Article

Traces in Current Genomes of a Primitive RNA Ring

Jacques Demongeot *

Posted Date: 10 April 2025

doi: 10.20944/preprints202504.0857.v1

Keywords: origin of life; evolution; amino acid RNA interaction; nucleotide motifs; evolution

 doi:10.20944/preprints202504.0857.v1

Article

Traces in Current Genomes of a Primitive RNA Ring

Jacques Demongeot

University of Grenoble Alpes, AGEIS EA 7407, Faculty of Medicine, 38700 La Tronche, France;
jacques.demongeot@univ-grenoble-alpes.fr

Abstract: (1) Background: Previous theoretical studies have provided arguments for the existence of a ring or hairpin RNA that could have served as a primitive informational and functional molecule at the origin of life. The present article consists of searching in current genomes for RNAs closest to this ring in terms of occurrence of similar nucleotide motifs. (2) Methods: In searching for the smallest possible ring/hairpin RNA capable of interacting with amino acids in the construction of the peptides of the primitive living world, we found a circular docosamer RNA molecule (length 22), which we called AL ring (for Alpha or Archetypal Loop). Then, we started to systematically track AL relics in current genomes in the form of motifs like pentamers or pairs of consecutive codons in common with AL. (3) Results: The sequence correspondence between AL and RNA sequences of organisms from different kingdoms of life (Archaea, Bacteria and Eukarya) was found with high statistical significance with a frequency gradient depending on both the antiquity of the species and the functional necessity of the genes. (4) Conclusions: Considering the suitability of AL as a candidate for being a primitive sequence, and the evolution of the different species considered, we can consider the AL RNA ring as a possible actor that favored the appearance of life on Earth.

Keywords: origin of life; evolution; amino acid RNA interaction; nucleotide motifs; evolution

1. Introduction

For 55 years, considerable efforts, both theoretical and experimental [1–16], have been made to demonstrate that before the emergence of the ribosomal machinery, molecular assemblies involving RNA molecules and amino acids could have given rise to the first peptides. In this article, we focus on an RNA molecule, a candidate for the role of peptide catalyst at the origin of life. To find it, we selected four criteria from information theory and arrived at a unique RNA molecule, which we called AL (for Archetypal Loop or Alpha Loop), in which we discovered 18 biological properties concerning its fit with the sequences and motifs of current genomes. Section 2 will present the materials and methods, followed by Section 3 with the results obtained, and the final sections with discussion, conclusion, and outlook.

2. Material and Methods

2.1. The Stereochemical Theory of the Origin of Life

Considered by Eigen [8,9] as the first "function" of life, proteinogenesis requires adequate production of peptides, an absolute necessity for evolution, as suggested in 1951 by Bernal who said that this process could be favored on very fine clay deposits such as montmorillonite [10]. As a "polymerization catalyst", montmorillonite would indeed have the consequence of decreasing the content of free amino acids following their polymerization [1–7]. In 1963, Ponnamperuma and his collaborators described the formation of ATP under possible primitive terrestrial conditions [15], and in 1995, they proposed the interactions between amino acids and nucleotides as a possible physicochemical basis for the origin of the genetic code [16]. All of these observations form the experimental corpus of the stereochemical theory of the origin of life [17–19]. Shapiro [20] admitted that "life began in a mixture of simple organic molecules, with possible participation of minerals",

 doi:10.20944/preprints202504.0857.v1

but with Bernhardt [21] he was critical of the montmorillonite hypothesis, the alternative (or complement) being hydrothermal chimneys, i.e., cracks between tectonic plates with discharges of geothermally heated water [22]. Yarus for his part defended the idea of a catalytic role of RNA rings promoting peptide bonds between amino acids [23–25] and recent work has emphasized the role of lipids in the very early stages of life [26–29].

2.2. Theoretical Criteria

Four theoretical criteria for a primordial RNA ring (AL) to be a candidate for primordial catalysis of peptide biosynthesis are optimal combinatorial properties, already identified and published [30–60], and can be summarized as follows:

1) The AL must satisfy the principle "be as short as possible and contain at least one codon per synonymy class of the genetic code",

2) The AL codon sequence obtained with overlap after 3 turns of the ring must begin with the start codon and end with the stop codon,

3) The AL must have a hairpin configuration in balance with its circular shape, and this hairpin must have a minimum head length (3 nt) and a maximum number (9) of codon pairs;

4) If multiple rings possess properties 1) to 3), they must have a single barycenter for classical inter-ring distances (circular Hamming, permutation, and editing distances), i.e., the AL ring.

2.3. Progressive Deciphering of the AL Ring

The discovery of the AL ring occurred in four stages. In 1975, a first 22-nucleotide ring satisfying criteria 1) and 3) was discovered among the 4^{22} possible rings of length 22 [30,31]. This ring was called the cyclic code because it represented a non-degenerate summary of the complete genetic code with 64 codons. The ring had a short hairpin configuration with only 6 hybridized nucleotide pairs (in red and blue, nucleotides in green being not hybridized):

GCCAT TCAG A A-3'

TGGTA TCAG T A-5'

In 1983, a second ring called C3 (because it presented 3 zones of hybridization) was published [32] with a longer hairpin (8 hybridized pairs) starting with AUG and having UGA among its codons, but not at the end after 3 laps:

TGGTG AA GA C G-3'

ACCAT AA CT T C-5'

In 1996, a third ring called AB (for Ancestral Basal) has been discovered [33]. He had 9 hybridized pairs, but not contiguous, and verified criteria 1) and 2):

GCCATTCAAG A-3'

TGGTAAGTAT C-5'

In 2004 [34], it was established that criterion 1) had no solution for a cycle of length 20 or 21, but only of length 22, for which existed 29,520 solutions (out of the 4^{22} possible solutions) containing only one repeated codon AXN repeat codon, with X being G for 52% of the solutions. In 2006 [35], an attempt to explain the degeneracy of the genetic code from a non-degenerate cyclic code was proposed. In 2007 [36], it was finally shown that among the 29,520 solutions, only 25 cycles satisfied criteria 1) and 3) with the existence of a hairpin of 9 or more nucleotides, of which only 19 encompassed both the start and a stop codon, and 9 satisfied criterion 2). By calculating several distances (e.g., circular Hamming distance, permutation distance, and edit distance), the singular ring called AL (for ALpha or Archetypal Loop) ATGGTACTGCCATTCAAGATGA had a minimal average distance to the other 18, thus acting as their unique barycenter satisfying all criteria 1) to 4):

TGCCATTCA A -3'

G CATGGTAAGT A -5'

2.4. AL-Codon-Counter, an Algorithm for Finding AL Traces in Current Genomes

 doi:10.20944/preprints202504.0857.v1

The AL-Codon-Counter algorithm detects in any RNA sequence certain motifs in common with AL, using, for example, a sliding window of five nucleotides to find pentameric motifs [60]. This algorithm systematically analyzes the sequence by shifting the window by one nucleotide at a time, thus capturing all possible pentamers. Once identified, the pentamers are mapped onto the RNA sequence and the distances between consecutive occurrences of each motif are calculated. This process includes handling overlapping pentamers and taking into account cases where pentamers are close to each other. The program calculates various distance measures, including the mean, median, standard deviation, and mode of distribution of these motifs, thus providing useful statistical information. In addition, pentamer analyzis is extended to the study of evolutionary trends, particularly hypothetical remnants of the ancestral AL ring. The program's statistical framework allows for comparison of pentamer distributions across multiple species, potentially revealing the evolutionary conservation of identified motifs (see for example Figure 1).

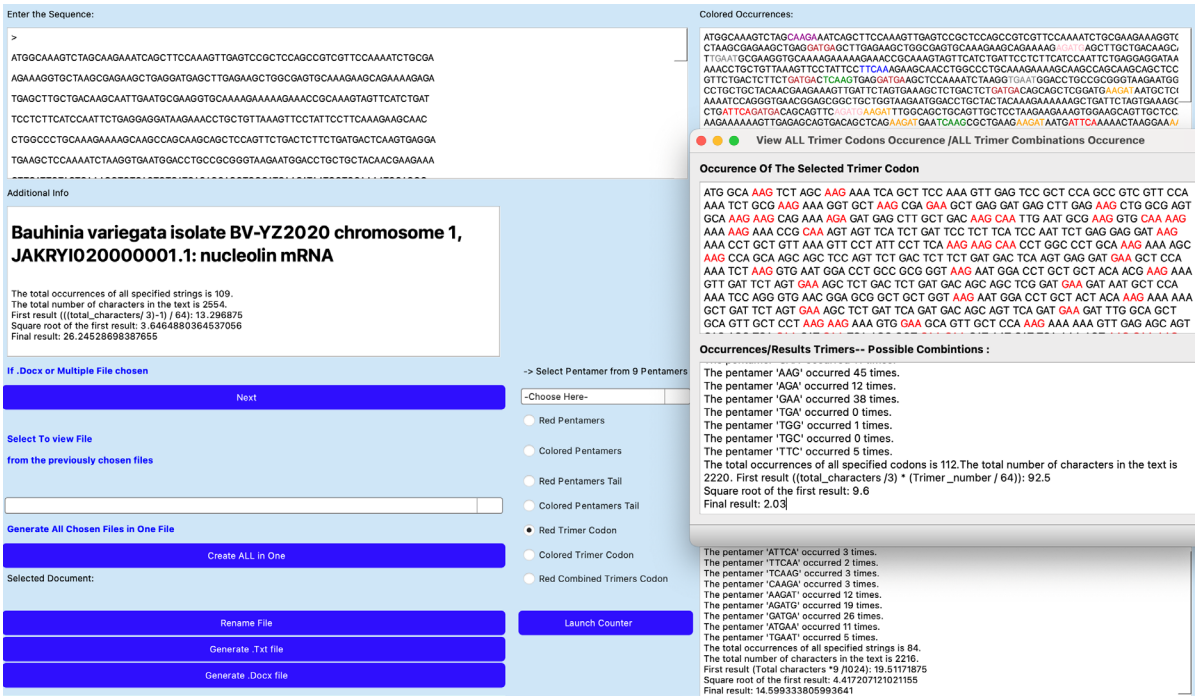

Figure 1. Screen of AL-Codon-Counter application showing the processing of the mRNA sequence of nucleolin gene from Bauhinia variegate. It shows an AL-proximity of 14.6 (resp. 26.2) due to pentamers (resp. pairs of codons) common between AL and the mRNA sequence.

The 9 pentamers of the head of the hairpin form of AL all have at least one nucleotide linked to a nucleotide of the AGA head, which explains their fragility and their tendency to detach from AL, explaining the fact that they are found in RNAs involved in evolution with a decreasing frequency as we move away from the origin of life. These 9 pentamers are: AUUCA, UUCAA, UCAAG, CAAGA, AAGAU, AGAUG, GAUGA, AUGAA, UGAAU. The proximity to AL P_{PAL} of an RNA sequence is obtained by calculating the number n_o of these 9 pentamers observed in the RNA sequence, and the expected number n_e equal to the expectation of a binomial distribution B(n,p), where n is the size of the RNA sequence minus 4 and p = 9/1024 (the probability of observing all 9 pentamers). The proximity P_{PAL} is then equal to twice the number of standard deviations $(npq)^{1/2}$ in the interval $[n_e, n_0]$, which allows the calculation of the probability p to observe n_o by chance using the Gaussian approximation of the binomial distribution: for example, if P_{PAL}=8, then $p < 5\ 10^{-5}$. Another proximity called P_{PAL} Doublet is calculated by counting the number of pairs (or doublets) of successive codons in a given mRNA. If the occurrence of a bond between two amino acids has been initially due to their reversible weak link to a pair of codons close on AL acting as a proto-ribosome, this pair could occur with a significant frequency in the mRNA. To show that, the number observed along the mRNA of pairs of AL codons (such as ATT, CAA, GAT, GAA, CCA, AGA, ACT or AAT,

 doi:10.20944/preprints202504.0857.v1

TAC or AAG) is calculated, as well as its expected number and then, P_{PAL} Doublet is obtained in the same way as P_{PAL}.

3. Results

3.1. Biological Properties

In the following, some biological properties of the circular and hairpin forms of AL (Figure 2) will be explored, with reference to current genomes, in which some of these properties persist, such as the survival of common motifs between sequences of these genomes and the AL sequence.

Figure 1. Circular form of the ring AL fitting the loops of the tRNA-Gly^{GCC} of Methanococcus maripaludis (bottom) with inside the ring, the hairpin configurations of AL (left) and anti-AL (right).

The biological properties of AL correspond to the 20 following optimal properties:

1) All dinucleotides appear in AL, except CG, the least frequent dinucleotide in Archaea [61] and archaeal virus genomes [62]. Among the AL codons, 12 belong to the set of the 20 most frequent codons of chloroplasts [63],

2) AL fits well the loops of tRNA-Gly^{GCC} of Arabis alpina mitochondrion [64], and more generally, the set of the most invariant nucleotides essentially located in the loops of the tRNAs from the database GtRNAdb (Figure 2).

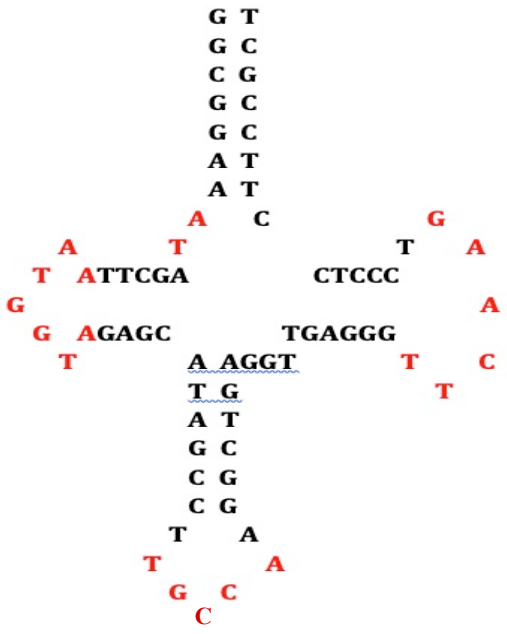

Figure 2. tRNA-Gly^{GCC} of Arabis alpina mitochondrion [65,66].

3) AL among the rings verifying criteria 1) to 4) is the closest in mean edit distance to all tRNAs of GtRNAdb species [65] belonging to the three domains of life, Archaea, Bacteria and Eukarya, whose list of full names is given on Figure 3, list of short names and sequences on Figure 4 and phylogenetic tree on Figure 3,

4) 50% of the tRNAs of GtRNAdb have the edit distance of their loops to AL less than 4 [37],

5) The 4 domains of any tRNA (3 loops and one articulation pivot) are ranked in their natural order inside AL.

Salinarchaeum Harcht, Sulfobacillus acidophilus, Thermococcus, Thermofilum carboxyditrophus, Acidobacterium capsulatum, Blochmannia endosymbiont, Candidatus Carsonella ruddii, Erysipelothrix

 doi:10.20944/preprints202504.0857.v1

rhusiopathiae, Faecalitalea cylindroides, Mycoplasma agalactiae, Posidoniimonas polymericola, Thuidium tamariscinum, Sphaerobolus stellatus, Angelica biserrata, Arabidopsis thaliana, Bauhinia binata, Chara braunii, Cuscuta campestris, Ceratodon purpureus, Gossypium raimondii, Nicotiana tabacum, Oenothera villaricae, Solanum lycopersicum, Sphagnum fallax, Sphagnum magellanicum, Spirogloea muscicola, Triticum aestivum, Vitis vinifera, Alligator mississippiensis, Balaenoptera acutorostrata scammoni, Bos taurus, Procavia capensis, Rhytidiadelphus loreus, Sorex araneus, Spodoptera frugiperda, Strongylocentrotus purpuratus, Taeniopygia guttata, Tarsius syrichta, Trichechus manatus, Tursiops truncatus, Vicugna pacos, Xenopus tropicalis.

Figure 3. List of species from the three domains of life Archaea (mauve), Bacteria (green) to Eukarya with Fungi (violet), Plants (red) and Animals (blue).

Sali GCGTCGGTAGTGTAGTGGTATCACGTGACCCTGCCACGGTCGCAaCCCGAGTTCAAATCTCGGCCGACGCA
Therm GCGGTGGTAGTCTAGCCtGGTctAGGACAGCGGCCTGCCACGCCGCTGGCCCGGGTTCAAATCCCGGCCACCG
Therc GCGGCCGTAGTCTAGTCtGGTAGGATGGCGGCCTGCCACGCCGCAGAaCCCGGGTTCAAATCCCGGCGGCCG
Sulf GCCGTAGTCTAGCATGGAtAGGACGCCTGCCTGCCACGCAGGAGGtCCCGGGTTCAAATCCCGGCGGCCGCA
Acid
GCGGGAGTAGCTCAGTGGTAGAGCATCGCCTTGCCAAGGCGAGGGtCGCGGGTTCAAGTCCCGTCTCCCGCTCCA
Blocen GCGGGAATAGCTCAGTTGGTAGAGTACAACCTTGCCAAGGTTGGGGtCGCGAGTTCAAGTCTCGTTTCCCGCT
Cand GCGAAAGTATCTTAATGGTAAAGTATCACCTTGCCATGGTGAAAGtTGCGAGTTCAAGTCTCGTCTTTCGCT
Erys GCAGGTGTAGTTCAATGGTAGAACACGACCTTGCCAAGGTTGAGGCGGGGGTTCAATTCCCCTCACCTGCTCCA
Faec GCAGATGTAGTTCAATGGTAGAACACAGCCTTGCCAAGGCTGATaCGGGGGTTCAATTCCCCTCATCTGCTCCA
Myco GCAAATGTAGTTCAATGGTAGAACACCAGCTTGCCATGCTGGATaCGGGGGTTCAATTCCCCTCATTTGCTCCA
Popo GCGGGTGTAGTTCAATGGTAGAACGAAAGCTTGCCAAGCTTTAGGCGAGGGTTCGATTCCCTTCACCCGCTC
Ttama GCGGAAATAGCTTAATGGTAGAGTATAGCCTTGCCAAGGCTGAGGtTGAGGGTTCAAGTCCCTTTTTCCGCT
Spha GCATTAATGGGGTAATGGTAACCTTAGTCGTTGCCATCGACTAGcCGCGAGTTCAATTCTTGCTTAGTGCA
Angebi
GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTGAGGTTGAGGGTTCAAGTCCCTTCTTCCGCT
Athal GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTAAGGtTGAGGGTTCAAGTCCCTCCTTCCGCT
Baubi
GCGGATATAGTCGAATGGTAAAATTTCTCTTTGCCAAGGAGAAGACGCGGGTTCGATTCCCGCTATCCGCC
Cbrau GCGGAAATAGCTTAATGGTAGAGTATAGCCTTGCCAAGGCTAAGGtTGAGGGTTCAAATCCCTTTTTCCGCT
Ccamp GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTGAGGtTGAGGGTTCAAGTC CCTCCTTCCGCT
Cpurp GCGGAAATAGCTTAATGGTAGAGTATAGCCTTGCCAAGGCTGAGGtTGAGGGTTCAAGTCCCTTTTTCCGCT
Graim GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTGAGGtTGAGGGTTCAAGTCCCTCCTTCCGCT
Ntaba GCGGAAATAGATTAATGGTAGATCATAGCCTTGCCAAGGCTGAGGtTGAGGGTTCAAGTCCCTCCTTCCGCT
Oevill GCGGAAATAGCTTAATGGTAGAGCGTAGCCTTGCCAAGGCTGAGGTTGAGGGTTCAAGTCCCTCCTTCCGCT
Sfall GCGGAAATAACTTAATGGTAGAGTATAGCCTTGCCAAGGCTGAGGtTGAGGGTTCAAGTCCCTTTTTTCGCT
Slyco GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTAAGGtTGAGGGTTCAAGTCCCTCCTTCCGCT
Smage GCGGAAATAGCTTAATGGTAGAGTATAGCCTTGCCAAGGCTGAGGtTGAGGGTTCAAGTCCCTTTTTCCGCT
Smusc
GCGGATGTAGCTCAATGGTAGAGTatgtgtATAGCCTTGCCAAGGCTGAGGtTGAGGGTTCAAGTCCCGTTTTCAGCT
Taest GCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCATGGTACAGaCCTGGGTTCAATTCCTGGCTGGTGGA
Vvini GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTGAGGtTGAGGGTTCAAGTCCCTCCTTCCGCT
Amiss GCATTGGTGGTTCAATGGTAGAATTCCTGCCTGCCATGCAGGAGaCCTGGGTTCAATTCCCAGCCAATGCA
Bacut TCCCTGGTGGCACAGTGGTTGAGAATCTGCCTGCCAgTGCAGGGGGaCACACATTCAAGCCCTGgTCCAGGAAG
Btaur TCCTAGGTGGCTCAATGGTaAAAATACTTGCCTGCCAaAGCAGGAGaCTCAGGTTCAATCCCTGGTCCAGGAA
Pcape GGACTAATGGCATAGTGGTgAAAGTGTTTGGCTGCCAACCAAAAGGtCAGTGATTCAAACCCACTagctGTTTT
Rlore GCGGAAATAGCTTAATGGTAGAGTATAGCCTGCCAAGGCTGAGGtTGAGGGTTCAAGTCCCTTTTTCCGCT
Rnorv GaGAgATGGCTCAATGGTtaAGAGCACTGACTGCCGtgctcTTCCAGAGGcCGTGAGTTCAAGTCTCAGTAACCA
Saran GCTcGATTCAGTGGTAGAATTCTCGGCTGCCACGTGGGAGGCCCGGGTTCAATTCCCGGcccaTGCA
Sfrug GCACGGTTGGCGCAGTGGCTGGGCAACCGCCTGCCGCGCCATGTGtGGTGGGTTCAATTCCCACACGGTGCA
Spurp GCATCGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCGGGGGaCCCGGGTTCAATTCCCGGCCGATGCA
Tgutt GCCCTGGTGGCTCCGTGGTAGAATTCTGCCTGCCACGGCGGCAgCCTGGGTTCAATTCCCGGCAGAGGCA
Tmana TCCTGGGTAGTACAAATGGTAACGCACTCGGCTGCCAACCGAAAGGtTGGTGGTTCAAACCCACCCCCAGAAG
Tsyri ACATGGGTAGTTCAGTGGTAGAATTCTCGCCTGCCACACAGGAGGCCCGGATTCAATTACTAACCCATGCA
Ttrun GCATTGGTGATTCAGTGGTAGAATTCTTGCCTGCCACGTGGGAGGCCAGGGTTCAATTCCCAGCCAATGCA
Vpaco GGAGGTATAGCTCAGTGGTAGAGTGCATGCCTGCCATGCACAAGGtCCTGGGTTCAATCCCCAGTACCTCCA
Xtrop GCATTGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCAATTCCCGGCCAATGCA

 doi:10.20944/preprints202504.0857.v1

Figure 4. Sequences of tRNA-Gly^{GCC} from species belonging to the three domains of life, Archaea, Bacteria and Eukarya.

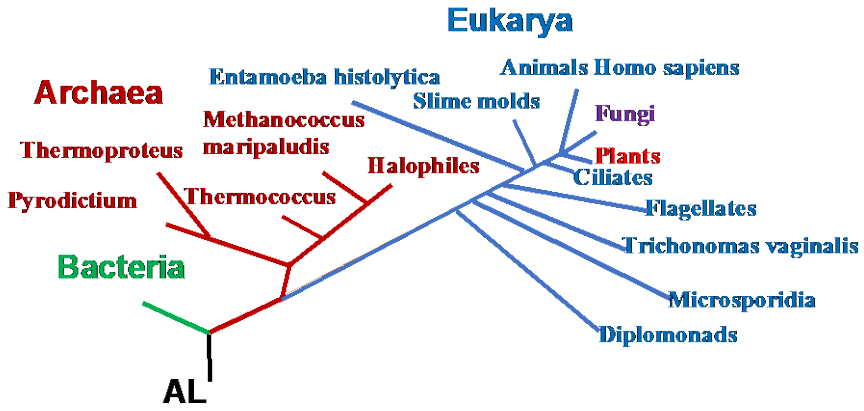

Figure 5. Phylogeny of the 3 domains of life, Archaea, Bacteria and Eukarya, with indication of some species.

6) The average edit distance from AL to 20,000 different randomized versions of randomly repeated microRNAs preserving length 22 and base composition of AL is significantly larger [37] than the average edit distance of AL to the real microRNAs from the database miRBase [66],

7) AL has at least 15 common nucleotides with the barycenter of these 20,000 repmirs, whose edit distance to AL is less than 7 [37],

8) AL fragments match exon/intron boundary [67,68] with sequences 5'-3' GGTAC or 3'-5' TGAATGG (Figure 6),

Figure 6. exon/intron boundary (Left from [67], Right from [68]). The black arrow indicates the splicing site.

9) AL matches with Hamming and edit distances ≤ 2 with at least 43 tRNA-Gly from GtRNAdb from the 3 domains of life, Archaea, Bacteria and Eukarya (Figure 4),

10) In the anticodon position, AL has "GCC" suggested as the first anticodon, because it "anticodes" for the simplest amino-acid, the glycine,

12) AL aligns with the main articulation pivot "AUG" and allows the pairing TGG-ΨCA, needed between the D- and TΨ-loops of tRNAs for their 3d-folding,

13) AL matches well with many not coding genomes from viral origin [38],

14) AL matches well with many microRNAs [40], IRE and YUNR loops [42] as well as circular RNAs [42],

15) AL contains twice all the most unexpected dimers as defined by P.P. Slonimski [73],

16) There is an experimental evidence of direct RNA-amino acid interactions with AL-pentamers GCCAU [74] and AUGGU [75–78],

17) The CRISp-R cas9 system shows in the guide RNA sequences the occurrence of AL-heptamers like GAAUGGU [79] and AAGAUGA [80],

18) Complete genome of one the oldest Bacteria, Cyanobacterium aponinum, contains a significant proportion of AL-codons from the set {CCA, ATT, CAA, AAG, GAT, AGA, GAA, AAT}, such as the distance between observed and expected numbers of such codons is more than 212 standard deviations (cf. Supplementary material 1), and complete genomes of Methanococcus maripaludis (Archaea), Dojkabacteria bacterium (Bacteria), Clitoria ternatea plasmid and Oenothera

 doi:10.20944/preprints202504.0857.v1

villaricae chloroplast (plants), and mitochondrion of Jaculus jaculus (mammal) have their proximity PPAL Doublets in their decreasing order in evolution, i.e., 312.5, 224.1, 93, 92.4 and 4.7 (cf. Supplementary material 2),

19) The AL heptamer TCAAGAT is part of the palindromes located upstream of replicase genes in Rhodobacterales repABC-9 replicons, and in replication units of the alphaproteobacterial plasmids [81],

20) Twelve hexadecameric peptide sequences of 16 amino acids from MVLPFKMNGTAIQDEW to IQDEWYCHSRMVLPFK corresponding to 16 successive codons without overlap on AL (see Figure 7 Top) are observed in 332 proteins with a probability of observing that by chance equal to 4 $10^{-12} \pm 3\ 10^{-6}$, these proteins having been selected by NCBI Blast [66] from 117,262,330 protein sequences with a total number of 42,988,570,095 amino-acids. Among these 332 proteins, many come from extremophiles of the Rhodobacterales family, like Roseivivax marinus, Ponticoccus litoralis, Thiobacimonas profunda and Tropicibacter naphthalenivorans.

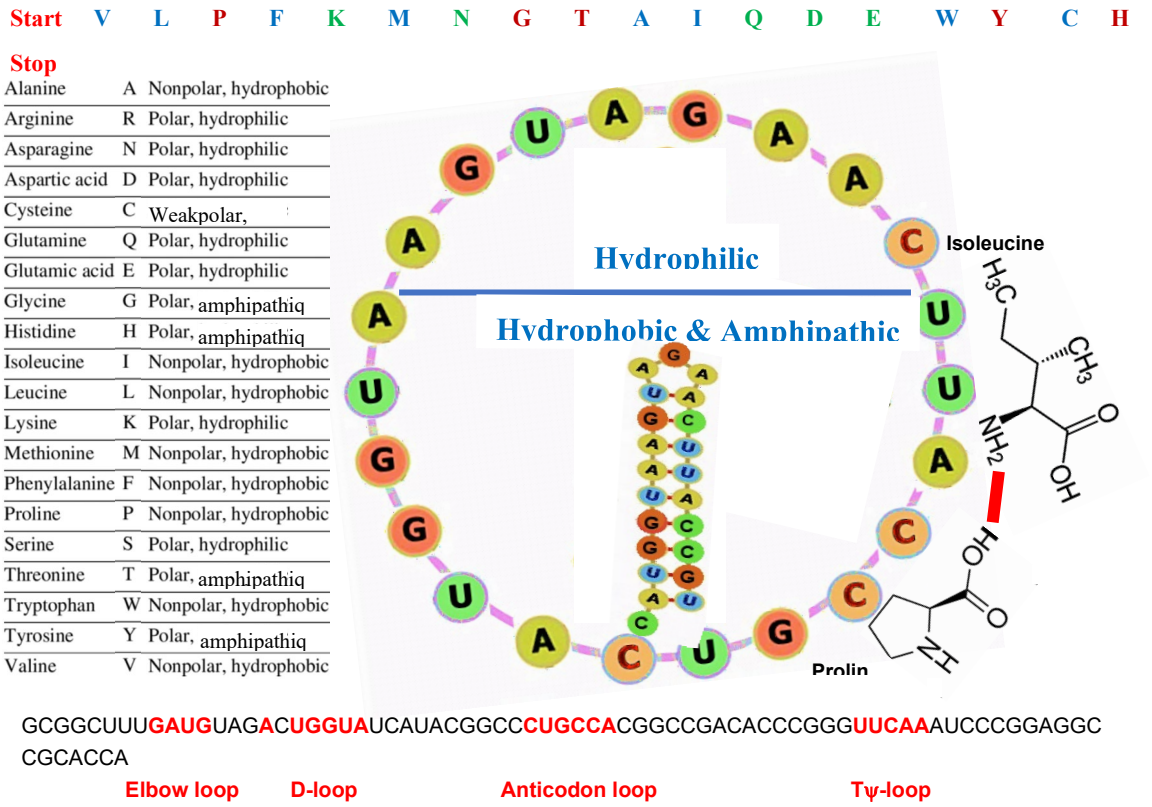

Figure 7. The evolutionary machinery. On the top, the succession of codons without overlap and corresponding amino acids of AL. On the left, the amino-acid polarities. On the right, AL in the catalytic function of its circular form, where the codons CCA for proline and AUU for isoleucine temporarily attract their amino acids through weak electromagnetic binding, promoting the creation of a strong peptide bond between them. The lower part of AL contains codons (Met, Trp, Gly, Val, Tyr, Thr, Leu, Cys, Ala, Pro, His, Ileu, Phe) corresponding to hydrophobic amino acids (Met, Val, Leu, Cys, Ala, Pro, Ileu, Phe) or amphipathic amino acids (His, Thr, Tyr). The upper part contains codons corresponding to hydrophilic amino acids (Ser, Gln, Lys, Arg, Asp, Glu, Asn), plus START (AUG) and STOP (UGA) codons.

3.2. Functional properties of AL

The four basic functional activities considered in the following are membrane transport (ATPase, translocase), proteolysis (FtsH), translation (ribosomal RNAs and proteins, and aminoacyl-tRNA ligases), and RNA synthesis (RNA polymerase, helicase and gyrase) [82–84]. The AL RNA capable of

 doi:10.20944/preprints202504.0857.v1

replication as in a "quine" program, leaves functional traces in current "vital" RNAs, from its two forms in equilibrium: a reactive circular form, the circAL (circular or ring AL) and a stable short hairpin form, the shAL (or short hairpin AL). The circAL hybridizes with its complement constructed by affinity of AL nucleotides with their complements (A with U; G with C). This antiAL complement exhibits the same stable hairpin characteristics (identical to shAL, except for the head and tail), in equilibrium with a circular form capable of restoring AL by the same process (Figure 2). The main function of AL could have been that of a "protoribosome" favoring peptide bonds between amino-acids interacting with its codons (Figure 7), as predicted by Ponnamperuma [16] and experimented by Tamura and Schimmel [75–78], and Yarus [25].

3.3. Searching for AL Motifs in Current Genomes

The 8 pentamers at the head of the hairpin form of AL possess all at least one nucleotide having a link with a nucleotide of the AGA head, which causes their fragility and the fact that they are observed in the RNAs of many species during evolution with a frequency less and less important as we move away from the origin of life. These pentamers are the following: AUUCA, UUCAA, UCAAG, CAAGA, AAGAU, AGAUG, GAUGA, AUGAA, UGAAU. The proximities to AL, P_{PAL} and P_{PAL} Doublet, are calculated on Table 1 for tRNA-Gly and mRNA sequences of the genes of gyrase, helicase, translocase, ATPase, RNA polymerase, Gly-tRNA ligase, PFK, FtsH and rprotein L18, for 5 species from the oldest to the earliest : Methanococcus maripaludis (Mm), Trichomonas vaginalis (Tri), Entamoeba histolytica (Ent), Saccharomyces cerevisiae (SC) and Homo sapiens (HS) (see Supplementary material 3).

Table 1. n_o (resp. n_e) is the observed (resp. expected) number of pentamers belonging to the head of the hairpin form of AL, and P_{PAL} is twice the number of empirical standard deviations of n_e contained in (n_o - n_e). The calculation is identical for the P_{PAL} Doublet. P_{PAL} and P_{PAL} Doublet are measures of the proximity of RNAs (tRNAs or mRNAs of the 5 observed species) to AL.

 doi:10.20944/preprints202504.0857.v1

RNAs	Species	n_o	N	n_e	(σ_e)	$Pp_{AL} = 2(n_o-n_e)/\sigma_e$	Mean Pp_{AL}	Pp_{AL} Doublet	Mean Pp_{AL} Doublet
rprotein L18	HS	14	639	5.7	(2.38)	7	10.1	4	8
	SC	19	536	4.7	(2.2)	13.1		7.4	
	Ent	9	536	4.71	(2.17)	4		5	
	Tri	25	562	5	(2.23)	18		7.1	
	Mm	8	227	2	(1.42)	8.4		16.2	
mRNA FtsH	HS	36	1918	16.9	(4.1)	9.3	11.4	1.3	13
	SC	53	2968	26.1	(5.1)	10.52		4.4	
	Ent	21	944	8.3	(2.9)	8.8		27.1	
	Tri	58	1598	14.1	(3.8)	23.4		23.4	
	Mm	23	1457	12.8	(3.6)	5.7		9	
mRNA PFK	HS	43	3036	26.7	(5.2)	6.3	10.6	8.6	13.5
	SC	79	2960	26	(5.1)	23.4 20.8		1.2 26.8	
	Ent	21	1413	12.5	(3.5)	4.8		2	
	Tri	25	1286	11.3	(3.4)	8.1		9.5	
	Mm	35	1385	12.2	(3.5)	13		20.7	
mRNA Gly-tRNA ligase	HS	34	2230	19.6	(4.4)	6.5	10.8	6.1	19.5
	SC	42	1856	16.3	(4)	12.7		20.1	
	Ent	44	1880	16.5	(4.07)	13.6		36.6	
	Tri	39	1946	17.1	(4.14)	10.6		8.7	
	Mm	36	1721	15.2	(3.9)	10.8		26	
mRNA RNA polymerase	HS	75	3959	34.8	(5.9)	5.5	13.5	8.6	18.9
	SC	30	1316	11.6	(3.4)	14.5		16	
	Ent	72	3194	28.1	(5.3)	10		11.7	
	Tri	29	1040	9.2	(3)	23.2		25	
	Mm	45	2351	20.7	(5.5)	14.4		33.3	
mRNA ATPase	HS	78	3414	30	(3.9)	17.5	15.7	19.1	21.6
	SC	42	1850	16.3	(4)	12.8		14.3	
	Ent	52	1832	16.1	(4)	17.9		30.1	
	Tri	35	1366	12	(3.5)	13.2		10.4	
	Mm	98	2978	26.2	(5.1)	28		39.4	
mRNA Translocase	HS	15	1027	9.1	(3)	4	17.9	4	19.1
	SC	133	4856	42.7	(6.53)	27.6		23	
	Ent	106	3002	26.4	(5.14)	31		38.1	
	Tri	33	1066	9.4	(3)	15.4		14.8	
	Mm	20	1325	11.7	(3.4)	4.9		15.3	
mRNA Helicase	HS	73	2716	23.9	(4.9)	20.1	20.5	16.4	19.7
	SC	110	4541	39.9	(6.32)	22.2		16.6	
	Ent	63	2573	22.6	(4.76)	17		30.9	
	Tri	46	1256	11.1	(3.33)	21		11.3	
	Mm	56	2236	19.6	(4.4)	22.2		23.2	
mRNA Gyrase	HS	138	5691	50	(7.1)	24.9	32.5	26.2	29.9
	SC	122	4283	37.7	(6.14)	27.4		34.6	
	Ent	176	4046	35.6	(6)	47		55.7	
	Tri	162	4376	38.5	(6.2)	39.8		33.6	
	Mm	66	1094	9.65	(3.1)	23.2		24	
tRNA-Gly	HS	18	22	0.19	(0.44)	81	84.6		
	SC	17	22	0.19	(0.44)	76.4			
	Ent	19	22	0.19	(0.44)	85.5			
	Tri	19	22	0.19	(0.44)	85.5			
	Mm	21	22	0.19	(0.44)	94.6			
RNA AL		22	22	0.19	(0.44)	100	100		

Among the 5 species of Table 1, Methanococcus maripaludis (Mm), Trichomonas vaginalis (Tri), Entameoba histolytica (Ent), Saccharomyces cerevisiae (SC), and Homo sapiens (HS), the most frequent pairs of consecutive codons in the mRNA of their gyrase have been calculated (Table 2), showing that the most frequent are those corresponding to identical or close (but without overlap) AL codons corresponding to hydrophilic amino-acids (GAA-GAA, GAA-GAT, GAA-GAT, GAA-AGA, GAT-GAA) and a pair of codons corresponding to a pair of hydrophobic and hydrophilic acids (ATT-GAA). This observation reinforce the hypothesis of the primordial catalytic role of AL favoring the peptide synthesis at the origin of life.

 doi:10.20944/preprints202504.0857.v1

Table 2. Pairs of consecutive codons observed more than 2 times at least once among the 5 species of Table 1.

Doublet	Saccharo myces	Tricho monas	Methano coccus	Homo sapiens	Enta moeba	Total
ATT ATT	8		5	3	20	36
ATT CAA	5	4	11	2	14	36
ATT GAT	11		8	3	12	34
ATT GAA	17	3	27	16	25	88
ATT CCA	2	6	7	4	2	21
ATT ACT	10	3	3	1	4	21
ATT AGA	5	3	1	3	4	16
ATT TAC			3		2	5
CAA ATT	4		1	6	9	20
CAA CAA	5	4		2	7	18
CAA GAT	6	5		7	4	22
CAA GAA	12	6	6	11	11	46
CAA CCA					4	4
CAA AGA	4		4	6	9	23
CAA TAC	6			2	1	9
CAA ACT				7	5	12
GAT ATT	14		9	9	21	53
GAT GAT	14	15	4	19	26	78
GAT GAA	23	16	26	22	35	122
GAT CCA	3	3		5	4	15
GAT AGA	2	1	4	5	9	21
GAT ACT	6		7	1	6	20
GAT CAA	2		1	1	6	10
GAT TAC		5	5	2		12
GAA ATT	6	2	7	9	27	51
GAA CAA	11	2	7	6	12	38
GAA GAT	21	14	14	20	33	102
GAA GAA	33	16	37	20	43	149
GAA CCA	9	6	2	3	11	31
GAA ACT	10	1	6	8	8	33
GAA AGA	10	3	7	8	7	35
GAA TAC	10	4	4			18
CCA ATT	3	4	6	2	7	22
CCA GAA	6	3	4	8	7	28
CCA AGA		4	1	8	3	16
CCA GAT				8	3	11
CCA ACT				6	6	12
ACT CAA	3	4	1		9	17
ACT GAT			4	4	7	15
ACT GAA		3	6	8	5	22
ACT ATT	6		2	2	9	19
ACT CCA	9		1	7	4	21
ACT TAC			2	1		3
ACT ACT	7		2		4	13
AGA ATT	5	1	5	3	7	21
AGA GAT	12		5	6	11	34
AGA GAA	11	9	20	13	12	65
AGA CAA	9	11		4	2	26
AGA AGA	6		4	7	7	24
AGA CCA					2	2
AGA TAC			4	1		5
TAC ATT	3	4	1	2		10
TAC GAT	5		6	1	1	13
TAC GAA		5	10	1	1	17
TAC TAC		3	2	3		8
TAC AGA		3	2			5
TAC CCA		1	5	3	1	10

To conclude, it can be considered that the AL ring could belong to a family of ancient RNAs made from diverse RNA types involved at the Origin of Life (OL), these OL-RNAs close to AL, such as riboswitches, ribozymes, rRNAs, tRNAs, circRNAs and mRNAs of essential proteins, that are presumably close to ancestral RNAs. In perspectives, the data summarized on Figure 8 show the double dependency of the proximity to AL on the seniority on the species axis and of the functional necessity on the RNA-axis. The future work would concern more species from the 580.000 species of NCBI GenBank (formally described in October 2024) in order to confirm the tendencies shown in the present paper.

 doi:10.20944/preprints202504.0857.v1

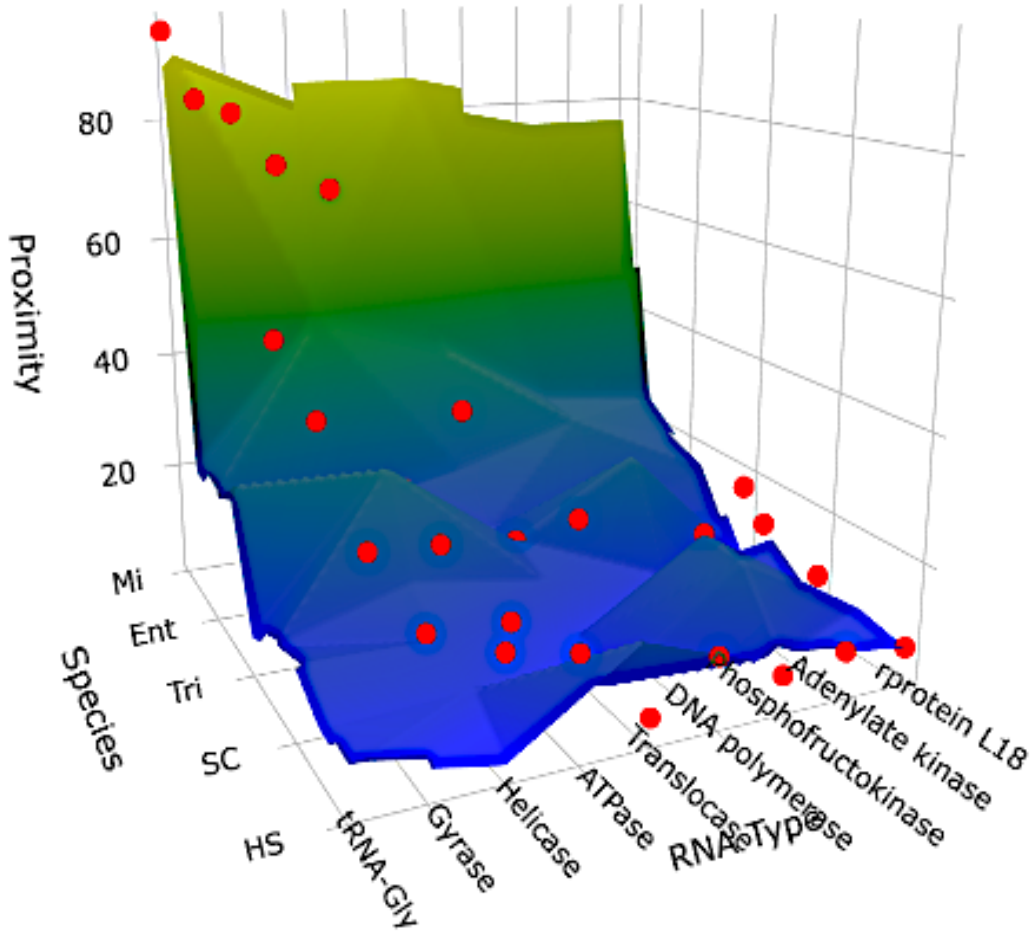

Figure 8. Surface representing the data of Table 1.

4. Conclusions

To support a network view of the origin of life, as discussed in 2018 by Fontecilla-Camps [85], Aguirre at al. [86] and Seligmann and Raoult [87], the AL ring can be proposed as a key in the primitive machinery building peptides (Figure 7). In this view, the boundary of this primordial functional « machine » able to build the first proteins, could be defined as a peptide gradient boundary, centered on the "proto-nucleus" AL. The amino acids confinement around AL could indeed favor the occurrence of peptide bounds, the machine functioning as a "proto-ribosome" into a "proto-membrane", close to a "proto-cell" with a network organization, each elements favoring the survival of the others. This approach stands as a solution of a variational problem in that peptide synthesis favored by AL was necessary to repair the proto-cell membrane made of hydrophobic peptides and lipids, which reciprocally ensured the integrity of the proto-nucleus, and so-protected it against denaturation. This mechanism is supported since one century by different works, theoretical as well as experimental: for example, in 1926, H.J. Muller already suggested that life began not as an enzyme, but as a gene [88]. The four amino acids: glycine, aspartic acid, asparagine, and serine have been claimed to have been coded by the first four triplets of the early, evolving genetic code [8,9], constituting the first class of amino acids selected following the min-max principle: "mean mutation error M equals information I", which uses the notion of information as proposed by Eigen [89]. In the theory of autopoiesis [90,91], the first living system is self-reproducing [92,93] and "continuously generates and specifies its own organization through its operation as a system of production of its own components, and does this in an endless turnover of components". Statistical

 doi:10.20944/preprints202504.0857.v1

and theoretical arguments about the role of the primitive RNAs in the progressive constitution of the genetic code [94–101].

As a singular prototype, this AL sequence should be useful to assess as a model matrix of future applications, ranging from synthetic biology used for producing proteins [102] to DNA computing [103]. As shown in this paper, the sequence AL and pentamers extracted from AL are indeed frequently retrieved as remnants in many genomes, notably in proteins essential for the protein translation and maintenance of the cell integrity (tRNA synthetases, RNA polymerases, tRNA nucleotidyl-transferases, lipids synthetases, CRISPR-Cas 9, etc.), which are considered as essential building blocks for cell survival.

Further studies should anew experimentally investigate the ring AL as a potential catalyzer of peptide synthesis [39] and search for its role in building protein and cell worlds after RNA world, and its role in consolidating the genetic code, in accordance with all the reference works establishing the present knowledge of the field, notably those concerning the evolution of the genetic code and of ancient ribo-nucleo-proteic structures like the ribosome or RNAs like the ribozymes [104–135].

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. S1: Supplementary material Biology 1, S2: Supplementary material Biology 2 and S3: Supplementary material Biology 3.

Author Contributions: Conceptualization, investigation and redaction, J.D.

Acknowledgments: We are indebted, for many helpful discussions on the existence of ancestral RNAs and of a proto-ribosome, to our masters and colleagues J. Besson (†), L. Demetrius, A. Moreira, J.D. Murray, V. Norris, H. Seligmann and R. Thom (†).

Conflicts of Interest: The author has no conflict of interest.

References

1. Paecht-Horowitz, M.; Berger, J.; Katchalsky, A. Prebiotic Synthesis of Polypeptides by Heterogeneous Polycondensation of Amino-acid Adenylates. *Nature* **1970**, *228*, 636-639.
2. Paecht-Horowitz, M.; Katchalsky, A. Synthesis of amino acyl-adenylates under prebiotic conditions. *J. Mol. Evol*. **1973**, *2*, 91-98.
3. Brack A. Polymerisation en phase aqueuse d'acides aminés sur des argiles. *Clay Minerals* **1976**, *11*, 117-120.
4. Crick, F. H. C., Brenner, S., Klug, A. & Pieczenik, G. A speculation on the origin of protein synthesis. *Orig. Life*, **1976**, *7*, 389–397.
5. Noller, H. F. Evolution of protein synthesis from an RNA world. *Cold Spring Harb. Perspect. Biol*. **2012**, *4*, a003681.
6. Jash, B.; Tremmel, P.; Jovanovic, D.; Richert, C. Single nucleotide translation without ribosomes. *Nat. Chem.* **2021**, *13*, 751–757.
7. Ishida, T. Simulation of the emergence of cell-like morphologies with evolutionary potential based on virtual molecular interactions. *Sci. Rep*. **2024**, *14*, 2086.
8. Eigen, M.; Schuster, P. The hypercycle: A principle of natural self-organization. Part C: The realistic hypercycle. *Naturwissenschaften* 1978, *65*, 341–369.
9. Eigen, M.; Winkler-Oswatitsch, R. Transfer-RNA: The early adaptor. *Naturwissenschaften* **1981**, *68*, 217–228.
10. Bernal, D. *The Physical Basis of Life*; RoutIedge and Kegan Paul: London, UK, 1951.
11. Miller, S.L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **1953**, *117*, 528–529.
12. Parker, E.T.; Cleaves, H.J.; Dworkin, J.P.; Glavin, D.P.; Callahan, M.; Aubrey, A.; Lazcano, A.; Bada, J.L. Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5526–5531.
13. Oró, J.; Kimball, A.P. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Arch. Biochem. Biophys.* **1961**, *94*, 217–227.

 doi:10.20944/preprints202504.0857.v1

14. Ferus, M.; Pietrucci, F.; Saitta, A.M.; Knížek, A.; Kubelík, P.; Ivanek, O.; Shestivska, V.; Civiš, S. Formation of nucleobases in a Miller–Urey reducing atmosphere. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4306–4311.
15. Ponnamperuma, C.; Sagan, C.; Mariner, R. Synthesis of adenosine triphosphate under possible primitive earth conditions. *Nature* **1963**, *199*, 222–226.
16. Hobish, M.K.; Wickramasinghe, N.S.; Ponnamperuma, C. Direct interaction between amino acids and nucleotides as a possible physicochemical basis for the origin of the genetic code. *Adv. Space Res.* **1995**, *15*, 365-382.
17. Caetano-Anolles, G.; Kim, K.M. The Origin and Evolution of the Archaeal Domain. Hindawi Publishing Corporation: London, 2014.
18. Di Giulio, M. On the origin of protein synthesis: A speculative model based on hairpin RNA structures. *J. Theor. Biol.* **1994**, *171*, 303–308.
19. Woese, C.R. A New Biology for a New Century. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 173–186.
20. Shapiro, R. Small Molecule Interactions were Central to the Origin of Life. *Q. Rev. Biol*. **2006**, *81*, 105-126.
21. Bernhardt, H.S. The RNA world hypothesis: The worst theory of the early evolution of life (except for all the others). *Biol. Direct* **2012**, *7*, 23.
22. Colín-García, M. Hydrothermal vents and prebiotic chemistry: a review. *Boletín de la Sociedad Geológica Mexicana* **2016**, *68*, 599-620.
23. Yarus, M. Life from an RNA world: the ancestor within. Harvard University Press: Cambridge Mass., 2010.
24. Yarus, M. Eighty routes to a ribonucleotide world; dispersion and stringency in the decisive selection. *RNA* **2018**, *24*, 1041-1055.
25. Yarus, M. On an RNA-membrane protogenome. *ArXiv* **2025**, 2502.00647
26. Lancet, D.; Zidovetzki, R.; Markovitch, O. Systems protobiology: Origin of life in lipid catalytic networks. *J. R. Soc. Interface* **2018**, *15*, 20180159.
27. Raine, D.J.; Norris, V. Lipid domain boundaries as prebiotic catalysts of peptide bond formation. *J. Theor. Biol*. **2007,** *246*, 176-185.
28. Kahana, A.; Lancet, D. Protobiotic Systems Chemistry Analyzed by Molecular Dynamics. *Life* **2019**, *9*, 38.
29. Caforio, A.; Driessen, A.J.M. Archaeal phospho-lipids: Structural properties and biosynthesis. *BBA-Mol. Cell Biol. Lipids* **2016**, *1862*, 1325-1339.
30. Demongeot, J. *Au sujet de quelques modèles stochastiques appliqués à la biologie*. PhD thesis, Université Joseph Fourier: Grenoble, 1975 (https://tel.archives-ouvertes.fr/tel-00286222).
31. Demongeot, J. Sur la possibilité de considérer le code génétique comme un code à enchaînement. *Revue de Biomaths* **1978,** *62*, 61-66.
32. Demongeot, J.; Besson, J. Code génétique et codes à enchaînement I. *C.R. Acad. Sc. III* **1983**, *296*, 807-810.
33. Demongeot, J.; Besson, J. Genetic code and cyclic codes II. *C.R. Acad. Sc. III* **1996,** *319*, 520-528.
34. Weil, G.; Heus, K.; Faraut, T.; Demongeot, J. An archetypal basic code for the primitive genome. *Theoret. Comp. Sc***. 2004**, *322*, 313-334.
35. Demongeot, J.; Elena, A.; Weil, G. Potential-Hamiltonian decomposition of cellular automata. Application to degeneracy of genetic code and cyclic codes III. *C. R. Acad. Sc. Biologies* **2006**,, *329*, 953-962.
36. Demongeot, J. *Primitive genome and RNA relics*. In: EMBC' 07. IEEE Proceedings: Piscataway, 6338-42, 2007.
37. Demongeot, J.; Moreira, A. A circular RNA at the origin of life. *J. Theor. Biol.* **2007**, *249*, 314-324.
38. Demongeot, J.; Drouet, E.; Moreira, A.; Rechoum, Y.; Sené, S. Micro-RNAs: viral genome and robustness of the genes expression in host. *Phil. Trans. Royal Soc. A*. **2009**, *367*, 4941-4965.
39. Demongeot, J.; Glade, N.; Moreira, A.; Vial, L. RNA relics and origin of life. *Int. J. Mol. Sci.* **2009**, *10*, 3420-3441.
40. Demongeot, J.; Hazgui, H.; Bandiera, S.; Cohen, O.; Henrion-Caude, A. MitomiRs, ChloromiRs and general modelling of the microRNA inhibition. *Acta Biotheoretica* **2013**, *61*, 367-383.
41. Demongeot, J. "Protoribosome" as new game of life. BioRxiv 2017 (https://dx.doi.org/10.1101/182055).
42. Demongeot, J.; Hazgui, H. The Poitiers school of mathematical and theoretical biology: Besson-Gavaudan-Schützenberger's conjectures on genetic code and RNA structures. *Acta Biotheoretica* **2016**, *64*, 403-426.
43. Demongeot, J.; Norris, V. Emergence of a "Cyclosome" in a Primitive Network Capable of Building "Infinite" Proteins. *Life (Basel)* **2019**, *9*, 51.

 doi:10.20944/preprints202504.0857.v1

44. Demongeot, J.; Seligmann, H. Theoretical minimal RNA rings recapitulate the order of the genetic code's codon-amino acid assignments. *J. Theor. Biol*. **2019**, *471*, 108–116.
45. Demongeot, J.; Seligmann, H. Spontaneous evolution of circular codes in theoretical minimal RNA rings. *Gene* **2019**, *705*, 95–102.
46. Demongeot, J.; Seligmann, H. More pieces of ancient than recent theoretical minimal proto-tRNA-like RNA rings in genes coding for tRNA synthetases. *J. Mol. Evol.* **2019**, *87*, 152-174.
47. Demongeot, J.; Seligmann, H. Bias for 3'-dominant codon directional asymmetry in theoretical minimal RNA rings. *J. Comput. Biol.* **2019**, *26*, 1003-1012.
48. Demongeot, J.; Seligmann, H. Theoretical minimal RNA rings designed according to coding constraints mimick deamination gradients. *Sci. Nat./Nat.* **2019**, *106*, 44.
49. Demongeot, J.; Seligmann, H. Pentamers with non-redundant frames: Bias for natural circular code codons. *J. Mol. Evol.* **2020**, *88*, 194–201.
50. Demongeot, J.; Seligmann, H. The primordial tRNA acceptor stem code from theoretical minimal RNA ring clusters. *BMC Genet.* **2020**, *21*, 7.
51. Demongeot, J.; Seligmann, H. Accretion history of large ribosomal subunits deduced from theoretical minimal RNA rings is congruent with histories derived from phylogenetic and structural methods. *Gene* **2020**, *738*, 144436.
52. Demongeot, J.; Seligmann, H. Deamination gradients within codons after 1<->2 position swap predict amino acid hydrophobicity and parallel β-sheet conformational preference. *Biosystems* **2020**, *192*, 104116.
53. Demongeot, J.; Seligmann, H. Theoretical minimal RNA rings maximizing coding information overwhelmingly start with the universal initiation codon AUG. *BioEssays* **2020**, *42*, 1900201.
54. Demongeot, J.; Henrion-Caude, A. The old and the new on the prebiotic conditions of the origin of life. *Biology (Basel)* **2020**, *9*, 88.
55. Demongeot, J.; Seligmann, H. Theoretical minimal RNA rings mimick molecular evolution before tRNA-mediated translation: codon-amino acid affinities increase from early to late RNA rings. *C. R. Acad. Sci. Biologies* **2020**, *343*, 111-122.
56. Demongeot, J.; Seligmann, H. Evolution of tRNA subelement accretion from small and large ribosomal RNAs. *Biosystems* **2022**, *193*, 104796.
57. Norris, V.; Demongeot, J. The Ring World hypothesis: the eversion of small, double-stranded polynucleotide circlets was at the origin of the double helix of DNA, the polymerisation of RNA and DNA, the triplet code, the twenty or so biological amino acids, and strand asymmetry. *Int. J. Mol. Sci.* **2022**, *23*, 12915.
58. Demongeot, J.; Thellier, M. Primitive oligomeric RNAs at the origins of life on Earth. *Int. J. Mol. Sci.* **2023**, *24*, 2274.
59. Demongeot, J.; Waku, J.; Cohen, O. Combinatorial and frequency properties of the ribosome ancestors. *AIMS MBE* **2023**, *21*, 884-902.
60. Ben Khalfallah, H.; Jelassi, M.; Rissaoui, H.; Barchouchi, M.; Baraille, C.; Gardes, J.; Demongeot, J. Information Gradient among Nucleotide Sequences of Essential RNAs from an Evolutionary Perspective. *Int. J. Mol. Sci*. **2024**, *25*, 7521.
61. Ben Khalfallah, H.; Jelassi, M.; Rachdi, M.; Demongeot, J. The AL-Codon-Counter Program: An Advanced Tool for Pentamer Analysis in RNA Sequences and Evolutionary Insights. In: SAI Computing Conference 2025, Lecture Notes in Networks & Systems, Springer Nature, New York (2025).
62. Staley, J.T. Domain Cell Theory supports the independent evolution of the Eukarya, Bacteria and Archaea and the Nuclear Compartment Commonality hypothesis. *Open Biol*. **2017**, *7*, 170041.
63. Li, S.; Yang, J. System analysis of synonymous codon usage biases in archaeal virus genomes. *J. Theor. Biol.* **2014**, *355*, 128-39.
64. Bahiri-Elitzur, S.; Tuller, T. Codon-based indices for modeling gene expression and transcript evolution. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2646-2663.
65. GtRNAdb. Available online: http://gtrnadb.ucsc.edu/(accessed on 22/02/2025).
66. NCBI Nucleotide. Available on line: https://www.ncbi.nlm.nih.gov/nucleotide (accessed on 22/02/2025).
67. MiRBase. Available online: http://www.mirbase.org/(accessed on 22/02/2025).

 doi:10.20944/preprints202504.0857.v1

68. Georg, R.C.; Stefani, R.M.; Gomes, S.L. Environmental stresses inhibit splicing in the aquatic fungus Blastocladiella emersonii. *BMC Microbiol*. **2009**, *9*, 231.
69. Rogozin, I.B.; Carmel, L.; Csuros, M.; Koonin, E.V. Origin and evolution of spliceosomal introns. *Biol. Direct*. **2012**, *7*, 11.
70. Brochier-Armanet, C.; Forterre, P.; Gribaldo, S. Phylogeny and evolution of the Archaea: One hundred genomes later. *Curr. Opin. Microbiol.* **2011**, *14*, 274–281.
71. Forterre, P. The Common Ancestor of Archaea and Eukarya Was Not an Archaeon. *Archaea* **2013**, *2013*, 372396.
72. Staley, J.T. Domain Cell Theory supports the independent evolution of the Eukarya, Bacteria and Archaea and the Nuclear Compartment Commonality hypothesis. *Open Biol.* **2017**, *7*, 170041.
73. Slonimski, P.P. Periodic oscillations of the genomic nucleotide sequences disclose major differences in the way of constructing homologous proteins from different procaryotic species. *Comptes Rendus Biologies* **2007**, *330*, 13-32.
74. Yarus M. The meaning of a minuscule ribozyme. *Phil. Trans R Soc Lond B Biol Sci*. **2011**, *366*, 2902-2909.
75. Tamura, K.; Schimmel, P. Oligonucleotide-directed peptide synthesis in a ribosome- and ribozyme-free system. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1393-1397.
76. Tamura, K.; Schimmel, P. Peptide synthesis with a template-like RNA guide and aminoacyl phosphate adaptors. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8666-8669.
77. Tamura, K.; Schimmel, P. Chiral-selective aminoacylation of an RNA minihelix. *Science* **2004**, 305, 1253.
78. Tamura, K.; Schimmel, P. Chiral-selective aminoacylation of an RNA minihelix: Mechanistic features and chiral suppression. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13750-13752.
79. Kim, N.; Kim, H.K.; Lee, S.; Seo, J.H; Choi, J.W.; Park, J.; Min, S.; Yoon, S.; Cho, S.; Kim, H.H. Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nature Biotechnology* **2020**, *38*, 1328-1336.
80. Miller, S.M.; Wang, T.; Randolph, P.B.; Arbab, M.; Shen, M.W.; Huang, T.P.; Matuszek, Z.; Newby, G.A.; Rees, H.A.; Liu, D.R. Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nature Biotechnology* **2020**, *38*, 471–481.
81. Petersen, J.; Brinkmann, H.; Pradella, S. Diversity and evolution of *repABC* type plasmids in *Rhodobacterales*. *Environ. Microbiol.* **2009**, *11*, 2627-2638.
82. Trifonov, E.N.; Bettecken, T. Sequence fossils, triplet expansion, and reconstruction of earliest codons. *Gene* **1997**, *205*, 1–6.
83. Trifonov, E.N. Consensus temporal order of amino acids and evolution of the triplet code. *Gene* **2000**, *261*, 139-151.
84. Sobolevsky, Y.; Guimarães, R.C.; Trifonov, E.N. Towards functional repertoire of the earliest proteins. *J. Biomol. Structure and Dynamics* **2013**, 31, 1293–1300.
85. Fontecilla-Camps, J. Geochemical Continuity and Catalyst/Cofactor Replacement in the Emergence and Evolution of Life. *Angew. Chem.* **2018**, *130*, 08438.
86. Aguirre, J.; Catalan, P.; Cuesta, J.A.; Manrubia, S. On the networked architecture of genotype spaces and its critical effects on molecular evolution. *Open Biol.* **2018**, *8*, 180069.
87. Seligmann, H.; Raoult, D. Stem-Loop RNA Hairpins in Giant Viruses: Invading rRNA-Like Repeats and a Template Free RNA. *Front. Microbiol.* **2018**, *9*, 101.
88. Muller, H.J. The gene as the basis of life. In Proceedings of the International Congress of Plant Sciences Ithaca 1926; Duggar, B.M., Ed.; Menasha: Banta, Wsconsin, 1929; pp. 897–921.
89. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, *58*, 465–523.
90. Maturana, H.R.; Varela, F.J. *Autopoiesis and Cognition: The Realization of the Living*; Reidel: Boston, MA, USA, 1980.
91. Bourgine, P.; Stewart, J. Autopoiesis and cognition. *Artif. Life* **2004**, *10*, 327–345.
92. Ono, N.; Ikegami, T. Self-maintenance and self-reproduction in an abstract cell model. *J. Theor. Biol*. **2000**, *206*, 243–253.

 doi:10.20944/preprints202504.0857.v1

93. Ono, N.; Ikegami, T. Artificial chemistry: Computational studies on the emergence of self-reproducing units. In *Proceedings of the 6th European conference on artificial life (ECAL'01), Prague, Czech Republic, September 2001*; Kelemen, J., Sosik, S., Eds.; Springer: Berlin, Germany, 2001; pp. 186–195.
94. Noble, D. Genes are not the blueprint for life. *Nature* **2019**, *626*, 254–255.
95. Noble, D.; Joyner, M. The physiology of evolution. *J. Physiol.* **2024**, *602*, 2361-2365.
96. Dufton, M.J. Genetic code synonym quotas and amino acid complexity: Cutting the cost of proteins? *J. Theor. Biol.* **1997**, *187*, 165–173.
97. Davis, B.K. Evolution of the genetic code. *Prog. Biophy. Mol. Biol.* **1999**, *72*, 157–243.
98. Wong, J.T.F. Coevolution theory of the genetic code at age thirty. *Bioessays* **2005**, *27*, 416–425.
99. Wong, J.T.F.; Ng, S.K.; Mat, W.K.; Hu, T.; Xue, H. Coevolution theory of the genetic code at age forty: Pathway to translation and synthetic life. *Life (Basel)* **2016**, *6*, 12.
100. Takeuchi, N.; Kaneko, K. The origin of the central dogma through conflicting multilevel selection. *Proc. R. Soc. B* **2019**, *286*, 20191359.
101. Fried, S.D.; Fujishima, K.; Makarov, M.; Cherepashuk, I.; Hlouchova, K. Peptides before and during the nucleotide world: an origins story emphasizing cooperation between proteins and nucleic acids. *J. R. Soc. Interface* **2022**, *19*, 20210641.
102. Chen R.; Wang, S.K.; Belk, J.A.; Amaya, L.; Li, Z.; Cardenas, A.; Abe, B.T.; Chen, C.K.; Wender, P.A.; Chang, H.Y. Engineering circular RNA for enhanced protein production. *Nat. Biotechnol.* **2023**, *41*, 262-272.
103. Levy, N.; Schabanel, N. ENSnano: A 3D Modeling Software for DNA Nanostructures. *DNA* **2021**, *5*, 1-5.
104. Michaud, M.; Cognat, V.; Duchêne, A.M.; Maréchal-Drouard, L. A global picture of tRNA genes in plant genomes. *Plant J.* **2011**, *66*, 80–93
105. Fonville, N.C.; Velmurugan, K.R.; Tae, H.; Vaksman, Z.; McIver, L.J.; Garner, H.R. Genomic leftovers: Identifying novel microsatellites, over-represented motifs and functional elements in the human genome. *Sci. Rep.* **2016**, *6*, 27722.
106. Fujishima, K.; Sugahara, J.; Tomita, M.; Kanai, A. Sequence Evidence in the Archaeal Genomes that tRNAs Emerged Through the Combination of Ancestral Genes as 59 and 39 tRNA Halves. *PLoS ONE* **2008**, *3*, e1622.
107. Spang, A.; Caceres, E.F.; Ettema, T.J.G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **2017**, *357*, eaaf3883.
108. Eme, L.; Spang, A.; Lombard, J.; Stairs, C.W.; Ettema, T.J.G. Archaea and the origin of eukaryotes. *Nature* **2017**, *15*, 711–723.
109. Legendre, M.; Fabre, E.; Poirot, O.; Jeudy, S.; Lartigue, A.; Alempic, J.M.; Beucher, L.; Philippe, N.; Bertaux, L.; Christo-Foroux, E.; Labadie, K.; Couté, Y.; Abergel, C.; Claverie, J.M. Diversity and evolution of the emerging Pandoraviridae family. *Nat. Commun.* **2018**, *9*, 2285.
110. Case, R.J.; Boucher, Y.; Dahllöf, I.; Holmström, C.; Doolittle, W.F.; Kjelleberg, S. Use of 16S rRNA and rpoB Genes as Molecular Markers for Microbial Ecology Studies. *Appl. Environ. Microbiol.* **2010**, *73*, 278–288.
111. Bartnik, E.; Borsuk, P. A glycine tRNA gene from lupine mitochondria. *Nucleic Acids Res.* **1986**, *14*, 2407.
112. Schlüter, K.; Fütterer, J.; Potrykus, I. Horizontal Gene Transfer from a Transgenic Potato Line to a Bacterial Pathogen (*Erwinia chrysanthemi*) Occurs-if at All-at an Extremely Low Frequency. *Biotechnology* **1995**, *13*, 1094–1098.
113. Becker, E.A.; Seitzer, P.M.; Tritt, A.; Larsen, D.; Krusor, M.; Yao, A.I.; Wu, D.; Madern, D.; Eisen, J.A.; Darling, A.E.; et al. Phylogenetically Driven Sequencing of Extremely Halophilic Archaea Reveals Strategies for Static and Dynamic Osmo-response. *PLoS Genet.* **2014**, *10*, e1004784.
114. Abrahão, J.; Silva, L.; Silva, L.S.; Khalil, J.Y.B.; Rodrigues, R.; Arantes, T.; Assis, F.; Boratto, P.; Andrade, M.; Kroon, E.G.; et al. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat. Commun.* **2018**, *9*, 749.
115. Buzayan, J.M.; Hampel, A.; Bruening, G. Nucleotide sequence and newly formed phosphodiester bond of spontaneously ligated satellite tobacco ringspot virus RNA. *Nucleic Acids Res.* **1986**, *14*, 9729–9743.
116. Salter, J.; Krucinska, J.;Alam, S.; Grum-Tokars, V.; Wedekind, J.E. Water in the Active Site of an All-RNA Hairpin Ribozyme and Effects of Gua8 Base Variants on the Geometry of Phosphoryl Transfer. *Biochemistry* **2006**, *45*, 686–700.

 doi:10.20944/preprints202504.0857.v1

117. Pérez-Ruiz, M.; Barroso-delJesus, A.; Berzal-Herranz, A. Specificity of the Hairpin Ribozyme. *J. Biol. Chem.* **1999**, *274*, 29376–29380.
118. Müller, U.F. Design and Experimental Evolution of trans-Splicing Group I Intron Ribozymes. *Molecules* **2017**, *22*, 75.
119. Paul, N.; Joyce, G.F. A self-replicating ligase ribozyme. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12733–12740.
120. Perreault, J.; Weinberg, Z.; Roth, A.; Popescu, O.; Chartrand, P.; Ferbeyre, G.; Breaker, R.R. Identification of Hammerhead Ribozymes in All Domains of Life Reveals Novel Structural Variations. *PLoS Comput. Biol.* **2011**, *7*, e1002031.
121. Hammann, C.; Luptak, A.; Perreault, J.; De La Peña, M. The ubiquitous hammerhead ribozyme. *RNA* **2012**, *18*, 871–885.
122. Harris, K.A.; Lünse, C.E.; Li, S.; Brewer, K.I.; Breaker, R.R. Biochemical analysis of hatchet self-cleaving ribozymes. *RNA* **2015**, *21*, 1–7.
123. Agmon, I.C. Could a Proto-Ribosome Emerge Spontaneously in the PrebioticWorld? *Molecules* **2016**, *21*, 1701.
124. Arquès, D.G.; Michel, C.J. A complementary circular code in the protein coding genes. *J. Theor. Biol.* **1996**, *182*, 45–58.
125. Dila, G.; Ripp, R.; Mayer, C.; Poch, O.; Michel, C.J.; Thompson, J.D. Circular code motifs in the ribosome: A missing link in the evolution of translation? *RNA* **2019**, *25*, 1714–1730.
126. Kim, Y.; Opron, K.; Burton, Z.F. A tRNA- and Anticodon-Centric View of the Evolution of Aminoacyl-tRNA Synthetases, tRNAomes, and the Genetic Code. *Life (Basel)* **2019**, *9*, 37.
127. Kunnev, D.; Gospodinov, A. Possible Emergence of Sequence Specific RNA Aminoacylation via Peptide Intermediary to Initiate Darwinian Evolution and Code Through Origin of Life. *Life (Basel)* **2018**, *8*, 44.
128. Seligmann, H. Protein Sequences Recapitulate Genetic Code Evolution. *Computational and Structural Biotechnology Journal* **2018**, *16*, 177–189.
129. Zaia, D.A.; Zaia, C.T.; De Santana, H. Which amino acids should be used in prebiotic chemistry studies? *Orig. Life Evol. Biosph.* **2008**, *38*, 469–488.
130. Robinson, R. Jump-starting a cellular world: Investigating the origin of life, from soup to networks. *PLoS Biol.* **2005**, *3*, e396.
131. Beringer, M.; Rodnina, M.V. Importance of tRNA interactions with 23S rRNA for peptide bond formation on the ribosome: Studies with substrate analogs. *Biol. Chem.* **2007**, *388*, 687–691.
132. Koonin, E.V.; Novozhilov, A.S. Origin and evolution of the genetic code: The universal enigma. *Iubmb Life* **2009**, *61*, 99–111.
133. Rodin, A.S.; Szathmáry, E.; Rodin, S.N. On origin of genetic code and tRNA before translation. *Biol. Direct.* **2011**, *6*, 14.
134. Koonin, E.V. Frozen Accident Pushing 50: Stereochemistry, Expansion, and Chance in the Evolution of the Genetic Code. *Life (Basel)* **2017**, *7*, 22.
135. Gonzalez, D.L.; Giannerini, S.; Rosa, R. On the origin of degeneracy in the genetic code. *Interface Focus* **2019**, *9*, 20190038.