
Gly

CGG
Gly→Pro

UGG
Gly→Pro

CGA
Ser

UGA
Ser

GGA
Ser

GGC
Ala

GGU
Thr

GCC
Arg

GCU
Arg

GCG
Arg

CGC
Ala

UGC
Ala

CGU
Thr

UGU
Thr

GCA
Cys

ACU
X

ACC
Trp

Module 3

Self-complementary, 5´Y

Nonself-complementary

Self-complementary, 5´G

B

Module 2

| GAG Leu | CUC Asp→Glu |
| UUC Asp→Glu |
| CUU Asn→Lys |
| GAA Leu→Phe | UUU Asn→Lys |

CAG Leu | UAG Leu | CAA Leu | UAA Leu

GUC Gln | GUU Gln

AUC X | AUU X

Self-complementary, 5´Y

Module 4

GUG His | CAC Val | UAC Val

GUA Tyr | CAU Ile→Met | UAU Ile

UUG Asn | CUG Asp

GAC Val | GAU Ile

Nonself-complementary
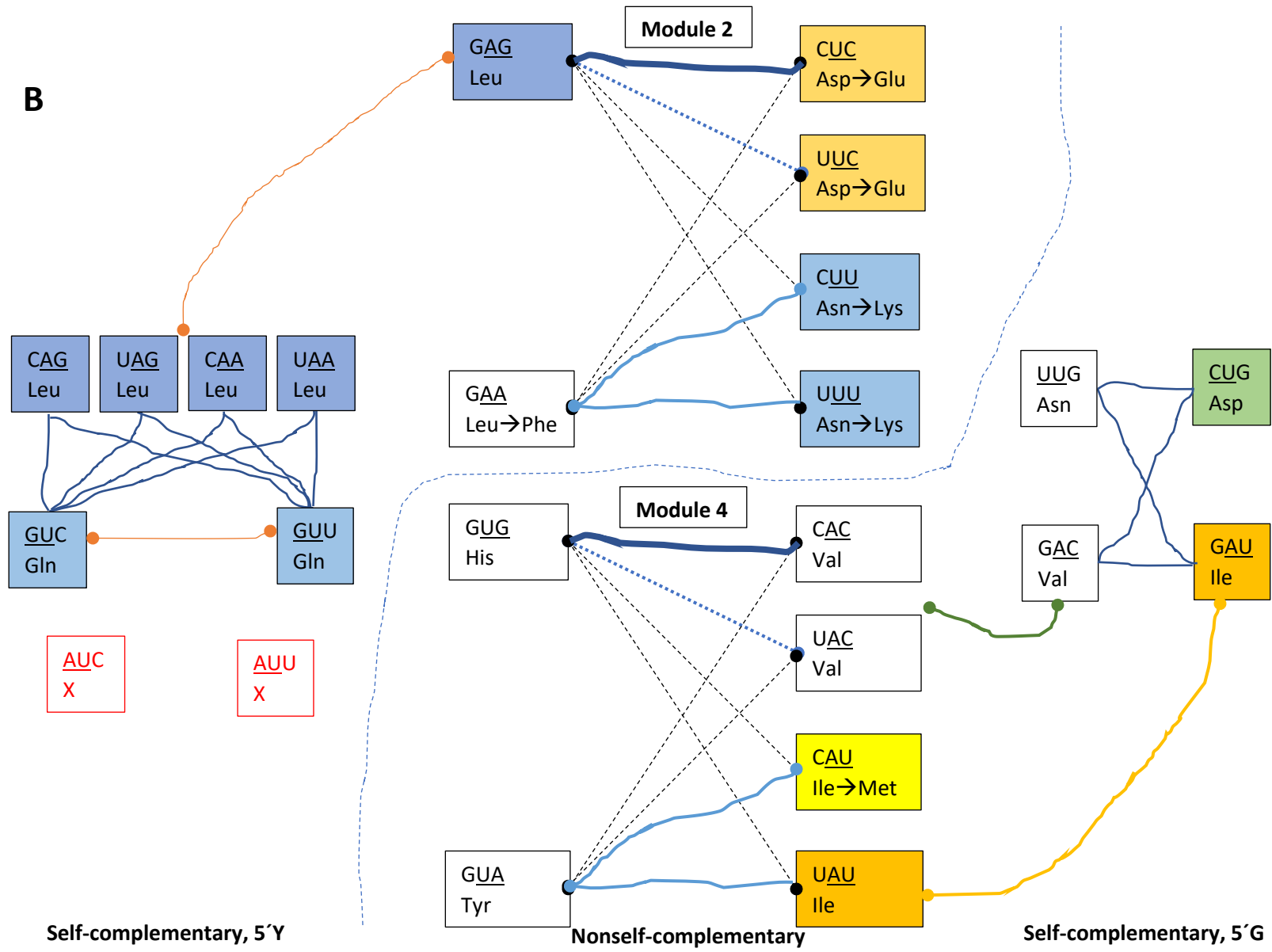
Self-complementary, 5´G

Table 2 Composition of Multi-Synthetase Complexes along the evolutionary scale

Modules		Archaebacteria *Methanothermobacter thermoautotrophicum*		Apicomplexa *Toxoplasma gondii*	Euglenozoa *Trypanosoma brucei*	Fungi *Saccharomyces cerevisiae*	Nematoda *Caenorhabditis elegans*			Insecta *Drosophila melanogaster* Crustacea *Artemia salina* Mammalia Four spp		
1 Gly Ser			Ser									
2 Leu Asp Asn		Leu					Leu			Leu	Asp	
1, 2, +: Phe, Glu Pro Lys Arg		Pro Lys	Arg	Glu	Pro	Glu	Glu	Lys	Arg	Glu Pro	Lys	Arg
3 (Arg) Ala Cys Trp Thr												
4 Tyr His Gln Val Ile Met				Tyr Gln Met	Gln Met	Met	Ile	Val	Gln Met	Ile Met		Gln
Synthetase class	II Σ 8	PK	S		P			K		P	DK	
	I Σ23	L	R	EYQM	QM	EM	LEI	V	RQM	LEIM		RQ
Central base	G, C Σ 7	P	SR		P				R	P		R
	A, U Σ24	LK		EYQM	QM	EM	LEI	KV	QM	LEIM	DK	Q
Auxiliary p43 p38 p18				p43	YbaK p43 p43c	p43		p38	p43	p18	p38	p43

Table 2. Composition of Multi-Synthetase Complexes along the evolutionary scale. An intermediate step of encoding is added between the homogeneous and the mixed sector (modules 1 and 2, +) to indicate the maturation of the central metabolic pathways and full occupation of the homogeneous sector, beyond the Gly-Ser Cycle. Arg belongs to both modules 1 and 3, but was added at the end of the homogeneous sector. It is indicated that the aggregation involves mostly synthetases for the Glu family of amino acids, class I enzymes, and pairs of anticodons with central A:U, which reflects absence of systems involving amino acids from Module 3. Oldest auxiliary protein entering the composition of the complex is p43, youngest p18. [19]
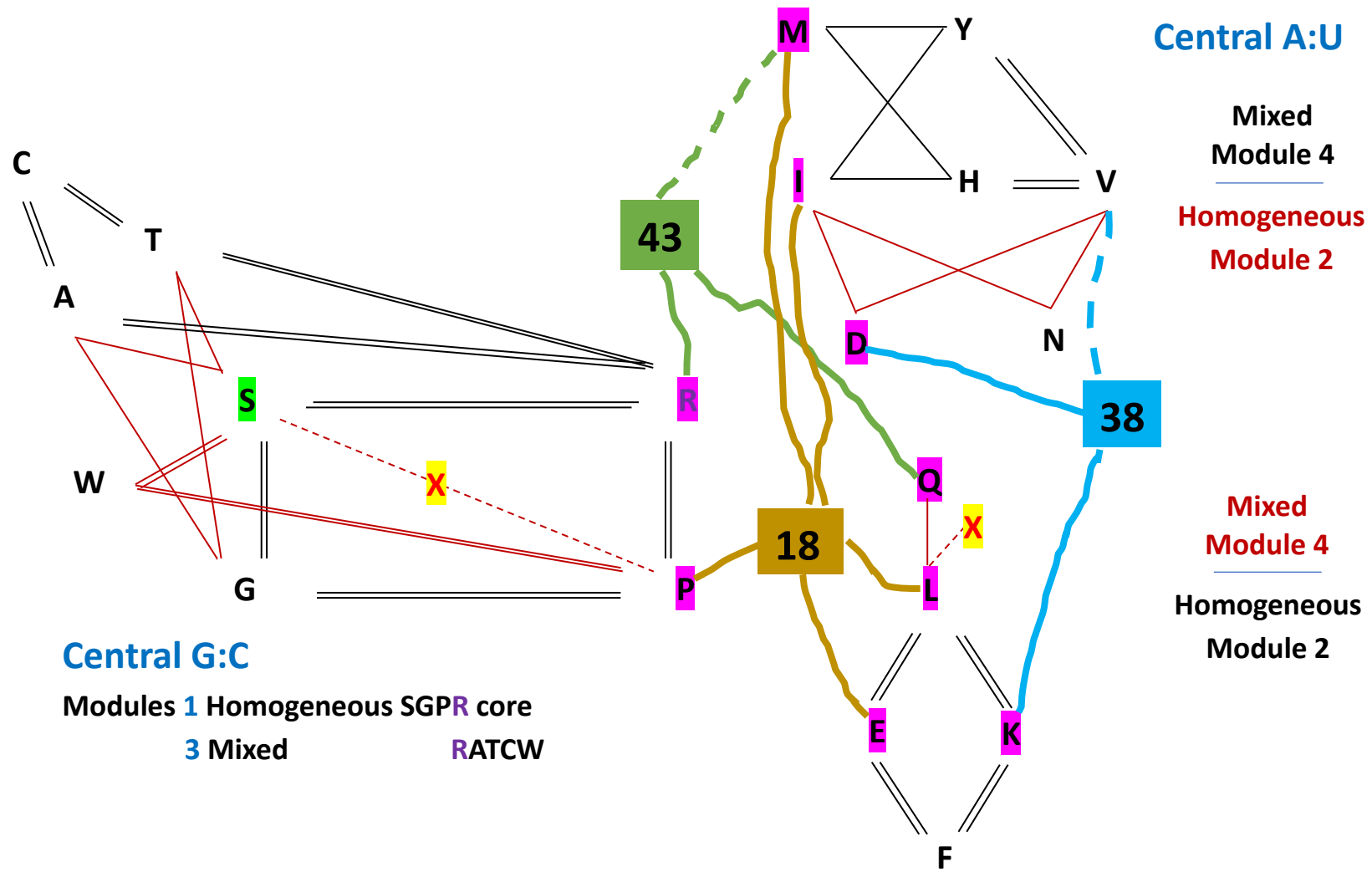
Figure 9 Network graph of the connections between synthetases mediated by tRNA pairs and by protein:protein binding in the Multi-Synthetase Complex

It has been shown above that the tRNA dimers compose a few types of small networks that are lowly connected, one possible strong limitation being the constraint imposed by the strict central base pairing rule. These are partially joined by the synthetase wide degeneracy, when specificity is based on sequence features that are distributed along the tRNA and synthetase molecules without recognition of the anticodon or when it is directed to the principal dinucleotide of the anticodon. These are the cases of the hexa- and tetracodonic attributions (Figure 8).

Dimers place different synthetases in contact and propitiate integration through binding Beyond these integrative advancements based on synthetase degeneracy, a new level of integration comes with the possibility offered by the dimers, when they associate tRNAs belonging to different synthetases. The proteins placed in close contact may develop binding sites through evolution of adjustments in these sites, in case the association, which was initially driven by the tRNAs, proves beneficial to the system [2, 6]. This process may be placed in the context of internalization or endogenization of the benefits of an external influence into the genetic sequences: selection in favor of variants, in the interacting sites, whose effects will mimic the work of the external factor and may now do without it.

Table 3 Connections between synthetases facilitated by the pairing of their correspondent tRNAs and mediated by ribonucleoprotein aggregation

A	In nonself-complementary modules	Σ pairs	In self-complementary modules	Σ pairs	Σ Σ	Synthetases, X
Central G:C, NSC Modules 1 (homogeneous pDiN) and 3 (mixed pDiN)						
Ser	Arg 2, Gly 2	4	Arg 4, Trp 2, Ala 1, Thr 1	8	12	5, X^{UCA}
Arg	Ser 2, Pro 2, Ala 2, Thr 2	8	Ser 4, Pro 4	8	16	4
Ala	Cys 2, Arg 2	4	Ser 1, Gly 1	2	6	4
Thr	Cys 2, Arg 2	4	Ser 1, Gly 1	2	6	4
Gly	Pro 2, Ser 2	4	Ala 1, Thr 1	2	6	4
Pro	Arg 2, Gly 2	4	Arg 4	4	8	2, X^{UCA}
Cys	Ala 2. Thr 2	4			4	2
Trp			Pro 2, Ser 2	4	4	2
8	7 occurrences	32	Trp new out of 7	30	62	
Central A:U, NSC Module 2 (homogeneous pDiN)						
Leu	Glu 2, Lys 2	4	Gln 8	8	12	3, X^{YUA}
Gln			Leu 8	8	8	1
Glu	Leu 2, Phe 2	4			4	2
Lys	Leu 2, Phe 2	4			4	2
Phe	Glu 2, Lys 2	4			4	2
5	4	16	Gln new out of 2	16	32	
Central A:U, NSC Module 4 (mixed pDiN)						
Ile	His 1, Tyr 1	4	Asn 1, Val 1	2	6	4
Val	His 2, Tyr 2	4	Asp 1, Ile 1	2	6	4
Tyr	Ile 1, Val 2, Met 1	4			4	3
His	Ile 1, Val 2, Met 1	4			4	4
Met	His 1, Tyr 1	2			2	2
Asp			Asn 1, Val 1	2	2	2
Asn			Asp 1, Ile 1	2	2	2
7	5	18	Asp, Asn new out of 4	8	26	
20	16 occurrences	66	4 new out of 13	54	120	

Table 3 (cont.)

B HOW TO DISTRIBUTE SYNTHETASES TO COMPOSE THE MULTI-SYNTHETASE COMPLEX				
Global connectivity through anticodon pairs		Central G:C Modules 1 + 3 Homogeneous principal dinucleotide core GPS**R** plus Mixed **R**ATCW	Central A:U Mixed Module 4 VMIYH plus Homogeneous 2: DN	Central A:U Homogeneous Module 2 LEFK plus Mixed 4: Q
Three-node cycles		SGA, SGT, SRA, SRT	-	-
Four-node cycles		SRPG, SRPW, GACT, SACT, RACT	VYIH, VYMH, MYIH, VNID	LEFK
Number of nodes connected to a node	3 + 2 1	A T S G R P C W -	Y V H I M D N -	L E K F Q
Anticodon pairs 120 / 20 synthetases = 6		62 / 8 = 7.75	26 / 7 = 3.71	32 / 5 = 6.4
Synthetases in MSC 9 / 20 = 0.45		2 / 8 = 0.25	3 / 7 = 0.43	4 / 5 = 0.8
Synthetases not in MSC		Ser auto-aggregated Cys Trp Gly Ala Thr	Val Tyr His Asn	Phe

Table 3. Connections between synthetases facilitated by the pairing of their correspondent tRNAs and mediated by ribonucleoprotein aggregation. **(A)** Number of anticodon pairs formed by the specificity in the left column with other specificities. The auto-aggregated SerRS is highlighted green. The synthetases belonging to the MSC of the mammalian type are highlighted blue. These detailed data are summarized in the graph sketched in Figure 9. The isolation of the subnetworks is overcome by the cohesive properties of the synthetases and the auxiliary proteins. Synthetase specificity is indicated to have started wide, now shown by the hexacodonics, then entering partial reduction when directed to the principal dinucleotides – forming the tetracodonics, up to the finer tuning of the di- and mono-specificities. **(B)** Only a few hints can be extracted from an evaluation of the information on the tRNA dimers and on the synthetases aggregated into the MSC. **B1.** It is possible that the presently seen structure is still under evolution, but indications may be that the integration by aggregation should be partial. Would this mean: don't over-integrate but leave some specificities free for independent and autonomous work and regulation? While we could not extract regularities referring to specific qualitative amino acid properties, it is possible to infer some quantifications for further evaluation. The central G:C set of pairs is hyperconnected so that it conceded the least number of enzymes to the aggregate and its highest connected specificity SerRS was excluded from hetero-aggregation via auto-aggregation. The two central A:U subnetwoks are lowly connected through tRNA pairs and became more strongly integrated into the system via synthetase aggregation.

Some quantitative observations are consistent with this possibility. Data on the distribution of dimers, the whole matrix of anticodes traverse, were overlaid upon the data on the constitution of the Multi-Synthetase Complexes (MSC). The MSCs that have been observed across the evolutionary scale build an apparent succession of increasing size, expressed as the number of different enzyme specificities that associate into a complex (Table 2). It is noticed that data on plants are lacking from the compilation. The peak sizes are in the Bilateria animal groups, with a variant containing eight enzymes and two auxiliary proteins in the worm *C. elegans*, and another of wider distribution (Crustacea, Insecta, Mammals) composing nine enzymes with three auxiliary proteins [18, 19].

Our counts of the connections between enzyme specificities that would be propitiated by the dimers are detailed in Table 3, where observations on the enzymes that compose the mammalian type of MSC are marked in blue, as well as the auto-aggregated SerRS marked in green. These observations are sketched in a network graph format in Figure 9.

All synthetases that compose the different MSCs were counted, reaching a sum of 31 occurrences. Among these, it has already been pointed out the excess (3/4) of class I enzymes (just 8 class II). It is now realized that this corresponds to the excess of central A and central U attributions (just 7 central G and central C). A possible rationale, merely quantitative, would be that the aggregation into the MSC would be of help to the synthetase activity via stabilization inside the MSC. It is not possible to detail suggestions on the mechanisms since there is only the hint given by the known low ΔG of the central A:U dimers. The stabilization would be of help also in view of the relative isolation of the two small central A:U subnetworks, relative to the tight integration of the central G:C single network.

The scarcity of the central G:C attributions is not easy to explain, even making room for suspicion on the organization of the mixed principal dinucleotide sector. The enzymes lacking in the MSCs correspond mostly to module 3. There are five among the eight synthetases of the central G and C kind missing from the MSCs, GACWT, the last four belonging to module 3 (50%), while there are only three missing among the 12 of the central A and U kind (25%). Module 3 is also the only one not showing an independent evidence of the organization based on anticodon pairs (Figure 1). On the contrary, the termination codes are distributed along the NYA hemirow, which have given room to the proposition of the 'windmill' organization, with different topologies for the modules between the two sectors (17).

Such kinds of aggregate cytoplasmic organizations are usually interpreted as having the function of bypassing the need for transcriptional or translational regulation, therewith providing for quick and strong responses to necessities of the material stored in the aggregates. Some of the material stored may be used as such, which is the case of tRNAs and synthetases when they are recruited for the translational function. The model says that tRNAs would not be naked but associated with proteins: they would leave the MSC as aminoacyl-tRNA already in association with EF1A and, after the translational function, they leave the ribosomes in association with synthetases and get into the MSC again. The MSC aggregate may therefore have a dynamic and variable constitution, so that it is not expected to present precise stoichiometry.

Other functions of synthetases that participate in MSC are non-translational and much varied, a large part of them related to fragments of their proteolytic processing. It is said that these would depend on the release of the synthetases from the MSC so that they could be activated, meaning that the enzymes integrated into the MSC would be in a precursor form with respect to the non-translational function [6, 30]. The same might be said of the tRNAs, whose fragments may acquire different functions [1, 22, 25, 33]. While not being able to rationalize in detail the whole set of functional qualities compiled until now, a generic homeostatic participation of the MSC materials is proposed. It is typical of complex systems to offer challenges to explanatory endeavors.

Conflicts of interest No conflicts of interest are involved with the present communication.

References

1. Balatti V, Nigita G, Veneziano D, Drusco A, Stein GS, Messier TL, Farina NH, Lian JB, Tomasello L, Liu CG, Palamarchuk A, Hart JR, Bell C, Carosi M, Pescarmona E, Perracchio L, Diodoro M, Russo A, Antenucci A, Visca P, Ciardi A, Harris CC, Vogt PK, Pekarsky Y, Croce CM (2017) tsRNA signatures in cancer. PNAS 114 (30) 8071–8076 www.pnas.org/cgi/doi/10.1073/pnas.1706908114

2. Cho HY, Maeng SJ, Cho HJ, Choi YS, Chung JM, Lee S, Kim HK, Kim JH, Eom CY, Kim YG, Guo M, Jung HS, Kang BS, Kim S (2015) Assembly of multi-tRNA synthetase complex via heterotetrameric glutathione transferase-homology domains. The Journal of Biological Chemistry 290, 29313-29328 doi: 10.1074/jbc.M115.690867

3. Colussi, T.M.; Constantino, D.A.; Hammond, J.A.; Ruehle, G.M.; Nix, J.C.; Kieft, J.S. (2014) The structural basis of tRNA mimicry and conformational plasticity by a viral RNA. Nature, 511, 366–369

4. Diss G, Arsenault IG, Coté AMD, Vignaud H, Ascencio DI, Berger CM, Landry CR et al. (2017) Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. Science 355, 630–634

5. Donoghue PCJ, Mark A, Purnell MA (2005) Genome duplication, extinction and vertebrate evolution. Trends in Ecology and Evolution **20**, 312–319 doi:10.1016/j.tree.2005.04.008

6. Fang P, Guo M (2017) Structural characterization of human aminoacyl-tRNA synthetases for translational and nontranslational functions. Methods 113: 83–90 http://dx.doi.org/10.1016/j.ymeth.2016.11.014

7. Francis, B.R. (2015) The hypothesis that the genetic code originated in coupled synthesis of proteins and the evolutionary predecessors of nucleic acids in primitive cells. Life, 5, 467–505

8. Fung, A.W.S.; Payoe, R.; Fahlman, R.D. (2016) Perspectives and insights into the competition for aminoacyl-tRNAs between the translational machinery and for tRNA-dependent non-ribosomal peptide bond formation. Life , 6, 2

9. Goudry, M.; Saugnet, L.; Belin, P.; Thai, R.; Amoureux, R.; Tellier, C.; Tuphile, K.; Jacqet, M.; Braud, S.; Courçon, M.; et al. (2009) Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. Nat. Chem. Biol., 5, 414–420.

10. Grosjean H, Houssier C (1990) Codon recognition: evaluation of the effects of modified bases in the anticodon loop of tRNA using the temperature-jump relaxation method. In: Gehrke CW, Kuo KCT (eds) Chromatography and modification of nucleotides. Elsevier, Amsterdam, pp A255–A295

11. Gruic-Sovulj I, Landeka I, Söll D, Weygand-Durasevic I (2002) tRNA-dependent amino acid discrimination by yeast seryl-tRNA synthetase. Eur J Biochem 269:5271-5279

12. Guimarães, R.C. (1996) Anti-complementary order in the genetic coding system. Int. Conf. Orig. Life, 26, 435–436

13. Guimarães, R.C. (2012) Mutuality in discrete and compositional information: Perspectives for synthetic genetic codes. Cogn. Comput., 4, 115–139

14. Guimarães, R.C. (2013) Formation of the genetic dode—Review and update as of November 2012. Available online: http://www.icb.ufmg.br/labs/lbem/pdf/GMRTgeneticodeNov12.pdf. All original

publications are available online: https://www.researchgate.net/profile/Romeu_Guimaraes (both sites accessed on August 2017)

15. Guimarães RC (2015) Emergence of information patterns: In the quantum and biochemical realms. Quantum Biosyst., 6, 148–159

16. Guimarães RC (2017) Self-Referential encoding on modules of anticodon pairs—roots of the biological flow system. Life, 7, 16; doi:10.3390/life7020016 www.mdpi.com/journal/life

17. Guimarães, R.C.; Moreira, C.H.C.; Farias, S.T. (2008) A self-referential model for the formation of the genetic code. Theory Biosci., 127, 249–270

18. Havrylenko S, Legouis R, Negrutskii B, Mirande M (2011) *Caenorhabditis elegans* evolves a new architecture for the Multi-aminoacyl-tRNA Synthetase Complex. Journal of Biological Chemistry 286 (32): 28476–28487 DOI 10.1074/jbc.M111.254037

19. Havrylenko S, Mirande M (2015) Aminoacyl-tRNA Synthetase Complexes in evolution. *Int. J. Mol. Sci. 16*, 6571-6594; doi:10.3390/ijms16036571

20. Hordijk, W.; Steel, M. (2017) Chasing the tail: The emergence of autocatalytic networks. Biosystems, 152, 1–10

21. Iranzo J, Puigb P, Lobkovsky AE, Wolf YI, Koonin EV (2016) Inevitability of genetic parasites. Genome Biol. Evol. 8(9):2856–2869. doi: 10.1093/gbe/evw193

22. Keam SP, Sobala A, Have St, Hutvagner G 2017 tRNA-Derived RNA Fragments Associate with Human Multisynthetase Complex (MSC) and Modulate Ribosomal Protein Translation. *J Proteome Res*, *16*, 413–420 doi: 10.1021/acs.jproteome.6b00267

23. Kenkel, C.D.; Matz, M.V. (2016) Gene expression plasticity as a mechanism of coral adaptation to a variable environment. Nat. Ecol. Evol., 1, 0014.

24. Lehmann, J.; Libchaber, A. (2008) Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon. RNA, 14, 1264–1269

25. Millán PF, Schelcher C, Chihade J, Masquida B, Giegé P, Sauter C 2016 Transfer RNA: From pioneering crystallographic studies to contemporary tRNA biology. Archives of Biochemistry and Biophysics 602, 95-105 https://doi.org/10.1016/j.abb.2016.03.005

26. Mocibob, M.; Ivic, N.; Bilokapic, S.; Maier, T.; Luic, M.; Ban, N.; Durasevic, I.W. (2010) Homologs of aminoacyl-tRNA synthetases acylate carrier proteins and provide a link between ribosomal and nonribosomal peptide synthesis. Proc. Natl. Acad. Sci. USA, 107, 14585–14590

27. Moras, D.; Dock, A.C.; Dumas, P.; Westhof, E.; Romby, P.; Ebel, J.-P.; Giegé, R. (1986) Anticodon-anticodon interaction induces conformational changes in tRNA: Yeast tRNAASP, a model for tRNA-mRNA recognition. Proc. Natl. Acad. Sci. USA, 83, 932–936

28. Moutiez, M.; Schmidt, E.; Seguin, J.; Thai, R.; Favry, E.; Belin, P.; Mechulan, Y.; Goudry, M. (2014) Unravelling the mechanism of non-ribosomal peptide synthesis by cyclopeptide synthases. Nat. Commun., 5, 5141–5146

29. Murren, C.J.; Auld, J.R.; Callahan, H.; Ghalambor, C.K.; Handelsman, C.A.; Heskel, M.A.; Kingsolver, J.G.;

Maclean, H.J.; Masel, J.; Maughan, H.; et al. (2015) Constraints on the evolution of phenotypic plasticity: Limits and costs of phenotype and plasticity. Heredity, 115, 293–301

30. Ognjenović J, Simonović M (2017) Human aminoacyl-tRNA synthetases in diseases of the nervous system. RNA Biology 23:1-12. doi: 10.1080/15476286.2017.1330245

31. Park JW, Yan ZZ, Loh H, Will SA, Zwierlein MW (2017) Second-scale nuclear spin coherence time of ultracold ^{23}Na^{40}K molecules. Science 357, 372–375

32. Ptashne, M. (2013) Epigenetics: core misconcept. Proc. Nat. Acad. Sci. USA, 110, 7101–7103

33. Schimmel P (2017) The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. Nature Reviews Molecular Cell Biology doi:10.1038/nrm.2017.77

34. Schlosshauer M (2014) The quantum-to-classical transition and decoherence. In Aspelmeyer M, Calarco T, Eisert J, Schmidt-Kaler F (eds.), Handbook of Quantum Information. Springer, Berlin/Heidelberg

35. Seligmann H, Ganesh W (2017) Genetic code optimization for cotranslational protein folding: codon directional asymmetry correlates with antiparallel betasheets, tRNA synthetase classes. Computational and Structural Biotechnology Journal, 15:412–424 doi: 10.1016/j.csbj.2017.08.001

36. Souza GM, Lüttge U (2015) Stability as a Phenomenon Emergent from Plasticity–Complexity–Diversity in Eco-physiology. In: Lüttge U, Beyschlag W (eds.), Progress in Botany 76, 211-239 Springer International Publishing Switzerland, doi 10.1007/978-3-319-08807-5_9

37. Widmann J, Di Giulio M, Yarus M, Knight R (2005) tRNA creation by hairpin duplication. J Mol Evol 61:524–530

38. Xia T, SantaLucia JJr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C, Turner DH (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry 37:14719-14735

39. Zurek WH (1991) Decoherence and the transition from quantum to classical. Physics Today, October, 36-44

40. Zurek WH (2002) Decoherence and the transition from quantum to classical – revisited. Los Alamos Science 27: 2-25