Article

Traces of a Primitive RNA Ring in Current Genomes

Jacques Demongeot

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

Simple Summary: The search for RNA molecules at the origin of life is a major challenge for understanding the primordial stages of evolution. Here, we propose a small RNA that could have served as a catalyst for the formation of the first peptides and that has left numerous traces in current genomes.

Abstract: (1) Background: Previous theoretical studies have provided arguments for the existence of a circular 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 primitive RNA in terms of the occurrence of similar nucleotide motifs. (2) Methods: In searching for the smallest possible 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 (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 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 that is 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 ALpha or Archetypal Loop), in which we discovered 18 biological properties concerning its fit with the sequences and motifs of current genomes. 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", 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 simple RNA structures (like internal and bulge loops) 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]. Section 2 will present the materials and methods, followed by Section 3 with the results obtained and the discussion. The final section will be devoted to the conclusion and outlook.

2. Materials and Methods

2.1. Theoretical Criteria

Four theoretical criteria for a primordial RNA called AL (for ALpha or Archetypal Loop) to be a candidate for primordial catalysis of peptide biosynthesis have already been identified as optimal combinatorial properties and published [30–60]. The concept of ring for the structure of AL can be considered as a sort of circular consensus capable of embedding all possible genetic encodings, whose properties 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 its circular form (the theoretical 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.

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

The AL-Codon-Counter algorithm detects certain motifs in any RNA sequence that are 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 (see Figure 1).

In Figure 1, the process of calculation 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 analysis is extended to the study of evolutionary trends, particularly hypothetical remnants of the ancestral AL.

The program's statistical framework allows for comparison of pentamer distributions across multiple species, potentially revealing the evolutionary conservation of identified motifs. 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. This also explains 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, and 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 this 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 (i.e., the number of the possible pentamers in the sequence) and $p = 9/1024$ (the probability of observing all 9 pentamers by chance). 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 is directly related to the probability P to observe n_o pentamers by chance. If we use the Gaussian approximation of the binomial distribution, for example, $P_{PAL} = 8$ corresponds to $P < 5 \times 10^{-5}$. More generally, P is obtained thanks to Gaussian Tables or calculated using the classical approximations of the Gaussian repartition function.

Figure 1. Screenshot 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.

Another proximity called P_{PAL} Doublet can be calculated by counting the number of pairs (or doublets) of successive codons in a given mRNA sequence. If the occurrence of a bond between two amino acids is due to their reversible weak link to codons close on AL acting as a proto-ribosome, this pair could occur with a significant frequency in the mRNA. To show this, the number observed along the mRNA of pairs of AL codons (such as ATT, CAA, GAT, GAA, CCA, AGA, ACT or AAT, 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}. Both will be calculated for species belonging to the three domains of life, Archaea, Bacteria, and Eukarya.

3. Results and Discussion

3.1. Presentation of AL

The discovery of the structure of AL occurred in four stages. In 1975, the 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 six hybridized nucleotide pairs (in red and blue, nucleotides in green being not hybridized):

$5'$-**GCCAT TCAG A A**
TGGTA TCAG TA

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

$5'$-**TGGTGAA GA C G**
ACCAT AA CT T C

In 1996, a third ring called AB (for Ancestral Basal) was discovered [33]. It has nine hybridized pairs, but they are not contiguous, and verified criteria (1) and (2):

$5'$-**GCCATTCAAG A**
TGGTAAGTAT C

In 2004 [34], it was established that criterion (1) had no solution for a cycle of length 20 or 21, but only for a length of 22, for which there were 29,520 solutions (out of the 4^{22} possible solutions) containing only one repeated codon AXN, 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 nine or more nucleotides, of which only 19 encompassed both a start and 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 and satisfying all criteria (1) to (4):

$5'$-**TGCCATTCAA**
G
CATGGTAAGTA

Then, the two configurations (circular and hairpin) of AL are summarized in Figure 2, where two hairpins appear inside the circular configuration, one corresponding to AL on the left and the other to the complement of AL (in the complementarity A/U, G/C), which also has a hairpin form on the right.

3.2. Structural Properties of AL

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. This is the case for the loops of the transfer RNA (D-loop, anticodon-loop, and T ψ-loop), which are well conserved during evolution, unlike the

branches of the clover leaf structure of the tRNAs, which are very variable. Examples are given for various species of the three domains of life with their full names given on Figure 3A in various colors: Archaea (mauve), Bacteria (green), Eukarya with Fungi (violet), Plants (red), Animals (blue), and short names (in the same order) with the sequences of their transfer RNA tRNA-Gly^{GCC} given on Figure 3B.
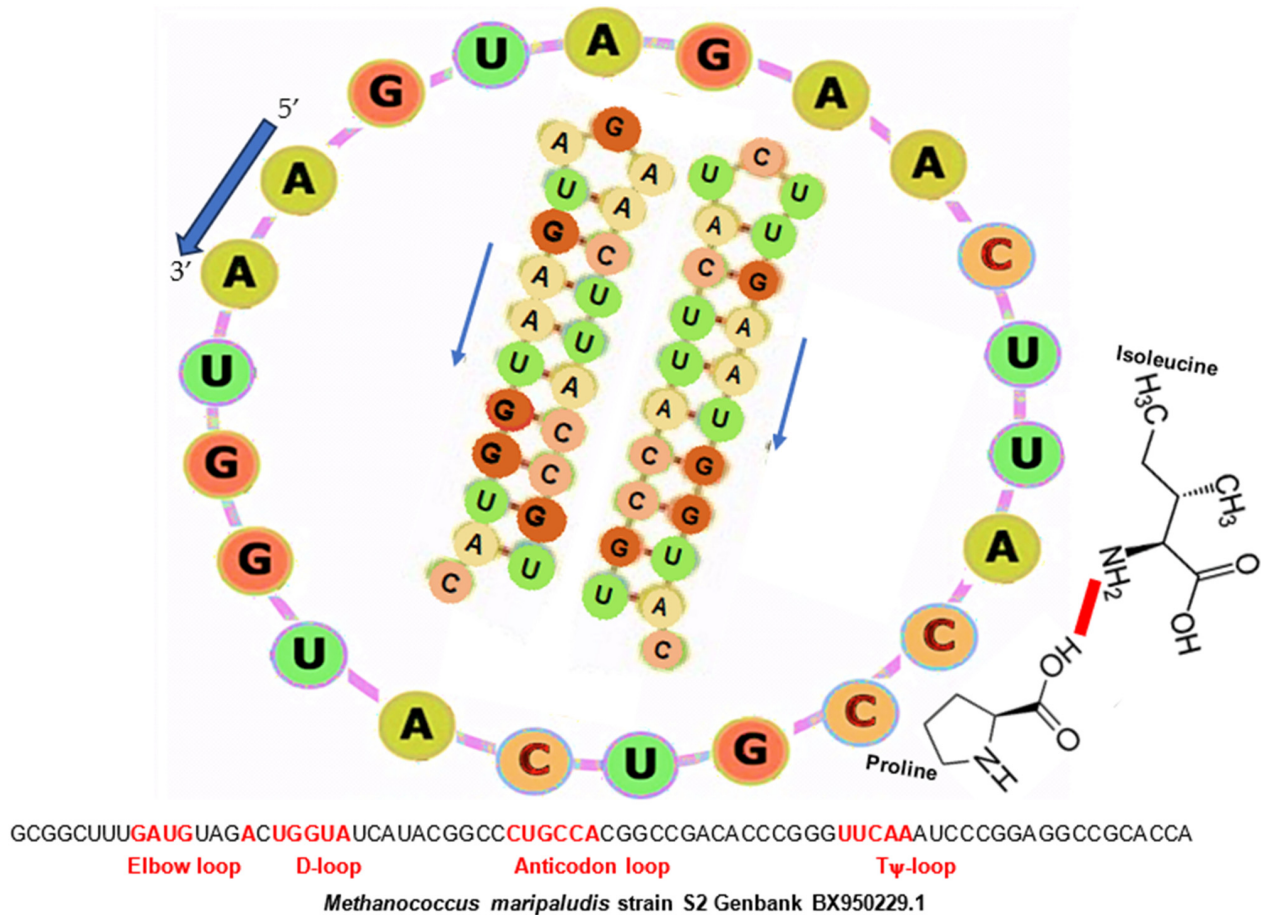

GCGGCUUU**GAUG**UAG**ACUGGUA**UCAUACGGCC**CUGCCA**CGGCCGACACCCGGG**UUCAA**AUCCCGGAGGCCGCACCA
Elbow loop **D-loop** **Anticodon loop** **Tψ-loop**
Methanococcus maripaludis strain S2 Genbank BX950229.1

Figure 2. Circular form of AL fitting the loops (in red) of the tRNA-Gly^{GCC} of *Methanococcus maripaludis* (bottom). Inside the ring, the hairpin configurations of AL (left) and anti-AL (right), and outside the ring, a dipeptide whose synthesis could be favored by the proximity to AL of the amino acids constituting the dipeptide. Blue arrows indicate the 5′-3′ sense.

The other structural properties of AL correspond to the following optimal characteristics:

(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 located in the loops (in red) of the tRNAs from the database GtRNAdb (Figure 4).

(3) Among the rings verifying criteria (1) to (4) m, AL 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 full and short names, along with the sequences of their tRNA-GlyGCC, are given in Figure 2 and the phylogenetic tree in Figure 5.

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

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

A

Salinarchaeum Harcht, Thermococcus, Thermofilum carboxyditrophus, Sulfobacillus acidophilus, Acidobacterium capsulatum, Blochmannia endosymbiont, Candidatus Carsonella ruddii, Erysipelothrix 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, Sphagnum fallax, Solanum lycopersicum, Sphagnum magellanicum, Spirogloea muscicola, Triticum aestivum, Vitis vinifera, Alligator mississippiensis, Balaenoptera acutorostrata scammoni, Bos taurus, Procavia capensis, Rhytidiadelphus loreus, Rattus norvegicus, Sorex araneus, Spodoptera frugiperda, Strongylocentrotus purpuratus, Taeniopygia guttata, Trichechus manatus, Tarsius syrichta, Tursiops truncatus, Vicugna pacos, Xenopus tropicalis.

B

Sali	GCGTCGGTAGTGTAGTGGTATCACGTGACCCTGCCACGGTCGCAaCCCGAGTTCAAATCTCGGCCGACGCA
Therm	GCGGTGGTAGTCTAGCCtGGTctAGGACAGCGGCCTGCCACGCCGCTGGCCCGGGTTCAAATCCCGGCCACCG
Therc	GCGGCCGTAGTCTAGTCtGGTAGGATGGCGGCCTGCCACGCCGCAGAaCCCGGGTTCAAATCCCGGCGGCCG
Sulf	GCCGTAGTCTAGCATGGAtaGGACGCCTGCCTGCCACGCAGGAGGtCCCGGGTTCAAATCCCGGCGGCCGCA
Acid	GCGGGAGTAGCTCAGTGGTAGAGCATCGCCTTGCCAAGGCGAGGGtCGCGGGTTCAAGTCCCGTCTCCCGCTCCA
Blocen	GCGGGAATAGCTCAGTTGGTAGAGTACAACCTTGCCAAGGTTGGGGtCGCGAGTTCAAGTCTCGTTTCCCGCT
Cand	GCGAAAGTATCTTAATGGTAAAGTATCACCTTGCCATGGTGAAAGtTGCGAGTTCAAGTCTCGTCTTTCGCT
Erys	GCAGGTGTAGTTCAATGGTAGAACACGACCTTGCCAAGGTTGAGGCGGGGGTTCAATTCCCCTCACCTGCTCCA
Faec	GCAGATGTAGTTCAATGGTAGAACACAGCCTTGCCAAGGCTGATaCGGGGGTTCAATTCCCCTCATCTGCTCCA
Myco	GCAAATGTAGTTCAATGGTAGAACACCAGCTTGCCATGCTGGATaCGGGGGTTCAATTCCCCTCATTTGCTCCA
Popo	GCGGGTGTAGTTCAATGGTAGAACGAAAGCTTGCCAAGCTTTAGGCGAGGGTTCGATTCCCTTCACCCGCTC
Ttama	GCGGAAATAGCTTAATGGTAGAGTATAGCCTTGCCAAGGCTGAGGtGAGGGTTCAAGTCCCTTTTTCCGCT
Spha	GCATTAATGGGGTAATGGTAACCTTAGTCGTTGCCATCGACTAGcCGCGAGTTCAATTCTTGCTTAGTGCA
Angebi	GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTGAGGTGAGGGTTCAAGTCCCTTCTTCCGCT
Athal	GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTAAGGtTGAGGGTTCAAGTCCCTCCTTCCGCT
Baubi	GCGGATATAGTCGAATGGTAAAATTTCTCTTTTGCCAAGGAGAAGAACGCGGGTTCGATTCCCGCTATCCGCC
Cbrau	GCGGAAATAGCTTAATGGTAGAGTATAGCCTTGCCAAGGCTAAGGtGAGGGTTCAAATCCCTTTTTCCGCT
Ccamp	GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTGAGGtGAGGGTTCAAGTC CCTCCTTCCGCT
Cpurp	GCGGAAATAGCTTAATGGTAGAGTATAGCCTTGCCAAGGCTGAGGTGAGGGTTCAAGTCCCTTTTTCCGCT
Graim	GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTGAGGtGAGGGTTCAAGTCCCTCCTTCCGCT
Ntaba	GCGGAAATAGATTAATGGTAGATCATAGCCTTGCCAAGGCTGAGGtGAGGGTTCAAGTCCCCTCCTTCCGCT
Oevill	GCGGAAATAGCTTAATGGTAGAGCGTAGCCTTGCCAAGGCTGAGGTTGAGGGTTCAAGTCCCTCCTTCCGCT
Sfall	GCGGAAATAACTTAATGGTAGAGTATAGCCTTGCCAAGGCTGAGGtGAGGGTTCAAGTCCCTTTTTTCGCT
Slyco	GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTAAGGtGAGGGTTCAAGTCCCTCCTTCCGCT
Smage	GCGGAAATAGCTTAATGGTAGAGTATAGCCTTGCCAAGGCTGAGGtGAGGGTTCAAGTCCCTTTTTCCGCT
Smusc	GCGGATGTAGCTCAATGGTAGAGTatgtgtATAGCCTTGCCAAGGCTGAGGtGAGGGTTCAAGTCCCGTTTTCAGCT
Taest	GCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCATGGTACAGaCCTGGGTTCAATTCCTGGCTGGTGGA
Vvini	GCGGAAATAGCTTAATGGTAGAGCATAGCCTTGCCAAGGCTGAGGtGAGGGTTCAAGTCCCTCCTTCCGCT
Amiss	GCATTGGTGGTTCAATGGTAGAATTCCTGCCTGCCATGCAGGAGaCCTGGGTTCAATTCCCGCCAATGCA
Bacut	TCCCTGGTGGCACAGTGGTTGAGAATCTGCCTGCCAgTGCAGGGGGaCACACATTCAAGCCCTGgTCCAGGAAG
Btaur	TCCTAGGTGGCTCAATGGTaAAAATACTTGCCTGCCAaAGCAGGAGaCTCAGGTTCAATCCCTGGTCCAGGAA
Pcape	GGACTAATGGCATAGTGGTgAAAGTGTTTGGCTGCCAACCAAAAGGtCAGTGATTCAAACCCACTagctGTTTT
Rlore	GCGGAAATAGCTTAATGGTAGAGTATAGCCTGCCAAGGCTGAGGtGAGGGTTCAAGTCCCTTTTTCCGCT
Rnorv	GaGGAgATGGCTCAATGGTtaAGAGCACTGACTGCCGtgctcTTCCAGAGGcCGTGAGTTCAAGTCTCAGTAACCA
Saran	GCTcGATTCAGTGGTAGAATTCTCGGCTGCCACGTGGGAGGCCCGGGTTCAATTCCCGGcccaTGCA
Sfrug	GCACGGTTGGCGCAGTGGCTGGGCAACCGCCTGCCGCGCCATGTGtGGTGGGTTCAATTCCCACACGGTGCA
Spurp	GCATCGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCGGGGGaCCCGGGTTCAATTCCCGGCCGATGCA
Tgutt	GCCCTGGTGGCTCCGTGGTAGAATTCTGCCTGCCACGGCGGCAgCCTGGGTTCAATTCCCGGCAGAGGCA
Tmana	TCCTGGGTAGTACAAATGGTAACGCACTCGGCTGCCAACCGAAAGGtTGGTGGTTCAAACCCACCCCCAGAAG
Tsyri	ACATGGGTAGTTCAGTGGTAGAATTCTCGCCTGCCACACAGGAGGCCCGGATTCAATTACTAACCCATGCA
Ttrun	GCATTGGTGATTCAGTGGTAGAATTCTTGCCTGCCACGTGGGAGGCCAGGGTTCAATTCCCAGCCAATGCA
Vpaco	GGAGGTATAGCTCAGTGGTAGAGTGCATGCCTGCCATGCACAAGGtCCTGGGTTCAATCCCCAGTACCTCCA
Xtrop	GCATTGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCGGGAGGCCCGGGTTCAATTCCCGGCCAATGCA

Figure 3. (**A**) List of species from the three domains of life: Archaea (mauve), Bacteria (green), Eukarya with Fungi (violet), Plants (red), and Animals (blue); (**B**) List of short names of the species in the same order as in (**A**) with the sequences of their transfer RNA, tRNA-Gly^{GCC}.

Figure 4. tRNA-Gly^{GCC} of *Arabis alpina* mitochondrion with loops in red [65,66].
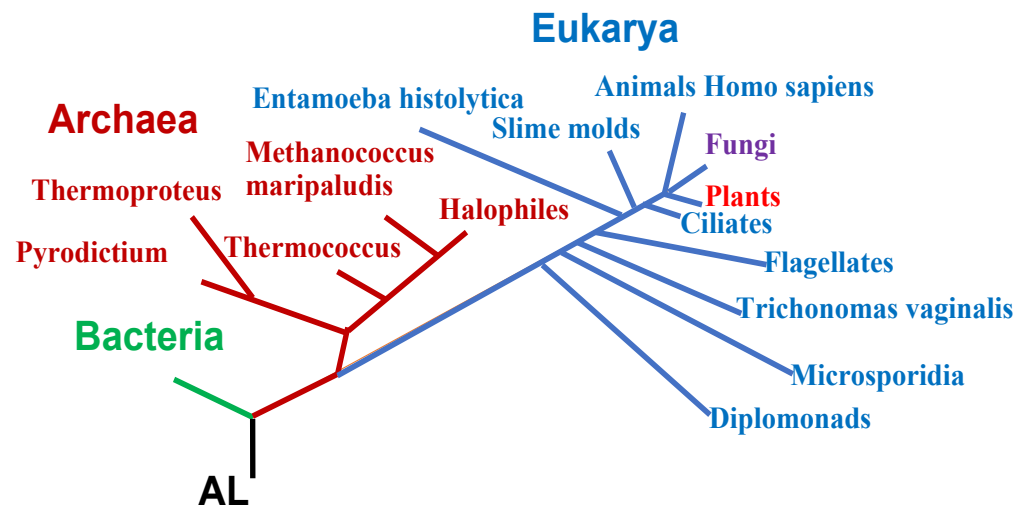
Figure 5. Phylogeny of the 3 domains of life—Archaea, Bacteria, and Eukarya—with indication of some species.

(6) The average edit distance from AL to 20,000 different randomized versions of randomly repeated microRNAs preserving length 22 and nucleotide composition of AL is significantly larger [37] than the average edit distance of AL to the real microRNAs from the database miRBase [67].

(7) AL has at least 15 common nucleotides with the barycenter of these 20,000 randomly repeated microRNAs of length 22 and with the same nucleotide composition as AL, 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 three domains of life, Archaea, Bacteria, and Eukarya (Figure 4 and [69–72]).

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

(11) 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.

(12) AL matches well with many non-coding genomes from viral origin [38].

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

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

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

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

(17) 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 S1), and complete genomes of *Methanococcus maripaludis* (Archaea), *Dojkabacteria bacterium* (Bacteria), *Clitoria ternatea* plasmid, and *Oenothera villaricae* chloroplast (Plants), and mitochondrion of *Jaculus jaculus* (mammal) have their proximity P_{PAL} Doublets in decreasing order in evolution, i.e., 312.5, 224.1, 93, 92.4, and 4.7 (cf. Supplementary Material S2).

(18) 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].

(19) Twelve hexadecameric peptide sequences of 16 amino acids from MVLPFKM-NGTAIQDEW 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 $4 \times 10^{-12} \pm 3 \times 10^{-6}$, 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*.
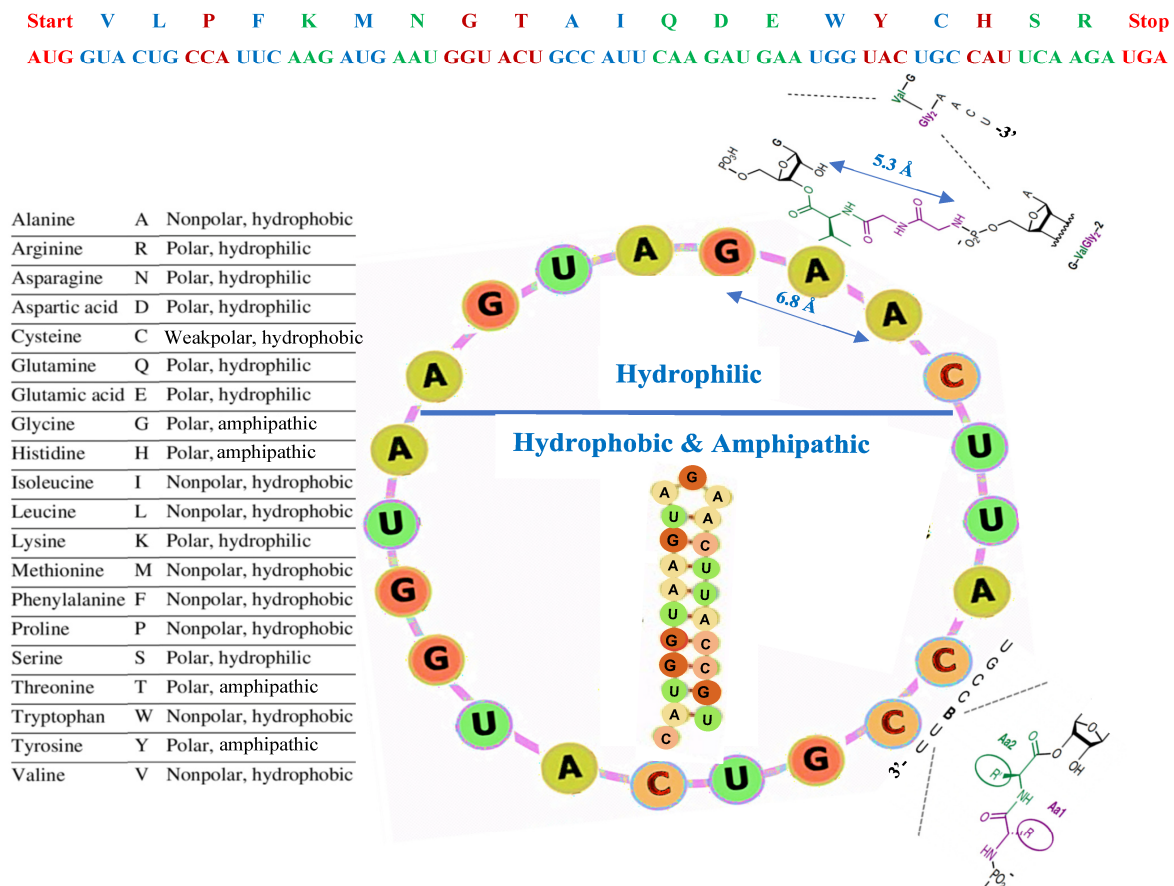
Figure 7. The evolutionary machinery. On top, the succession of codons without overlap and corresponding amino acids of AL. On the left, the amino acid polarities. On the right, AL in the catalytic function of its circular form, where the codons CCA for proline and AUU for isoleucine temporarily attract 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. The representations of RNA–amino acid complexes are adapted from [6].

3.3. Functional Properties of AL

The circular form of AL hybridizes with its complement anti-AL, which exhibits the same stable hairpin characteristics (identical to AL hairpin, except for the head), in equilibrium with a circular form capable of restoring AL by the same process (Figure 3). 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 with first by Tamura and Schimmel [75–78] and Yarus [25].

Traces of AL in current ribosomes can be found by considering the 5S rRNA of 10 bacteria (Figure 8 and [82]). The third quarter of the sequence is folded like a hairpin, and the tail of this hairpin is similar to a fragment of the tail part of the AL hairpin (in red in the table in Figure 8). Eigen [8,9] demonstrated that the early genetic code (PGC) consisted of nucleotides with the pattern RNY, where R represents purines (A/G), Y pyrimidines (C/U), and N any nucleotide (A/C/G/U), and that this RNY pattern was frequent in ribosomal RNA subunit 5S (5S rRNA) for more than 200 varied species [83]. On Figure 8, the 30 nt third quarter of the 5S rRNAs of 10 bacteria shows the RNY-encoded pattern, namely on the tail of its hairpin configuration (Figure 8A), which corresponds to the tail part of the AL hairpin, which has also codons with RNY patterns.

Code	Species name	5S 30 nt subsequence
>NR_075107.1	Aquifex aeolicus	GCAGTTAAGCCCCCCAGCGCCG**ATGTACTG**
>NR_103286.1	Bacillus subtilis	GAAGTTAAGCTCTTCAGCGCCG**ATGGTA**GTC
>pdb|1XBP|9	Deinococcus radiodurans	GTCGTGAAACACAGCAGCGCCA**ATGATACT**C
>MZ852211.1	Borreliella burgdorferi	TTTTATTTTTATGTTTTTTAAATAT**TGGT**GTTT
>NR_075284.1	Escherichia coli	GAAGTGAAACACAGTATCGCCG**ATGGTAGTG**
>NR_075332.1	Candidatus Pelagibacter	GAAGTCAAGCCCTTCAGCGCC**AATGGTACT**T
>NR_075422.1	Shewanella piezotolerans	GAAGTGAAACACAGTATCGCCG**ATGGTAGTG**
>NR_103224.1	Thermotoga maritima	GCCGTTAAGCCGCCCTGGGCCG**ATGGTAGT**A
>M10821.1	Thermus aquaticus	GAAGTGAAACGCGCCAGCGCCG**ATGGTCACTG**
>NR_075167.1	Thermus thermophilus	GAAGTGAAACGCGCCAGCGCCG**ATGGTACTG**

Figure 8. (**A**) Consensus third quarter of 5S rRNA sequences from 10 bacteria with weak nucleotides (A and T/U) flanked by strong nucleotides (G and C); (**B**) third quarter of the sequence of 5S rRNA with fragment of the tail part of AL (in red); (**C**) Consensus sequence calculated for 26 Bacteria and Archaea.

3.4. Searching for AL Motifs in Current Genomes

The four basic functional activities with their main proteins considered in the following correspond to membrane transport (ATPase, translocase), proteolysis (FtsH), translation (ribosomal RNAs and proteins, and aminoacyl-tRNA ligases), and RNA synthesis (RNA polymerase, helicase, and gyrase) [83–85]. The AL RNA is capable of replication as in a "quine" informatics program, leaving functional traces in the current mRNAs of these proteins. The eight pentamers at the head of the hairpin form of AL all possess at least one nucleotide linked 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 decreasing frequency as we move away from the origin of life. These pentamers are the following: AUUCA, UUCAA, UCAAG, CAAGA, AAGAU, AGAUG, GAUGA, AUGAA, and UGAAU. The proximities to AL, P_{PAL}, and P_{PAL} Doublet are calculated in [42] as one 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 five species, from the oldest to the youngest: Methanococcus maripaludis (Mm), Trichomonas vaginalis (Tri), Entamoeba histolytica (Ent), Saccharomyces cerevisiae (SC), and Homo sapiens (HS) (see Supplementary Material S3).

Among the five species listed in 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). These calculations show that the most frequent are those corresponding to identical or close (but without overlap) AL codons, which correspond to hydrophilic amino acids (GAA-GAA, GAA-GAT, GAA-GAT, GAA-AGA, and GAT-GAA) and a pair of codons corresponding to a pair of hydrophobic and hydrophilic acids (ATT-GAA). This observation reinforces the hypothesis of the primordial catalytic role of AL in favoring peptide synthesis at the origin of life.

Table 1. n_o (resp. n_e) is the observed (resp. expected) number of pentamers (in red) belonging to the head of the hairpin form of AL, and P_{PAL} (in blue) is twice the number of empirical standard deviations σ_e of n_e contained in interval $[n_e, n_0]$. Mean P_{PAL} (in red) is the expectation of P_{PAL} for the 5 observed species. The calculation is identical for the P_{PAL} Doublet (in blue). P_{PAL} and P_{PAL} Doublet are measures of the proximity of RNAs (tRNAs or mRNAs of the 5 observed species) to AL.

Molecule	Species	n_o	N	n_e	(s_e)	$P_{PAL} = \frac{2(n_0 - n_e)}{\sigma_e}$	Mean P_{PAL}	P_{PAL} Doublet	Mean P_{PAL} Doublet
rprotein L18	HS	14	639	5.7	(2.4)	7	10.1	4	8
	SC	19	536	4.7	(2.2)	13.1		7.4	
	Ent	9	536	4.7	(2.2)	4		5	
	Tri	25	562	5	(2.2)	18		7.1	
	Mm	8	227	2	(1.4)	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)	10.5		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)	20.8		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	

Table 1. *Cont.*

Molecule	Species	n_o	N	n_e	(s_e)	$P_{PAL} = \frac{2(n_0 - n_e)}{\sigma_e}$	Mean P_{PAL}	P_{PAL} Doublet	Mean P_{PAL} Doublet
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	(4.1)	13.6		36.6	
	Tri	39	1946	17.1	(4.1)	10.6		8.7	
	Mm	36	1721	15.2	(3.9)	10.8		26	
mRNA DNA 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)	13.2		8.2	
	Mm	45	2351	20.7	(4.6)	10.7		18.2	
mRNA ATPase	HS	78	3414	30	(5.5)	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	(6.5)	27.6		23	
	Ent	106	3002	26.4	(5.1)	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	(6.3)	22.2		16.6	
	Ent	63	2573	22.6	(4.8)	17		30.9	
	Tri	46	1256	11.1	(3.3)	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	(6.1)	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	9.7	(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			
AL		22	22	0.19	(0.44)	100	100		

Table 2. Pairs of consecutive codons observed more than 2 times at least once among the 5 species of Table 1. Doublets appearing more than $\mu + 2\sigma = 61.3$ times (where $\mu = 29$ is the expectation of the totals for all considered pairs and $\sigma = 5.4$ the corresponding standard deviation) are in red.

Doublet	Homo Sapiens	Saccharomyces	Entamoeba	Trichomonas	Methanococcus	Total
ATT ATT	3	8	20	0	5	36
ATT CAA	2	5	14	4	11	36
ATT GAT	3	11	12	0	8	34

Table 2. *Cont.*

Doublet	Homo Sapiens	Saccharomyces	Entamoeba	Trichomonas	Methanococcus	Total
ATT GAA	16	17	25	3	27	88
ATT CCA	4	2	2	6	7	21
ATT ACT	1	10	4	3	3	21
ATT AGA	3	5	4	3	1	16
ATT TAC	0	0	2	0	3	5
CAA ATT	6	4	9	0	1	20
CAA CAA	2	5	7	4	0	18
CAA GAT	7	6	4	5	0	22
CAA GAA	11	12	11	6	6	46
CAA CCA	0	0	4	0	0	4
CAA AGA	6	4	9	0	4	23
CAA TAC	2	6	1	0	0	9
CAA ACT	7	0	5	0	0	12
GAT ATT	9	14	21	0	9	53
GAT GAT	19	14	26	15	4	78
GAT GAA	22	23	35	16	26	122
GAT CCA	5	3	4	3	0	15
GAT AGA	5	2	9	1	4	21
GAT ACT	1	6	6	0	7	20
GAT CAA	1	2	6	0	1	10
GAT TAC	2	0	0	5	5	12
GAA ATT	9	6	27	2	7	51
GAA CAA	6	11	12	2	7	38
GAA GAT	20	21	33	14	14	102
GAA GAA	20	33	43	16	37	149
GAA CCA	3	9	11	6	2	31
GAA ACT	8	10	8	1	6	33
GAA AGA	8	10	7	3	7	35
GAA TAC	0	10	0	4	4	18
CCA ATT	2	3	7	4	6	22
CCA GAA	8	6	7	3	4	28
CCA AGA	8	0	3	4	1	16
CCA GAT	8	0	3	0	0	11
CCA ACT	6	0	6	0	0	12
ACT CAA	0	3	9	4	1	17
ACT GAT	4	0	7	0	4	15
ACT GAA	8	0	5	3	6	22
ACT ATT	2	6	9	0	2	19
ACT CCA	7	9	4	0	1	21
ACT TAC	1	0	0	0	2	3
ACT ACT	0	7	4		2	13
AGA ATT	3	5	7	1	5	21
AGA GAT	6	12	11	0	5	34
AGA GAA	13	11	12	9	20	65
AGA CAA	4	9	2	11	0	26
AGA AGA	7	6	7	0	4	24
AGA CCA	0	0	2	0	0	2
AGA TAC	1	0	0	0	4	5

Table 2. *Cont.*

Doublet	Homo Sapiens	Saccharomyces	Entamoeba	Trichomonas	Methanococcus	Total
TAC ATT	2	3	0	4	1	0
TAC GAT	1	5	1	0	6	13
TAC GAA	1	0	1	5	10	17
TAC TAC	3	0	0	3	2	8
TAC AGA	0	0	0	3	2	5
TAC CCA	3	0	1	1	5	19

From Tables 1 and 2, it can be considered that AL 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 in Figure 9 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.
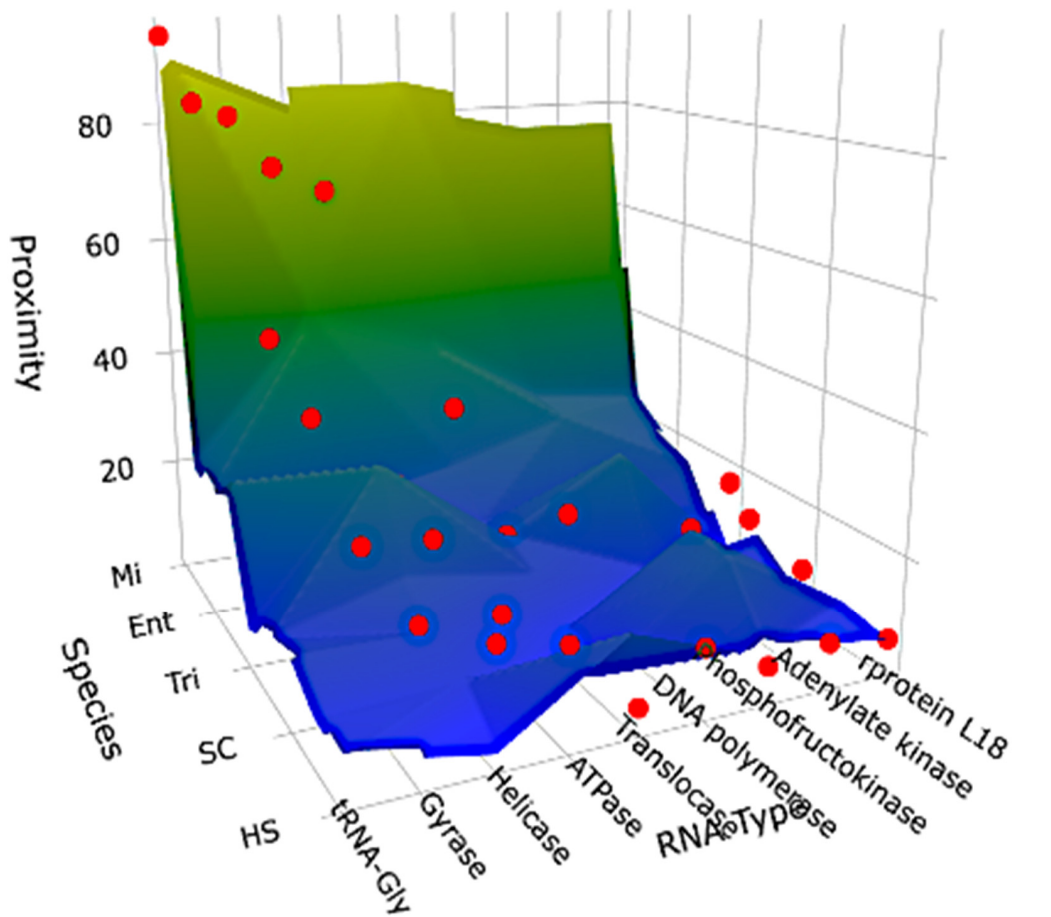
Figure 9. Surface representing the data of Table 1.

4. Conclusions

To support a network view of the origin of life, as discussed in 2018 by Fontecilla-Camps [86], Aguirre et al. [87], and Seligmann and Raoult [88], the AL RNA can be proposed as a key in the primitive machinery for 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 bonds, the machine functioning as a "proto-ribosome" into a "proto-membrane", close to a "proto-cell" with a network organization, each element favoring the survival of the others. This approach stands as a solution to 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 has been supported for a century by different works, theoretical as well as experimental: for example, in 1926, H.J. Muller suggested that life began not as an enzyme, but as a gene [89]. 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 [90,91]. In the theory of autopoiesis [92,93], the first living system is self-reproducing [94,95] 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 and theoretical arguments have been made about the role of the primitive RNAs in the progressive constitution of the genetic code [96–105].

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 [106] to DNA computing [107]. 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 essential building blocks for cell survival.

Further studies could experimentally investigate the AL RNA as a potential catalyst of peptide synthesis [39] and search for its role in building protein and cell worlds after the RNA world, and its possible role in consolidating the genetic code. This should be carried out 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-protein structures like the ribosome or RNAs like the ribozymes [108–136]. In particular, many peptide synthesis experiments have been carried out, with or without RNA template [137–166]. Those carried out with RNA template use sequences comparable to AL fragments, but in the reverse $3'$–$5'$ direction [137–161]. In the future, it will be necessary to systematically compare the efficiency of these amino acid polymerizations: in solution without an RNA template [162–166], in solution with an RNA template (an attempt using microcalorimetry encountered some technical obstacles in the past [39]), and with or without an RNA template in solid substrate (such as montmorillonite, initially proposed by A. Katchalsky [1]). The research presented in these articles is very important for understanding the synthesis of the first peptides, but also the progressive structuring of the genetic code, from its ancient versions to the current genetic code, from biochemical–statistical experimentation [167] to phylogenomic explorations [168–170], which help to explain the transition from a possible early 'operational' version to the current 'standard' genetic code.

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. [CrossRef] [PubMed]
2. Paecht-Horowitz, M.; Katchalsky, A. Synthesis of amino acyl-adenylates under prebiotic conditions. *J. Mol. Evol.* **1973**, *2*, 91–98. [CrossRef] [PubMed]
3. Brack, A. Polymerisation en phase aqueuse d'acides aminés sur des argiles. *Clay Miner.* **1976**, *11*, 117–120. [CrossRef]
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. [CrossRef] [PubMed]
5. Noller, H.F. Evolution of protein synthesis from an RNA world. *Cold Spring Harb. Perspect. Biol.* **2012**, *4*, a003681. [CrossRef]
6. Jash, B.; Tremmel, P.; Jovanovic, D.; Richert, C. Single nucleotide translation without ribosomes. *Nat. Chem.* **2021**, *13*, 751–757. [CrossRef]
7. Ishida, T. Simulation of the emergence of cell-like morphologies with evolutionary potential based on virtual molecular interactions. *Sci. Rep.* **2024**, *14*, 2086. [CrossRef]
8. Eigen, M.; Schuster, P. The hypercycle: A principle of natural self-organization. Part C: The realistic hypercycle. *Naturwissenschaften* **1978**, *65*, 341–369. [CrossRef]
9. Eigen, M.; Winkler-Oswatitsch, R. Transfer-RNA: The early adaptor. *Naturwissenschaften* **1981**, *68*, 217–228. [CrossRef]
10. Bernal, D. *The Physical Basis of Life*; Routledge 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. [CrossRef]
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. [CrossRef] [PubMed]
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. [CrossRef] [PubMed]
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. [CrossRef] [PubMed]
15. Ponnamperuma, C.; Sagan, C.; Mariner, R. Synthesis of adenosine triphosphate under possible primitive earth conditions. *Nature* **1963**, *199*, 222–226. [CrossRef]
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. [CrossRef]
17. Caetano-Anolles, G.; Kim, K.M. *The Origin and Evolution of the Archaeal Domain*; Hindawi Publishing Corporation: London, UK, 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. [CrossRef]
19. Woese, C.R. A New Biology for a New Century. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 173–186. [CrossRef]
20. Shapiro, R. Small Molecule Interactions were Central to the Origin of Life. *Q. Rev. Biol.* **2006**, *81*, 105–126. [CrossRef]
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. [CrossRef]
22. Colín-García, M. Hydrothermal vents and prebiotic chemistry: A review. *Bol. Soc. Geol. Mex.* **2016**, *68*, 599–620. [CrossRef]
23. Yarus, M. *Life from an RNA World: The Ancestor Within*; Harvard University Press: Cambridge, MA, USA, 2010.
24. Yarus, M. Eighty routes to a ribonucleotide world; dispersion and stringency in the decisive selection. *RNA* **2018**, *24*, 1041–1055. [CrossRef] [PubMed]
25. Yarus, M. On an RNA-membrane protogenome. *arXiv* **2025**, arXiv:2502.00647. [CrossRef]
26. Lancet, D.; Zidovetzki, R.; Markovitch, O. Systems protobiology: Origin of life in lipid catalytic networks. *J. R. Soc. Interface* **2018**, *15*, 20180159. [CrossRef] [PubMed]

27. Raine, D.J.; Norris, V. Lipid domain boundaries as prebiotic catalysts of peptide bond formation. *J. Theor. Biol.* **2007**, *246*, 176–185. [CrossRef]

28. Kahana, A.; Lancet, D. Protobiotic Systems Chemistry Analyzed by Molecular Dynamics. *Life* **2019**, *9*, 38. [CrossRef]

29. Caforio, A.; Driessen, A.J.M. Archaeal phospho-lipids: Structural properties and biosynthesis. *BBA-Mol. Cell Biol. Lipids* **2016**, *1862*, 1325–1339. [CrossRef]

30. Demongeot, J. Au Sujet de Quelques Modèles Stochastiques Appliqués à la Biologie. Ph.D. Thesis, Université Joseph Fourier, Grenoble, France, 1975. Available online: https://tel.archives-ouvertes.fr/tel-00286222 (accessed on 5 January 2025).

31. Demongeot, J. Sur la possibilité de considérer le code génétique comme un code à enchaînement. *Rev. 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. *Theor. Comp. Sc.* **2004**, *322*, 313–334. [CrossRef]

35. Demongeot, J.; Elena, A.; Weil, G. Potential-Hamiltonian decomposition of cellular automata. *Appl. Degeneracy Genet. Code Cycl. Codes III. C. R. Acad. Sc. Biol.* **2006**, *329*, 953–962.

36. Demongeot, J. Primitive genome and RNA relics. In Proceedings of the EMBC' 07, Lyon, France, 22–26 August 2007; IEEE Proceedings: Piscataway, NJ, USA, 2007; pp. 6338–6342.

37. Demongeot, J.; Moreira, A. A circular RNA at the origin of life. *J. Theor. Biol.* **2007**, *249*, 314–324. [CrossRef] [PubMed]

38. Demongeot, J.; Drouet, E.; Moreira, A.; Rechoum, Y.; Séné, S. Micro-RNAs: Viral genome and robustness of the genes expression in host. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **2009**, *367*, 4941–4965. [CrossRef] [PubMed]

39. Demongeot, J.; Glade, N.; Moreira, A.; Vial, L. RNA relics and origin of life. *Int. J. Mol. Sci.* **2009**, *10*, 3420–3441. [CrossRef]

40. Demongeot, J.; Hazgui, H.; Bandiera, S.; Cohen, O.; Henrion-Caude, A. MitomiRs, ChloromiRs and general modelling of the microRNA inhibition. *Acta Biotheor.* **2013**, *61*, 367–383. [CrossRef]

41. Demongeot, J. "Protoribosome" as new game of life. *BioRxiv* **2017**. [CrossRef]

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 Biotheor.* **2016**, *64*, 403–426. [CrossRef]

43. Demongeot, J.; Norris, V. Emergence of a "Cyclosome" in a Primitive Network Capable of Building "Infinite" Proteins. *Life* **2019**, *9*, 51. [CrossRef]

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. [CrossRef]

45. Demongeot, J.; Seligmann, H. Spontaneous evolution of circular codes in theoretical minimal RNA rings. *Gene* **2019**, *705*, 95–102. [CrossRef]

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. [CrossRef] [PubMed]

47. Demongeot, J.; Seligmann, H. Bias for 3′-dominant codon directional asymmetry in theoretical minimal RNA rings. *J. Comput. Biol.* **2019**, *26*, 1003–1012. [CrossRef] [PubMed]

48. Demongeot, J.; Seligmann, H. Theoretical minimal RNA rings designed according to coding constraints mimick deamination gradients. *Sci. Nat.* **2019**, *106*, 44. [CrossRef] [PubMed]

49. Demongeot, J.; Seligmann, H. Pentamers with non-redundant frames: Bias for natural circular code codons. *J. Mol. Evol.* **2020**, *88*, 194–201. [CrossRef]

50. Demongeot, J.; Seligmann, H. The primordial tRNA acceptor stem code from theoretical minimal RNA ring clusters. *BMC Genet.* **2020**, *21*, 7. [CrossRef]

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. [CrossRef]

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. [CrossRef]

54. Demongeot, J.; Henrion-Caude, A. The old and the new on the prebiotic conditions of the origin of life. *Biology* **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. *Comptes Rendus Biol.* **2020**, *343*, 111–122. [CrossRef]

56. Demongeot, J.; Seligmann, H. Evolution of tRNA subelement accretion from small and large ribosomal RNAs. *Biosystems* **2022**, *193*, 104796. [CrossRef] [PubMed]

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. [CrossRef] [PubMed]

58. Demongeot, J.; Thellier, M. Primitive oligomeric RNAs at the origins of life on Earth. *Int. J. Mol. Sci.* **2023**, *24*, 2274. [CrossRef] [PubMed]

59. Demongeot, J.; Waku, J.; Cohen, O. Combinatorial and frequency properties of the ribosome ancestors. *Math. Biosci. Eng.* **2023**, *21*, 884–902. [CrossRef]

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. [CrossRef]

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 Proceedings of the SAI Computing Conference 2025, London, UK, 19–20 June 2025; Lecture Notes in Networks & Systems. Springer Nature: New York, NY, USA, 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. [CrossRef]

63. Li, S.; Yang, J. System analysis of synonymous codon usage biases in archaeal virus genomes. *J. Theor. Biol.* **2014**, *355*, 128–139. [CrossRef]

64. Bahiri-Elitzur, S.; Tuller, T. Codon-based indices for modeling gene expression and transcript evolution. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2646–2663. [CrossRef]

65. GtRNAdb. Available online: http://gtrnadb.ucsc.edu/ (accessed on 22 February 2025).

66. NCBI Nucleotide. Available online: https://www.ncbi.nlm.nih.gov/nucleotide (accessed on 22 February 2025).

67. MiRBase. Available online: http://www.mirbase.org/ (accessed on 22 February 2025).

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. [CrossRef]

69. Rogozin, I.B.; Carmel, L.; Csuros, M.; Koonin, E.V. Origin and evolution of spliceosomal introns. *Biol. Direct.* **2012**, *7*, 11. [CrossRef]

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. [CrossRef] [PubMed]

71. Forterre, P. The Common Ancestor of Archaea and Eukarya Was Not an Archaeon. *Archaea* **2013**, *2013*, 372396. [CrossRef] [PubMed]

72. Chatterjee, S.; Yadav, S. The Origin of Prebiotic Information System in the Peptide/RNA World: A Simulation Model of the Evolution of Translation and the Genetic Code. *Life* **2019**, *9*, 25. [CrossRef]

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. Biol.* **2007**, *330*, 13–32. [CrossRef]

74. Yarus, M. The meaning of a minuscule ribozyme. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2011**, *366*, 2902–2909. [CrossRef] [PubMed]

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. [CrossRef]

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. [CrossRef]

77. Tamura, K.; Schimmel, P. Chiral-selective aminoacylation of an RNA minihelix. *Science* **2004**, *305*, 1253. [CrossRef]

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. [CrossRef]

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. *Nat. Biotechnol.* **2020**, *38*, 1328–1336. [CrossRef]

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. *Nat. Biotechnol.* **2020**, *38*, 471–481. [CrossRef] [PubMed]

81. Petersen, J.; Brinkmann, H.; Pradella, S. Diversity and evolution of *repABC* type plasmids in *Rhodobacterales*. *Environ. Microbiol.* **2009**, *11*, 2627–2638. [CrossRef] [PubMed]

82. Palacios-Pérez, M.; José, M.V. A Proposal of the Ur-RNAome. *Genes* **2023**, *14*, 2158. [CrossRef]

83. Trifonov, E.N. Consensus temporal order of amino acids and evolution of the triplet code. *Gene* **2000**, *261*, 139–151. [CrossRef]

84. Trifonov, E.N.; Bettecken, T. Sequence fossils, triplet expansion, and reconstruction of earliest codons. *Gene* **1997**, *205*, 1–6. [CrossRef]

85. Sobolevsky, Y.; Guimarães, R.C.; Trifonov, E.N. Towards functional repertoire of the earliest proteins. *J. Biomol. Struct. Dyn.* **2013**, *31*, 1293–1300. [CrossRef]

86. Fontecilla-Camps, J. Geochemical Continuity and Catalyst/Cofactor Replacement in the Emergence and Evolution of Life. *Angew. Chem.* **2018**, *130*, 08438. [CrossRef] [PubMed]

87. 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. [CrossRef]

88.   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. [CrossRef]

89.   Muller, H.J. The gene as the basis of life. In Proceedings of the International Congress of Plant Sciences, Ithaca, NY, USA, 16–23 August 1926; Duggar, B.M., Ed.; Menasha: Banta, WI, USA, 1929; pp. 897–921.

90.   Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **1971**, *58*, 465–523. [CrossRef]

91.   Eigen, M.; Lindemann, B.; Winkler-Oswatitsch, R.; Clarke, C.H. Pattern Analysis of 5S rRNA. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 2437–2441. [CrossRef] [PubMed]

92.   Bourgine, P.; Stewart, J. Autopoiesis and cognition. *Artif. Life* **2004**, *10*, 327–345. [CrossRef]

93.   Ono, N.; Ikegami, T. Self-maintenance and self-reproduction in an abstract cell model. *J. Theor. Biol.* **2000**, *206*, 243–253. [CrossRef] [PubMed]

94.   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, 10–14 September 2001; Kelemen, J., Sosik, S., Eds.; Springer: Berlin, Germany, 2001; pp. 186–195.

95.   Noble, D. Genes are not the blueprint for life. *Nature* **2019**, *626*, 254–255. [CrossRef]

96.   Noble, D.; Joyner, M. The physiology of evolution. *J. Physiol.* **2024**, *602*, 2361–2365. [CrossRef]

97.   Dufton, M.J. Genetic code synonym quotas and amino acid complexity: Cutting the cost of proteins? *J. Theor. Biol.* **1997**, *187*, 165–173. [CrossRef]

98.   Davis, B.K. Evolution of the genetic code. *Prog. Biophys. Mol. Biol.* **1999**, *72*, 157–243. [CrossRef]

99.   Wong, J.T.F. Coevolution theory of the genetic code at age thirty. *Bioessays* **2005**, *27*, 416–425. [CrossRef]

100.   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* **2016**, *6*, 12. [CrossRef]

101.   Takeuchi, N.; Kaneko, K. The origin of the central dogma through conflicting multilevel selection. *Proc. R. Soc. B* **2019**, *286*, 20191359. [CrossRef]

102.   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. [CrossRef] [PubMed]

103.   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. [CrossRef] [PubMed]

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. [CrossRef] [PubMed]

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. [CrossRef]

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. [CrossRef]

107.   Levy, N.; Schabanel, N. ENSnano: A 3D Modeling Software for DNA Nanostructures. *DNA* **2021**, *5*, 1–5.

108.   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. [CrossRef]

109.   Eme, L.; Spang, A.; Lombard, J.; Stairs, C.W.; Ettema, T.J.G. Archaea and the origin of eukaryotes. *Nature* **2017**, *15*, 711–723. [CrossRef]

110.   Legendre, M.; Fabre, E.; Poirot, O.; Jeudy, S.; Lartigue, A.; Alempic, J.M.; Beucher, L.; Philippe, N.; Bertaux, L.; Christo-Foroux, E.; et al. Diversity and evolution of the emerging Pandoraviridae family. *Nat. Commun.* **2018**, *9*, 2285. [CrossRef]

111.   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. [CrossRef]

112.   Bartnik, E.; Borsuk, P. A glycine tRNA gene from lupine mitochondria. *Nucleic Acids Res.* **1986**, *14*, 2407. [CrossRef] [PubMed]

113.   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. [CrossRef] [PubMed]

114.   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. [CrossRef] [PubMed]

115.   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. [CrossRef]

116.   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. [CrossRef]

117. 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. [CrossRef]

118. Pérez-Ruiz, M.; Barroso-delJesus, A.; Berzal-Herranz, A. Specificity of the Hairpin Ribozyme. *J. Biol. Chem.* **1999**, *274*, 29376–29380. [CrossRef]

119. Müller, U.F. Design and Experimental Evolution of trans-Splicing Group I Intron Ribozymes. *Molecules* **2017**, *22*, 75. [CrossRef]

120. Paul, N.; Joyce, G.F. A self-replicating ligase ribozyme. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12733–12740. [CrossRef]

121. 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. [CrossRef]

122. Hammann, C.; Luptak, A.; Perreault, J.; De La Peña, M. The ubiquitous hammerhead ribozyme. *RNA* **2012**, *18*, 871–885. [CrossRef]

123. Harris, K.A.; Lünse, C.E.; Li, S.; Brewer, K.I.; Breaker, R.R. Biochemical analysis of hatchet self-cleaving ribozymes. *RNA* **2015**, *21*, 1852–1858. [CrossRef]

124. Agmon, I.C. Could a Proto-Ribosome Emerge Spontaneously in the PrebioticWorld? *Molecules* **2016**, *21*, 1701. [CrossRef]

125. Arquès, D.G.; Michel, C.J. A complementary circular code in the protein coding genes. *J. Theor. Biol.* **1996**, *182*, 45–58. [CrossRef]

126. 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. [CrossRef]

127. 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* **2019**, *9*, 37. [CrossRef]

128. 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* **2018**, *8*, 44. [CrossRef]

129. Seligmann, H. Protein Sequences Recapitulate Genetic Code Evolution. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 177–189. [CrossRef]

130. 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. [CrossRef]

131. Robinson, R. Jump-starting a cellular world: Investigating the origin of life, from soup to networks. *PLoS Biol.* **2005**, *3*, e396. [CrossRef]

132. 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. [CrossRef]

133. Koonin, E.V.; Novozhilov, A.S. Origin and evolution of the genetic code: The universal enigma. *IUBMB Life* **2009**, *61*, 99–111. [CrossRef]

134. Rodin, A.S.; Szathmáry, E.; Rodin, S.N. On origin of genetic code and tRNA before translation. *Biol. Direct.* **2011**, *6*, 14. [CrossRef]

135. Koonin, E.V. Frozen Accident Pushing 50: Stereochemistry, Expansion, and Chance in the Evolution of the Genetic Code. *Life* **2017**, *7*, 22. [CrossRef]

136. Gonzalez, D.L.; Giannerini, S.; Rosa, R. On the origin of degeneracy in the genetic code. *Interface Focus* **2019**, *9*, 20190038. [CrossRef]

137. Chang, L.H.; Seitz, O. RNA-templated chemical synthesis of proapoptotic L- and d-peptides. *Bioorganic Med. Chem.* **2022**, *66*, 116786. [CrossRef]

138. Zhao, F.; Frandsen, M.; Capodaglio, S.; Sleiman, H.F. DNA-Mediated Peptide Assembly into Protein Mimics. *J. Am. Chem. Soc.* **2024**, *146*, 1946–1956. [CrossRef]

139. Radakovic, A.; Wright, T.H.; Lelyveld, V.S.; Szostak, J.W. A Potential Role for Aminoacylation in Primordial RNA Copying Chemistry. *Biochemistry* **2021**, *60*, 477–488. [CrossRef]

140. Węgrzyn, E.; Mejdrová, I.; Müller, F.M.; Nainytė, M.; Escobar, L.; Carell, T. RNA-Templated Peptide Bond Formation Promotes L-Homochirality. *Angew. Chem. Int. Ed. Engl.* **2024**, *63*, e202319235. [CrossRef]

141. Węgrzyn, E.; Mejdrová, I.; Carell, T. Gradual evolution of a homo-l-peptide world on homo-d-configured RNA and DNA. *Chem. Sci.* **2024**, *15*, 14171–14176. [CrossRef]

142. Kauffman, S.A.; Lehman, N. Mixed anhydrides at the intersection between peptide and RNA autocatalytic sets: Evolution of biological coding. *Interface Focus* **2023**, *13*, 20230009. [CrossRef]

143. Su, M.; Schmitt, C.; Liu, Z.; Roberts, S.J.; Liu, K.C.; Röder, K.; Jäschke, A.; Wales, D.J.; Sutherland, J.D. Triplet-Encoded Prebiotic RNA Aminoacylation. *J. Am. Chem. Soc.* **2023**, *145*, 15971–15980. [CrossRef]

144. Wickramasinghe, N.S.; Staves, M.P.; Lacey, J.C., Jr. Stereoselective, nonenzymatic, intramolecular transfer of amino acids. *Biochemistry* **1991**, *30*, 2768–2772. [CrossRef]

145. Gless, B.H.; Jones, E.; Labão-Almeida, C.; Tang, C.; Gottscheber, N.; Couto, R.; Bernardes, G.J.L. Conditional Activation of Protein Therapeutics by Templated Removal of Peptide Nucleic Acid Masking Groups. *Angew. Chem. Int. Ed. Engl.* **2025**, *64*, e202502268. [CrossRef]

146. Räuchle, M.; Leveau, G.; Richert, C. Synthesis of Peptido RNAs from Unprotected Peptides and Oligoribonucleotides via Coupling in Aqueous Solution. *Eur. J. Org. Chem.* **2020**, *2020*, 6966–6975. [CrossRef]

147. Altrichter, Y.; Bou-Dib, P.; Kuznia, C.; Seitz, O. Towards a templated reaction that translates RNA in cells into a proaptotic peptide-PNA conjugate. *J. Pept. Sci.* **2023**, *29*, e3477. [CrossRef]

148. Di Pisa, M.; Hauser, A.; Seitz, O. Maximizing Output in RNA-Programmed Peptidyl-Transfer Reactions. *ChemBioChem* **2017**, *18*, 872–879. [CrossRef]

149. Middel, S.; Panse, C.H.; Nawratil, S.; Diederichsen, U. Native Chemical Ligation Directed by Photocleavable Peptide Nucleic Acid (PNA) Templates. *ChemBioChem* **2017**, *18*, 2328–2332. [CrossRef]

150. Vázquez, O.; Seitz, O. Templated native chemical ligation: Peptide chemistry beyond protein synthesis. *J. Pept. Sci.* **2014**, *20*, 78–86. [CrossRef]

151. Saha, P.; Panda, D.; Dash, J. Nucleic acids as templates and catalysts in chemical reactions: Target-guided dynamic combinatorial chemistry and in situ click chemistry and DNA/RNA induced enantioselective reactions. *Chem. Soc. Rev.* **2023**, *52*, 4248–4291. [CrossRef]

152. Escobar, L. Covalent Linkages Used in Prebiotic Chemistry for RNA-Templated Amino Acid Transfer and Peptide Synthesis. *ChemSystemsChem* **2024**, *6*, e202400030. [CrossRef]

153. Guo, X.; Su, M. The Origin of Translation: Bridging the Nucleotides and Peptides. *Int. J. Mol. Sci.* **2022**, *24*, 197. [CrossRef]

154. Kuila, S.; Nanda, J. Cysteine-Based Dynamic Self-Assembly and Their Importance in the Origins of Life. *ChemSystemsChem* **2024**, *6*, e202400022. [CrossRef]

155. Ianeselli, A.; Salditt, A.; Mast, C.; Dieter Braun, D. Physical non-equilibria for prebiotic nucleic acid chemistry. *Nat. Rev. Phys.* **2023**, *5*, 185–195. [CrossRef]

156. Bandela, A.K.; Wagner, N.; Sadihov, H.; Morales-Reina, S.; Chotera-Ouda, A.; Basu, K.; Cohen-Luria, R.; de la Escosura, A.; Ashkenasy, G. Primitive selection of the fittest emerging through functional synergy in nucleopeptide networks. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2015285118. [CrossRef]

157. Sadihov-Hanoch, H.; Bandela, A.K.; Chotera-Ouda, A.; Ben David, O.; Cohen-Luria, R.; Lynn, D.G.; Ashkenasy, G. Dynamic exchange controls the assembly structure of nucleic-acid-peptide chimeras. *Soft Matter* **2023**, *19*, 3940–3945. [CrossRef]

158. Chotera, A.; Sadihov, H.; Cohen-Luria, R.; Monnard, P.A.; Ashkenasy, G. Functional Assemblies Emerging in Complex Mixtures of Peptides and Nucleic Acid-Peptide Chimeras. *Chemistry* **2018**, *24*, 10128–10135. [CrossRef]

159. Müller, F.; Escobar, L.; Xu, F.; Węgrzyn, E.; Nainytė, M.; Amatov, T.; Chan, C.Y.; Pichler, A.; Carell, T. A prebiotically plausible scenario of an RNA-peptide world. *Nature* **2022**, *605*, 279–284. [CrossRef]

160. Singer, J.N.; Müller, F.M.; Węgrzyn, E.; Hölzl, C.; Hurmiz, H.; Liu, C.; Escobar, L.; Carell, T. Loading of Amino Acids onto RNA in a Putative RNA-Peptide World. *Angew. Chem. Int. Ed. Engl.* **2023**, *62*, e202302360. [CrossRef]

161. Nainytė, M.; Müller, F.; Ganazzoli, G.; Chan, C.Y.; Crisp, A.; Globisch, D.; Carell, T. Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World. *Chemistry* **2020**, *26*, 14856–14860. [CrossRef]

162. Prosdocimi, F.; de Farias, S.T.; José, M.V. Prebiotic chemical refugia: Multifaceted scenario for the formation of biomolecules in primitive Earth. *Theory Biosci.* **2022**, *141*, 339–347. [CrossRef]

163. Turk, R.M.; Chumachenko, N.V.; Yarus, M. Multiple translational products from a five-nucleotide ribozyme. *Proc. Natl Acad. Sci. USA* **2010**, *107*, 4585–4589. [CrossRef]

164. Yang, Y.; Wang, Z.; Bai, J.; Qiao, H. Prebiotic Peptide Synthesis: How Did Longest Peptide Appear? *J. Mol. Evol.* **2025**, *93*, 193–211. [CrossRef]

165. Sutherland, J.D. The origin of life—Out of the blue. *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 104–121. [CrossRef]

166. Jash, B.; Richert, C. Templates direct the sequence-specific anchoring of the *C*-terminus of peptido RNAs. *Chem. Sci.* **2020**, *11*, 3487–3494. [CrossRef]

167. Carter, C.W., Jr.; Wills, P.R. The Roots of Genetic Coding in Aminoacyl-tRNA Synthetase Duality. *Annu. Rev. Biochem.* **2021**, *90*, 349–373. [CrossRef]

168. Caetano-Anollés, G.; Wang, M.; Caetano-Anollés, D. Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. *PLoS ONE* **2013**, *8*, e72225. [CrossRef]

169. Harish, A.; Caetano-Anollés, G. Ribosomal history reveals origins of modern protein synthesis. *PLoS ONE* **2012**, *7*, e32776. [CrossRef]

170. Mughal, F.; Caetano-Anollés, G. Evolution of intrinsic disorder in the structural domains of viral and cellular proteomes. *Sci. Rep.* **2025**, *15*, 2878. [CrossRef]