Footprints of a Singular 22-Nucleotide RNA Ring at the Origin of Life

Jacques Demongeot [1,*] and Alexandra Henrion-Caude [2]

[1] Laboratory AGEIS EA 7407, Faculty of Medicine, University of Grenoble Alpes, 38700 La Tronche, France
[2] Future of Research Team, SimplissimA International Research Institute, 39 rue saint Louis,
    Port-Louis 11324, Mauritius; a.caude@simplissima.org
[*] Correspondence: Jacques.Demongeot@univ-grenoble-alpes.fr

Abstract: (1) Background: Previous experimental observations and theoretical hypotheses have been providing insight into a hypothetical world where an RNA hairpin or ring may have debuted as the primary informational and functional molecule. We propose a model revisiting the architecture of RNA-peptide interactions at the origin of life through the evolutionary dynamics of RNA populations. (2) Methods: By performing a step-by-step computation of the smallest possible hairpin/ring RNA sequences compatible with building up a variety of peptides of the primitive network, we inferred the sequence of a singular docosameric RNA molecule, we call the ALPHA sequence. Then, we searched for any relics of the peptides made from ALPHA in sequences deposited in the different public databases. (3) Results: Sequence matching between ALPHA and sequences from organisms among the earliest forms of life on Earth were found at high statistical relevance. We hypothesize that the frequency of appearance of relics from ALPHA sequence in present genomes has a functional necessity. (4) Conclusions: Given the fitness of ALPHA as a supportive sequence of the framework of all existing theories, and the evolution of Archaea and giant viruses, it is anticipated that the unique properties of this singular archetypal ALPHA sequence should prove useful as a model matrix for future applications, ranging from synthetic biology to DNA computing.

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

1. Introduction

Elucidating the prebiotic ingredients from which life arose on Earth is one key focus of the origin-of-life research. Previous theoretical hypotheses have been providing insight into a hypothetical world where hairpin or ring of RNA molecules may have debuted as the primary informational and functional molecule [1–4]. Our model starts with revisiting the architecture of RNA–peptide interactions at the origin of life through the evolutionary dynamics of molecular evolution of RNA populations. By performing a step-by-step computation of the smallest possible hairpin/ring RNA sequences compatible with building up a variety of peptides of the primitive network, we inferred the sequence of a singular docosameric RNA molecule, which we refer to as the ALPHA sequence. In compliance with the stereochemical constraints of prebiotic conditions, we searched for any relics of the peptides made from ALPHA in sequences deposited in the different public databases. Sequence matching between ALPHA and sequences from organisms among the earliest forms of life on Earth were found at high statistical relevance.

1.1. Miller's Experiment as a Bridge between the Old towards the New Theories of the Origin of Life

In 1953, Stanley Miller, who was looking for the origin of the first biomolecules, made the most remarkable breakthrough by successfully generating five amino acids out of a chamber simply

containing a mixture of water and gas, and submitted to an electrical discharge [5]. To represent an environment with the four basic atoms of life (oxygen, hydrogen, nitrogen, and carbon), he had used gas made of methane (CH_4), ammonia (NH_3) and hydrogen, and found in the resulting solution the unexpected presence of: glycine, α- and β-alanine, aspartic acid, and α-aminobutyric acid. Later on, another set of amino acids were to be found by Jeffrey Bada, upon analyzing the products of another old experiment by Miller, dating back to 1958, and having assessed a hydrogen sulphide (H_2S)-containing mixture [6]. In 1961, Joan Oró used a solution of hydrogen cyanide (HCN) and ammonia solution, and found that adenine had been synthesized, in addition to amino acids [7]. Two years ago, an adaptation from Miller's experiment was conducted with a simple mixture of NH_3 + CO and H_2O, and resulted in all RNA nucleotides [8], concluding this extraordinary series of experiments that proved that the biomolecules, which are necessary to begin life, could be obtained from simple atmosphere exposed to electrical sparks. "For many years, researchers have asked themselves what preceded what—the proteins, the nucleic acids, or vice versa?" [1]. However, substituting 'protein' by 'function', and 'nucleic acids' by 'information', as proposed by Eigen [2,3], leads to rephrase the question *ad absurdum*, because "function" cannot occur in an organized manner unless "information" is present, and this "information" only acquires its meaning via the "function" for which it is coding. Hence, viewed by Eigen [3] as a "function", proteogenesis enabled to deduce a tRNA structure, which could be the ancestor of all the present tRNAs.

To attain the adequate concentration of the products, which would be an absolute necessity for evolution, it was suggested in 1951 by Bernal that adsorption could take place on very fine clay deposits. As a "polymerization catalyst", montmorillonite would result in decreasing the free amino acid content consequently to polymerization [4], and after a certain time, obtaining polypeptides at the same molecular weight distribution pattern as those produced if amino acid adenylates were polymerized. In fact, Ponnamperuma and his collaborators described the formation of ATP [9], and in 1995, the interactions between amino acids and nucleotides as a possible physicochemical basis for the origin of the genetic code [10]. Altogether, those observations actually form the experimental corpus of the stereochemical theory of the origin of life [11,12], and this was the origin of a long debate.

If Shapiro [13] admitted that "life began within a mixture of simple organic molecules, with possible participation by minerals" and that "a triple stem-loop (RNA) structure, containing 40–60 nucleotides, offered a reasonable hope of functioning as a replicase ribozyme", he and his colleague Bernhardt [14] were critical to the montmorillonite hypothesis and the stereochemical approach. Conversely, in line with Ponnamperuma, Yarus recently defended the idea of a catalytic role of RNA rings favoring the peptidic bonds between amino acids [15,16]. Accordingly, an illustration of his proposal was that: "a ready rationale exists for smaller individual amino acid sites, still side chain specific. These can be extreme single-ended sites, forced to be small because of the crowding of two sites produced by the short single covalent peptide bond between His and Phe" [15].

At last, a number of recent models have pointed to the role of a lipid component in the very early stages of the origin of RNA structures [17–21], which secondarily could have functioned independently [22]. Another possibility is that those lipids may result from the apparition of proteins catalyzing their synthesis, as in Archaea where a number of enzymes contribute to the biosynthesis of lipids [23]. In fact, Forterre and his coworkers have proposed that Archaea and Bacteria may have a common ancestor, referred to as LUCA, and that the co-evolution of Archaea, Bacteria and Eukarya may be driven in part by viruses (their phages and giant viruses) [24,25]. Accordingly, the complexity of Eukarya may have taken the advantage from retroviruses and large DNA viruses, whilst similar selection pressure could explain the somewhat similar evolution of the archaeal and bacterial mobilomes [25]. Coining the name of "Domain Cell Theory", Staley supported the independent evolution of Archaea, Bacteria and Eukarya, based on an evolution from three distinct cellular lineages [26].

Since the pioneer work by H.J. Muller in 1922, who claimed 20 years before the Beadle and Tatum's "one gene—one enzyme" hypothesis, that life began not as an enzyme but as a gene, the scientific community has been divided as to the origin of life along two hypotheses: the "genetics first"

and the "metabolism first", as accurately pointed out by Fontecilla-Camps [27]. The "genetics first" view, which is based on Miller's experiments and organic components from meteorites, proposes the "RNA World" concept as the origin of life. The "metabolism first" hypothesis posits that life began autotrophically on minerals and/or hydrothermal vents. While lending solid support to either hypothesis is still impossible, the "metabolism first" option may be further explored assuming a "continuous geochemical, catalytically dynamic process" [27]. In this context, the nucleotide synthesis that may have originated on a mineral surface, could have been later replaced by ATP.

In the same "metabolism first" view, Aguirre et al. [28] proposed that large neutral networks of genotypes formed mapping into sets of phenotypes having the same fitness, but several strong connected components. Despite the fact that an actual set of genotypes visited by an evolving population is rarely neutral, Aguirre suggests that nearly neutral mutations often increase the adaptive ability of finite populations, similarly to the model of Eigen and Schuster of quasi-species [2,3]. Because this old model of quasi-species offers the view of a large group of related genotypes in an environment with high mutation rate that may lead to hypercycles [2,3], it is now commonly assumed that the evolution of the first genome is based on the physicochemical properties of the amino acids that compose their proteins (i.e., their inertness [29], structural complexity [30,31] and stereochemical affinities between peptides and small RNAs [15,32,33]), that may have co-evolved with metabolic pathways [34] (e.g., N-fixation [35]) to parallel the metabolism of amino acids [36,37].

1.2. Stereo-Chemical Theory of Singular Docosameric Sequences

Forterre and his coworkers [24,25] have emitted the idea that Archaea and Bacteria have a common ancestor he called LUCA (for Last Universal Common Ancestor) and that the co-evolution of Archaea, Bacteria and Eukarya have been driven in part by Viruses (their phages and the giant viruses): "selection of different parts of the ancestral virosphere at the onset of the three domains played a critical role in shaping their respective biology. Eukarya probably evolved toward complexity with the help of retroviruses and large DNA viruses, whereas similar selection pressure (thermo-reduction) could explain why the archaeal and bacterial mobilomes somehow resemble each other". Staley [26] proposed in the framework of his Domain Cell Theory about the evolution of Archaea, Bacteria and Eukarya, that "when the domains of life evolved, each of the three domains evolved from separate and unique cellular lineages".

A recent model [38] in the spirit of the Eigen-Schuster's hypercycles [2,3] presents a plausible co-evolution of two types of replicating molecules, denoted P and Q: P can represent the primitive RNA molecules and Q the primitive DNA molecules. The dynamical evolution of P and Q leads to a stable stationary state, in which coexist the functional species P (ancestor of enzymes) and the informative species Q (ancestor of DNA). We will search in the following two analogues of P and Q, from the same primitive nucleotide sequences, one P in ring form able to catalyze the peptide formation from amino acids present among the early life components, and the second Q in hairpin form with the maximal thermo-dynamical stability able to store the memory of sequence P and being in chemical equilibrium with it.

All the previous theoretical hypotheses and experimental observations made it possible to develop present theories on the origin of protein translation in which the central role of hairpin or ring RNAs [39–41] was postulated and then supported by a large amount of research on the available genomic databases. In the present paper, dealing with the genomes (notably tRNAs) of plants, human, Archae and giant viruses, we will focus on the construction of an ancestral hairpin/ring called AL (for Ancestral Loop or ALPHA structure), which could have played an important role at the origin of the protein translation and we show the existence of relics of AL in present genomes, in a hypothetical framework consistent with the theories previously mentioned. The lack in our current knowledge of satisfactory models prompted us to find what kind of rationale could enable to reconcile a common ancestry to all present genomes. The bioinformatics tools for analyzing RNA, DNA and protein sequences, coupled with the availability of complete genome sequences [42–52], and dedicated

high-throughput databases, enabled us to postulate and decipher a sequence candidate, which could have played a role at the start of evolution. By performing a step-by-step computation of the smallest possible RNA sequence compatible with building up a variety of peptides, a small RNA model was identified called AL (ALPHA ring) thanks to a constraint propagation algorithm [39,40] searching a Hamiltonian path in a graph (Figure 1a) with the following properties:

- All dinucleotides should appear at least once (apart CG, because of CG suppression).
- Among the rings satisfying the principle "to be as short as possible and containing at least one codon of each amino acid", there is no solution for a length below 22 nucleotides. For the length 22, 29, 520 solutions (i.e., about 10^{-9} of the possible solutions) contain only one repeated codon AUN, N being G for 52% of the solutions.
- From these 29,520 solutions, 25 rings only satisfy with the formation of a hairpin 9-nucleotide-long or greater.
- From these 25 rings, 19 encompass both a start and a stop codon.
- Through the calculation of several distances (e.g., circular Hamming distance, permutation distance and edit distance), one singular ring (ALPHA ring) exhibits a minimum average distance as compared to the others. Only this sequence is thus acting as the barycenter of the set of the 18 others (Figure 1b).

The AL sequence is the following:

5′-AUGGUACUGCCAUUCAAGAUGA-3′

Using the Kinefold®algorithm [46], it is possible to exhibit the most thermodynamically stable hairpin formed from AL ring (Figure 1c) and an analysis of the primary structure of tRNAs like GlytRNA^{GCC} reveals that three well-characterized oligomers: a hexamer CUGCCA that is usually known as an anticodon loop, and two heptamers UUCAAGA and AAUGGUA, are respectively known as the Tψ-loop and the D-loop (cf. Figure 1d,e) of a high percentage of tRNAs (cf. the supplementary material 1).
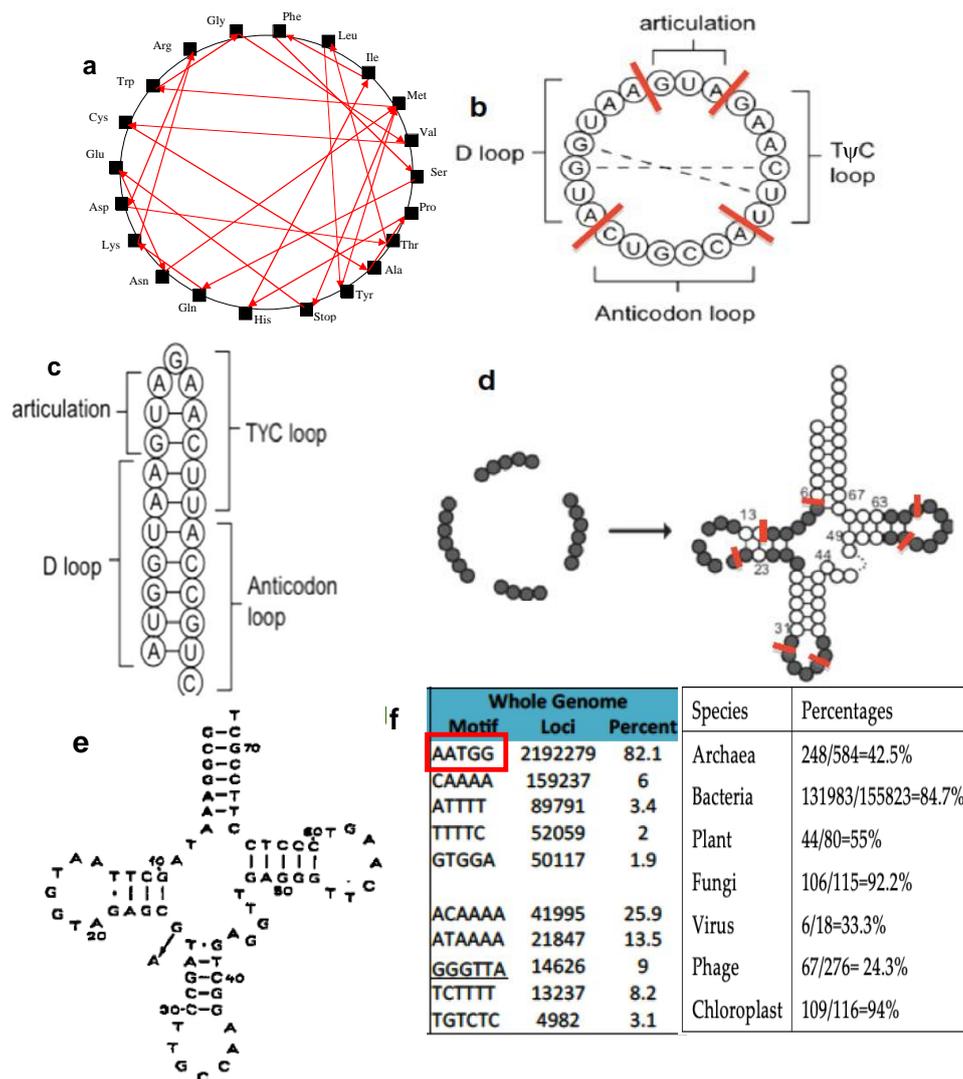
Figure 1. Computation of the Archetypal Loop AL. (**a**) Hamiltonian path in the graph having amino acids as vertices; (**b**) AL ring AUGGUACUGCCAUUCAAGAUG; (**c**) Optimal hairpin form of AL; (**d**) Urancestral tRNA-Gly [6], where nucleotides common with AL are indicated in black. (**e**) GlytRNA^GCC of *Œnothera coquimbensis* [53], whose loops (D-, anti-codon, articulation and TΨ-loops) fit quasi-perfectly AL; (**f**) pentamer frequencies in whole human genome [42] and percentages of tRNAs containing TGGTA and TTCNA in their D- and TΨ-loops, among tRNAs having NTGCCAN as an anticodon loop in different species of the tRNADB-CE database [51].

2. Materials and Methods

Searching the genomes of plants [52,53], human [54], Archaea [55–57], and giant viruses [58,59], we further searched the fitness of this ALPHA sequence, as a prototype, which could have played an important role at the origin of the protein translation. The recent discovery by Claverie and Abergel of three new strains of pandoraviruses solved the unresolved question on a common ancestor to the *Pandoraviridae* family [58]. Spontaneous neogenesis partially alleviates this new mystery but yet does not explain the diversity of those neogenes of unknown functions. Complying with previous theories including that built on Eigen and Schuster's model, we made the hypothesis that the proximity to a putative archetypal genome at the origin of life could explain the specificity of the Pandora family with respect to the rest of the whole giant virus set and the classification of Archaea [55–57], and measured the singular proximity of ALPHA by assessing the presence of oligomers derived from

ALPHA. This approach offered the possibility to confront the optimal properties of the ALPHA ring and to search for the existence of relics of ALPHA in present genomes.

We calculated the probability P_{22} of occurrence of the 22 pentamers derived from ALPHA ($100 \times P_{22}$ is called the ALPHA pentamer-proximity noted P_{22}) in the genomes of giant viruses, and compared it to 0.0215 + 0.0024 * as being the probability of random occurrences of a pentamer from ALPHA sequence (denoted AL5) for a genome of size 1,000,000 nucleotides-long (* indicating the upper limit of the 95%-confidence interval of the right-tailed test of significance of an empirical frequency). As shown in Figure 2, we observed two groups. In one group, which figures are noted in blue, the probability P_{22} was similar to those of a random occurrence ($P_{22} \leq 0.0222$). The group contained in particular the large family of Pandora viruses having a tendency to create neogenes (93% of their genes differ from all other known viral or cellular genes [58]). In the second group, in which the figures are noted in green, P_{22} was above 0.025 (significantly different from the purely random probability with $p < 0.5 \times 10^{-124}$).
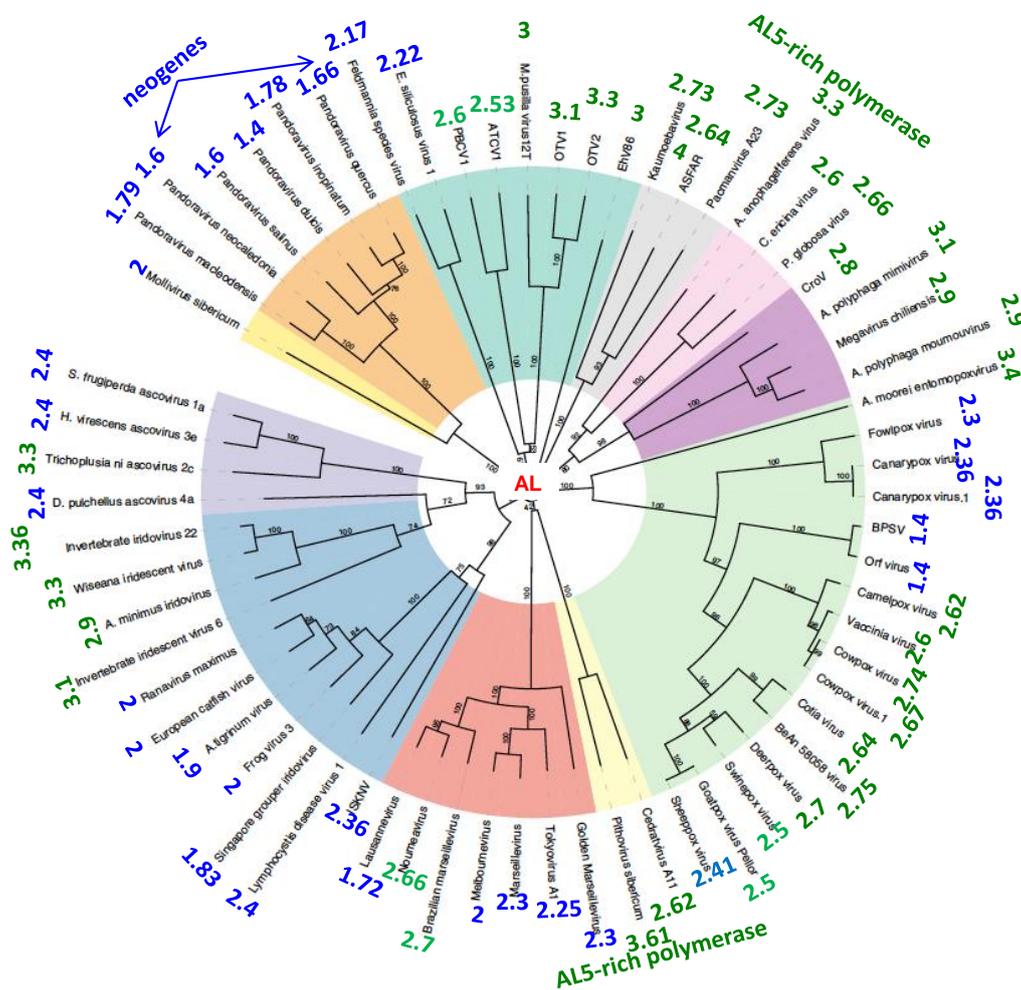

Figure 2. Giant viruses classification tree. The numbers at the periphery of the circular tree (from [58]) indicate the ALPHA pentamer-proximity P_{22} of the Giant viruses genomes to the ALPHA sequence.

3. Results

The search for ALPHA relics consisted in identifying in present genomes from different databases [42–52], the footprints of some ALPHA sub-sequences. By doing so, some remarkable facts arose such as the high similarity of ALPHA-derived sequences with the consensus sequences at exon/intron boundaries in complex eukaryotes, typically GGTAAGT and TTCAAG. Moreover, the footprints often matched with RNA genes, typically tRNAs, rRNAs, miRNAs,

and circRNAs (as assessed using NCBI sources, and GtRNAdb, miRBase and circBase, see Supplementary material 4).

Further sequence analysis of the primary structure of ALPHA revealed three well-characterized oligomers: the hexamer CUGCCA, which is usually known as the anticodon loop of tRNA-Gly's, and the two heptamers UUCAAGA and AAUGGUA, that are respectively known as the Tψ-loop and the D-loop of most tRNAs. This observation fitted nicely with the possibility of encountering the ALPHA sequence as a possible RNA present at origin of life. Indeed, the distribution of the pentamers frequencies within the genomes of Rfam database [43] shows the greatest survival probabilities for both pentamers coming from the most stable part of ALPHA, which correspond to the parts of the D-loop and Tψ-loop of many present tRNAs. We present in Figure 1e the example casted for *Œnothera coquimbensis* [52] and in Supplementary material 2, matches found with all known tRNAs.

3.1. ALPHA Remnants in the Genome of Archaea

Like with giant viruses, we assessed any proximity of the ALPHA with the well-known Archaea realm (Figure 3). Assessing their ALPHA pentamer-proximity P_9 (equal to the percentage of occurrence in their genomes of nine pentamers issued from ALPHA: ATTCA, TTCAA, TCAAG, CAAGA, AAGAT, AGATG, GATGA, ATGAA, TGAAT), Archaea fell into three categories. The first group (in blue) corresponded to $0.0109 \leq P_9 \leq 0.014$ (that differs significantly from the purely random probability with $p < 4 \times 10^{-22}$, with a genome of about 160,000 nucleotides-long), the second (in red) with $1.4 < P_9 \leq 1.8$ ($p < 10^{-122}$), and the third group (in green) with $P_9 > 1.8$ ($p \approx 1.5 \times 10^{-382}$).

Figure 3. Archaea classification tree. The numbers at the periphery of the circular tree (from [55,56]) indicate the ALPHA pentamer-proximity P_9 of the Archaea genomes to the subset of nine pentamers from ALPHA: {ATTCA, TTCAA, TCAAG, CAAGA, AAGAT, AGATG, GATGA, ATGAA, TGAAT}.

3.2. ALPHA Remnants in the Genome of Bacteria

By using the ALPHA pentamer-proximity P_{22} values for the 5S ribosomal RNAs (in red), Figure 4 (see also Supplementary material 3) shows that P_{22} explains partly a bacteria phylogeny based on the sequences of the 16S ribosomal RNA genes [59] with the following descending order: actinobacteria (mean $P_{22} = 6.7$), proteobacteria (mean $P_{22} = 4.4$), firmicutes without mycoplasmataceae (mean $P_{22} = 3.9$), mycoplasmataceae (mean $P_{22} = 1$), and cyanobacteria (mean $P_{22} = 0.45$), which corresponds to the classical order based on morphology and Gram stain. The upper limit U of significance of the observed frequencies of ALPHA pentamers for a 5S sequence of length 115 is equal to the upper bound of the 95% confidence interval of the right tailed test with H0 as pure random frequency $f = 22/1024 \approx 0.0214$; then, we have: $U = f + 1.64[f(1 - f)/115]^{1/2} \approx 0.043$, corresponding to $P_{22} = 4.3$.
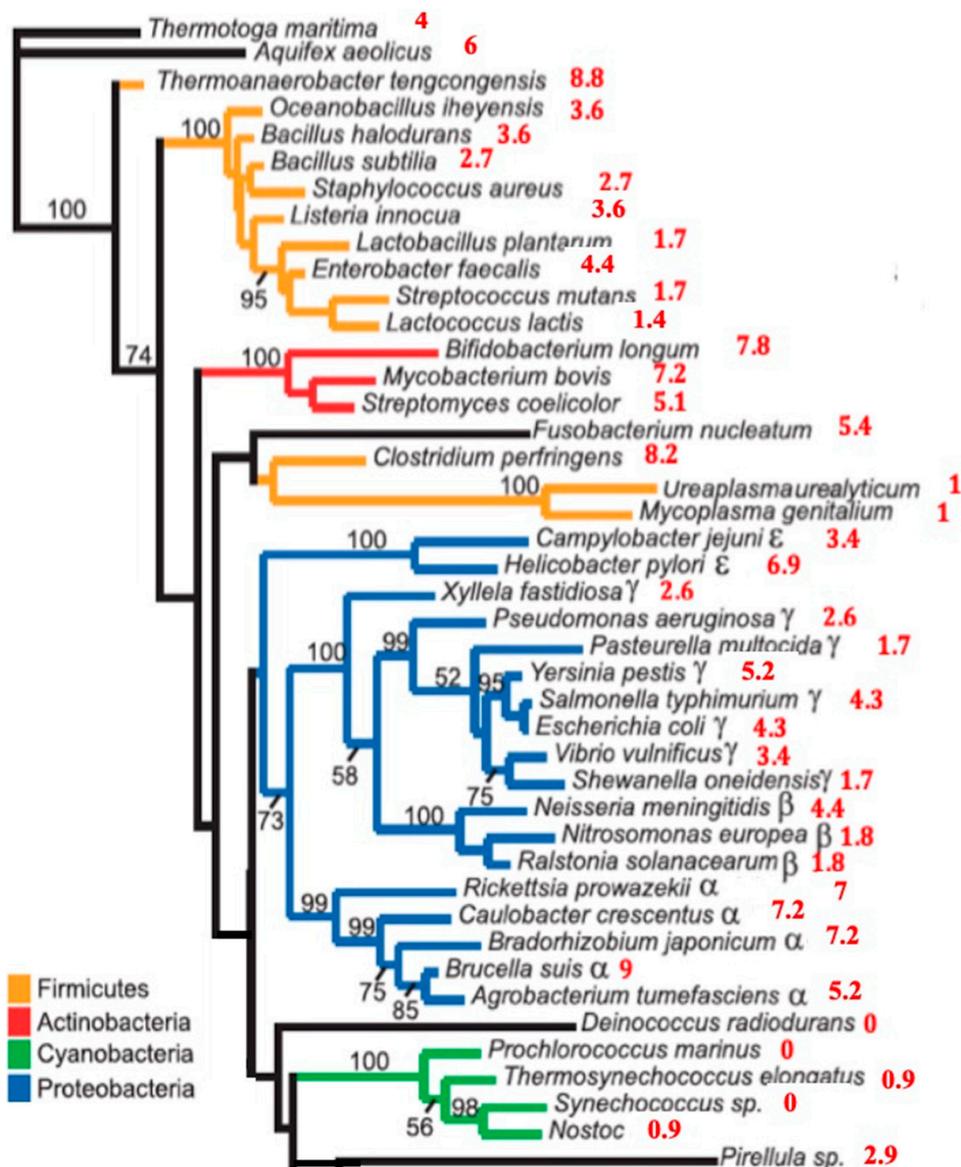
Figure 4. Bacteria phylogeny (from [60]) based on the sequences of the 16S ribosomal RNA genes. The numbers at the periphery of the phylogeny indicate the ALPHA pentamer-proximity P_{22} of the 5S ribosomal RNAs (in red).

3.3. ALPHA Remnants in Different Living Realms

Other remnants from the ring ALPHA were found by exploring the tRNA-Gly's of different living realms. As shown in Figure 5, the whole sequence of ALPHA, but with one distinct nucleotide, is encompassed within the lupine mt tRNA-Gly [60], and, as further presented in Figure 6, sequence similarity is found with 242 other vegetal species [42] and their commensal bacteria [61], and in particular the *Rickettsia prowazekii* genome (close to mitochondrial genomes), the Archae-like Halorubrum [62], and the giant-virus-like Tupanvirus, described as very ancient genomes [63], with the possibility of horizontal gene transfer between plants and bacteria [64]. Of note, the ALPHA-derived pentamer UGGUA (in bold) were also found twice within the 50 nucleotides-long sequence of the RNA catalytic domain harbored by the satellite RNA sequence of tobacco ringspot virus [65]:

5′-AAACAGAGAAGUCAACCAGAGAAACACACGUUG**UGGUA**UAUUACC**UGGUA**-3′

Likewise, a minimal RNA hairpin ribozyme, chain D, discovered 18 years later showed similarity with one ALPHA-derived hexamer and one pentamer spanning 19 bases: 5′-UCG**UGGUAC**AUUAC**CUGCC**-3′ [65]. Interestingly, the ALPHA-derived tetramer motif UGGU is a motif that is known to generally not be cleavable by ribozymes [66], and which could explain its ongoing presence in present ribozymes. Likewise, those ALPHA-derived pentamers were present in the D Chain of many hairpin ribozymes. This remnant presence of motifs could have been used to build simple RNA "cells", consisting of two ribozymes with concerted activity allowing RNA replication [67–72].
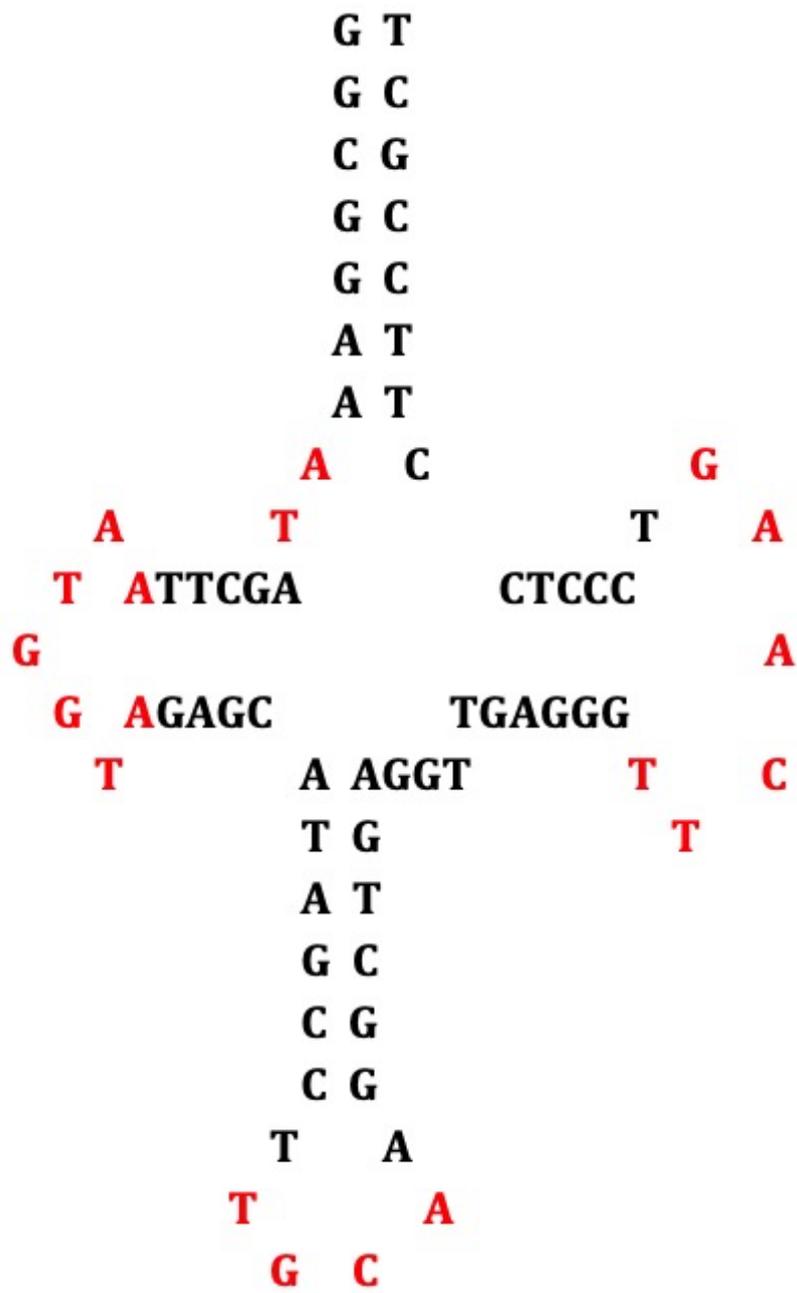
Figure 5. Lupine mitochondrial tRNA-Gly [61].

Vegetal species	Medicago truncatula mitochondrion, NCBI Reference Sequence: NC_029641.1 160317-160388 tRNA-Gly GCGGAAATAGCTT**AATGGTA**GAGCATAGCCT**TGCCA**AGGCTGAGGTTGAGGG**TTCAAG**TCCCTCCTTCCGC Medicago truncatula chloroplast NCBI Reference Sequence: NC_003119.8 99894-99964 tRNA-Gly GCGGATATAGTT**GAATGGTA**AAATTTCTCCT**TGCCA**AGGAGAAGACGCGGG**TTCGA**TTCCCGCTATCCGC Solanum tuberosum mitochondrial gene for tRNA-Gly (GCC) GenBank: AJ001728.1 GCGGAAATAGCTT**AATGGTA**AAGCATAGCCT**TGCCA**AGGCTGAGGTTGAGGG**TTCAAG**TCCCTCCTTCCGC Solanum tuberosum isolate RH89-039-16 chloroplast, tRNA-Gly GenBank: JF772171.1:37300-37370 GCGGATATGGTC**GAATGGTA**AAATTTCTCTT**TGCCA**AGGAGAAGATGCGGG**TTCGA**TTCCCGCTATCCGC Vicia faba mitochondrion, GenBank: KC189947.1 c336602-336531 tRNA-Gly GCGGAAATAGCTT**AATGGTA**GAGCATAGCCT**TGCCA**AGGCTGAGGTTGAGGG**TTCAAG**TCCCTCCTTCCGC Vicia faba plastid, complete genome GenBank: KF042344.1: c51711-51641 tRNA-Gly GCGGATATAGTTT**AATGGTA**TAATTTCTCTT**TGCCA**AGGAGAAGACGCGGG**TTCAA**TTCCCGCTATCCGC
Commensal bacteria Gram -	Sinorhizobium meliloti Rm41 complete genome GenBank: HE995405.1 167563-167637 tRNA-Gly GCGGGTGTAGTTCAG**TGGTA**GAACGCCAGCT**TGCCA**AGCTTGATGTCGTCGGT**TTCGA**TCCCGATCACCCGC Acetobacter pasteurianus subsp. pasteurianus strain SRCM101468 GenBank: CP021922.1: c537509-537436 tRNA Gly GCGGATGTAGTTCAATGGTAGAACGGCAGCT**TGCCA**AGCTGCATACGAGGG**TTCGA**TTCCCTTCATCCGC Acetobacter ascendens strain SRCM101447 chromosome, complete genome GenBank: CP021524.1 :c538411-538338 GCGGATGTAGTTCAATGGTAGAACGGCAGCT**TGCCA**AGGTGCATACGAGGG**TTCGA**TTCCCTTCATCCGC Pseudomonas sp. LAB-08 DNA GenBank: AP017423.1 : c2920760-2920685 tRNA-Gly GCGGGAAGCTCAGTTGGTAGAGCACGACCT**TGCCA**AGGTCGGGGTCGCGAGT**TCG**AGTCTCGTTTCCCGC
Proteobacteria	Rhizobium etli strain NXC12 chromosome, complete genomeGenBank: CP020906.1 : 548648-548721 tRNA-Gly GCGGGTGTAGCTC**AATGGTA**GAGCAGCAGCT**TGCCA**AGCTGAATACGAGGG**TTCGA**TTCCCTTCACCCGC Paraburkholderia phytofirmans PsJN chromosome 1, GenBank: CP001052.1 : c3267442-3267367 tRNA-Gly GCGGGAGTAGCTCAGTT**TGGTA**GAGCGCAACCT**TGCCA**AGGTTGAGGTCGCGAGT**TCG**AGCCTCGTCTCCCGC Rickettsia prowazekii str. NMRC Madrid E, complete genome GenBank: CP004888 473766-473839 tRNA-Gly (mt analogue) GCGGGTGTAGCTC**AATGGTA**GAGTTCCAGCCT**TGCCA**AGCTGGCTGTGGGG**TTCGA**TTCCCATCACCCGC
Archaea	Halorubrum saccharovorum DSM 1137 contig_70, NZ_AOJE01000070.1:63848-63918 tRNA-Gly GCGCCGGTGGTCT**AGTGGTA**GGACCTGAGCCT**TGCCA**AGCTCATGGCCCGGG**TTCAA**ATCCCGGCCGGC Halorubrum chaoviator strain DSM 19316, contig: Ga0170408_102, GenBank: FZNK01000002.1: 531509-531579 tRNA-Gly GCACCGGTAGTGTA**GTGGTA**TCACGCAACCT**TGCCA**TGGTTGCAACCCGAGT**TCAA**ATCTCGGCCGGTGC
Giant viruses	Tupanvirus deep ocean, partial genome GenBank: MF405918.1 : c297929-297858 tRNA-Gly GCTGTCGTAGTTT**AGTGGTA**GAATGGTTCGT**TGCCA**TCGAATCGGCAGGGG**TTCGA**TTCCCCTCGATAGCA

Figure 6. tRNA-Gly sequences from different living realms [42–52].

4. Discussion: The Proximity to ALPHA as Criterion of Primitivity

From our observations described above, ALPHA could well be the ancestor of tRNAs. Moreover, by searching the ALPHA proximity with additional molecules, as presented in Table 1, we found the interesting trend according to which the greatest pentameric proximity with ALPHA was encountered for those molecules that are considered as essential to the process of protein translation. Searching for primitive mechanisms of proteogenesis, Agmon [72] proposed a scenario describing the emergence of "life as we know it", i.e., "based on nucleic-acid and amino acid polymers that must include a proto-ribosome, which would have catalyzed the formation of a peptide bond between two amino acids and produced simple peptides." Aligned with this scenario and based on proximities defined throughout this work, we propose as proto-ribosome the ring/hairpin ALPHA as featured above, because it has with ribozymes, tRNAs and rRNAs a greater pentamer proximity than with other RNAs (see Table 1 and Supplementary material 2).

Other theories concerning the origin of life and ancestors of tRNA structure and function exist.

First, the circular code discovered by Arquès and Michel [73,74] contains a remarkable set X_1 of 20 highly frequent codons in RNAs involved in the protein translation:

X_1 = {**AAT (ATT), TAC (GTA), GAA (TTC), GAT (ATC)**, <u>**GCC (GGC)**</u>, <u>**GGT (ACC)**</u>, <u>**CTG (CAG)**</u>, GAC (GTC), GAG (CTC), GTT (AAC)}

Among the codons of X_1, 10 belong to ALPHA (in red) plus 4 to anti-ALPHA (in blue), and 12 of these 14 codons code all the 12 amino acids coded by X_1, which shows the close connection between ALPHA and the circular code.

Table 1. ALPHA proximity. The more the molecules considered are essential for protein translation, the greater is their ALPHA pentameric proximity P_{22} (calculated on the 22 pentamers of AL).

Molecular Family	5S RNA	SASP	Deaminase	Polymerase	cl. II Synthetase cl. I	Rprotein L7
Nb of species	100	100	50	30	3	30
Mean observed AL-proximity	4.45	4.5	4.5	3.5	3.5–2.9	3.4
Mean expected AL-proximity	2.1 ± 2 *	2.1 ± 2 *	2.1 ± 1.2 *	2.1 ± 0.9 *	2.1 ± 0.5 *	2.1 ± 1.25 *
% over max 0.90-confidence upper threshold *	42%	100%	48%	100%	96–53%	50%

* indicates the max (among species) of the 0.90-confidence upper threshold of expected P22.

Second, the tRNAs obtained from repeats like UAGCC [75] present loops close to ALPHA, with the following hexamer motifs:

- D-loop hexamer motif is CTGGTC for types I and II tRNA, and ATGGTA for ALPHA
- Anticodon-loop hexamer motif is CTanticodonA for type I tRNA and for ALPHA
- T-loop hexamer motif is TTCAAA for types I and II tRNA, and TTCAAG for ALPHA.

Hence, these hexamer motifs represent 18/22 bases in ALPHA and present similarities with types I and II evolving tRNAs of [75].

Third, in [76], the authors claimed that the protein translation mechanism emerged when the genetic code started to evolve due to a stabilizing effect on RNA–peptide complexes with bridge peptides present in polymerase motifs, like the QLSLF amino acid motif whose nucleotide sequence contains twice the ALPHA tetramer ACTG: CA(ACTG)TC(ACTG)TTC. The hybridization-induced proximity peptides of short amino-acetylated RNAs could have favored the emergence of random peptides [77,78], initiating the Darwinian transformation of the genetic code (Figure 7) in a harmonious co-evolution of the biological information and function. The present RNA ring theory is compatible with an RNA–peptides scenario in which peptides were synthesized thanks to ALPHA sequence.

Eventually, in [60], it is shown that ALPHA ring belongs to a family of ancient RNAs made from diverse RNA types, including replication origins (OL) and OL-like structure, riboswitches, ribozymes, rRNAs, and tRNAs, molecules that are presumably close to ancestral RNAs.
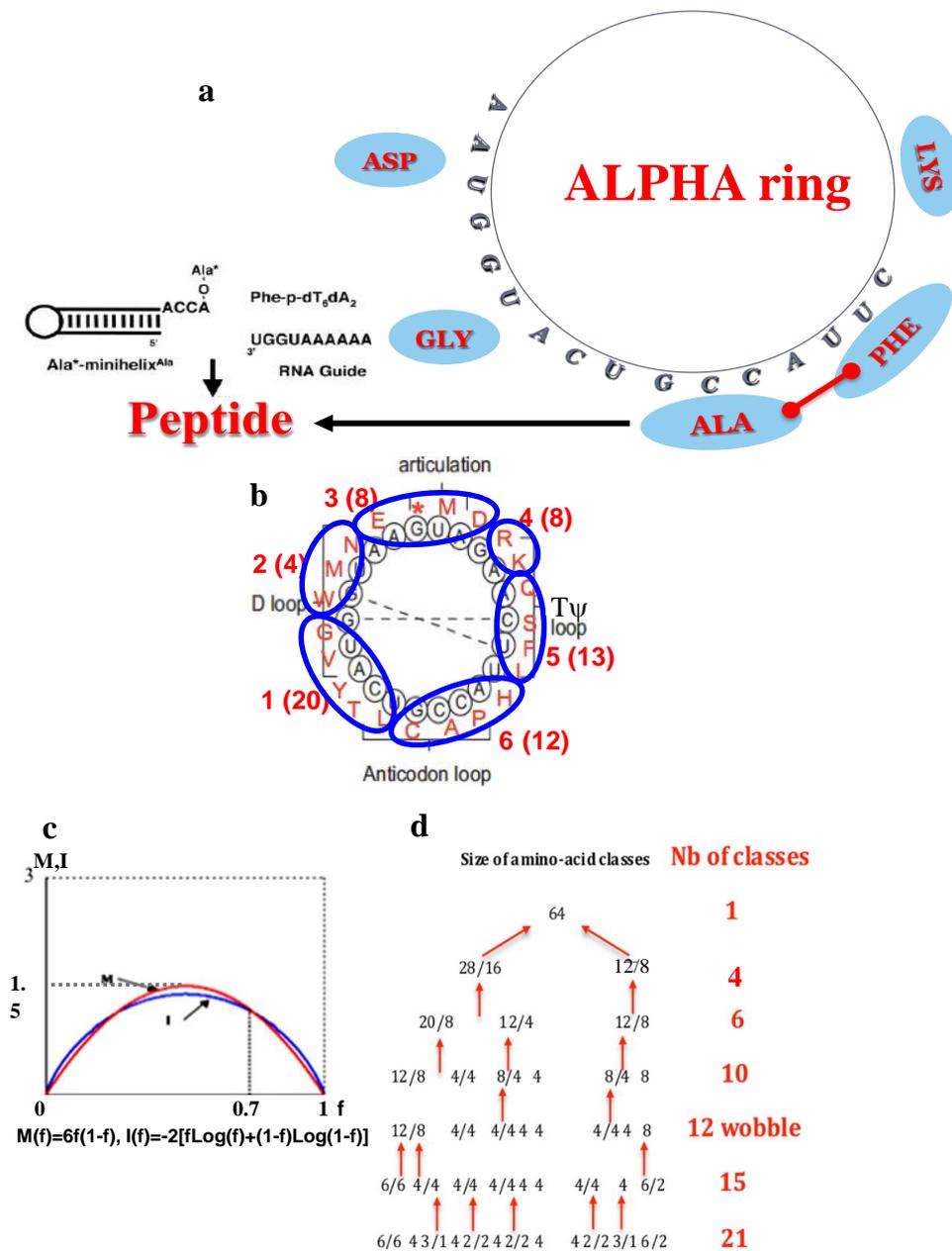
Figure 7. ALPHA ring as ribosome ancestor. (**a**) Primitive translation machinery centered on ALPHA ring with synthesis of Ala-Phe peptide on UGGUAA RNA guide [78] or on its complement in ALPHA GCCAUU. (**b**) ALPHA can be divided into six sub-sequences corresponding to the codons classes candidates for the second step of the descending partition, which follows (**c**) the min-max principle: "mean mutation error M equals information I" [79] and gives at step 4 (**d**) the "wobble" partition coding for the 11 early assigned amino acids plus a group of codons assigned to late amino acids [80–82].

5. Conclusions

To support the origin of life, a network view, as discussed by Aguirre at al. [28] and priorly by Seligmann and Raoult [83] can be proposed and centered on the ALPHA ring, as a key in the primitive machinery building peptides (Figure 7a,b). In this model, which could be viewed as a Mother Goose model, we hypothesize that the boundary of this first functional « machine », which was able to build peptides, could be defined as a peptidic gradient boundary, centered on the "proto-nucleus" ALPHA. The amino acids confinement around this protonucleus ALPHA could favor the occurrence of peptide bounds. This "organ" functioned as a "proto-ribosome" into a "proto-membrane", and thus

as a "proto-cell" with a circular organization. In fact, this model stands as a solution of a variational problem that is that peptide synthesis favored by ALPHA 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 by different works, theoretical as well as experimental:

- In 1926, H.J. Muller already suggested that life began not as an enzyme but as a gene [84]
- 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 [2], constituting the first class of amino acids (Figure 7d) selected following the min-max principle: "mean mutation error M equals information I" (Figure 7c), which uses the notion of information as proposed by Eigen [85]
- In the theory of autopoiesis [86,87], the first living system is self-reproducing [88,89] 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"
- Experimental evidence of direct RNA–amino acid interactions has been shown for Arginine by Yarus et al. [33] and Alanine by Tamura and Schimmel [90–92]
- Statistical and theoretical arguments about the role of the primitive RNAs in the progressive constitution of the genetic code are given in [93–97].

Strikingly, the unique properties exhibited by the ALPHA ring address a possible role at the origins of life, giving birth to footprints as molecular relics in the present structures involved in the ribosomal translation. As a theoretical singular prototype, this ALPHA sequence should be useful to assess as a model matrix of future applications, ranging from synthetic biology to DNA computing. The sequence ALPHA and pentamers extracted from ALPHA are indeed frequently retrieved as remnants in many genomes [40], notably in proteins essential for the protein translation and maintenance of the cell integrity (tRNA synthetases, polymerases, tRNA nucleotidyl transferases, lipids synthetases, CRISPR Cas 9, etc.), which are considered as essential building blocks for cell survival.

Further studies should experimentally investigate the ring ALPHA as a potential catalyzer of peptide synthesis and search for its role in building after the RNA world, the protein and cell worlds and its role in consolidation of the genetic code, in accordance with previous knowledge in the field [98–108].

References

1. Paecht-Horowitz, M. The origin of Life. *Angew. Chem.* **1973**, *12*, 349–356. [CrossRef] [PubMed]
2. Eigen, M.; Schuster, P. The hypercycle: A principle of natural self organization. Part C: The realistic hypercycle. *Naturwissenschaften* **1978**, *65*, 341–369. [CrossRef]
3. Eigen, M.; Winkler-Oswatitsch, R. Transfer-RNA: The early adaptor. *Naturwissenschaften* **1981**, *68*, 217–228. [CrossRef] [PubMed]
4. Bernal, D. *The Physical Basis of Life*; Routledge and Kegan Paul: London, UK, 1951.
5. Miller, S.L. A Production of Amino Acids under Possible Primitive Earth Conditions. *Science* **1953**, *117*, 528–529. [CrossRef] [PubMed]

6. 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]

7. 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]

8. 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]

9. Ponnamperuma, C.; Sagan, C.; Mariner, R. Synthesis of adenosine triphosphate under possible primitive earth conditions. *Nature* **1963**, *199*, 222–226. [CrossRef]

10. 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]

11. 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]

12. Woese, C.R. A New Biology for a New Century. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 173–186. [CrossRef] [PubMed]

13. Shapiro, R. Small Molecule Interactions were Central to the Origin of Life. *Q. Rev. Biol.* **2006**, *81*, 105–126. [CrossRef] [PubMed]

14. 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] [PubMed]

15. Yarus, M. The Genetic Code and RNA-Amino Acid Affinities. *Life* **2017**, *7*, 13. [CrossRef] [PubMed]

16. Yarus, M. Eighty routes to a ribonucleotide world; dispersion and stringency in the decisive selection. *RNA* **2018**, *24*, 1041–1055. [CrossRef]

17. Lancet, D.; Zidovetzki, R.; Markovitch, O. Systems protobiology: Origin of life in lipid catalytic networks. *J. R. Soc. Interface* **2018**, *15*, 20180159. [CrossRef]

18. Norris, V.; Raine, D. A fission–fusion origin for life. *Orig. Life Evol. Biosph.* **1998**, *28*, 523–537. [CrossRef]

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

20. Norris, V.; Loutelier-Bourhis, C.; Thierry, A. How did Metabolism and Genetic Replication Get Married? *Orig. Life Evol. Biosph.* **2012**, *42*, 487–495. [CrossRef]

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

22. Damer, B.; Deamer, D. The hot spring hypothesis for an origin of life. *Astrobiology* 2019. [CrossRef] [PubMed]

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

24. 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]

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

26. 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]

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

28. 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]

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

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

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

32. Yarus, M.; Christian, E.L. Genetic code origins. *Nature* **1989**, *342*, 349–350. [CrossRef] [PubMed]

33. Yarus, M.; Widmann, J.J.; Knight, R. RNA-amino acid binding: A stereochemical era for the genetic code. *J. Mol. Evol.* **2009**, *69*, 406–429. [CrossRef] [PubMed]

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

35. Elizondo Barron, J.; Pasini, R.J.; Davis, D.W.; Stuthman, D.D.; Graham, P.H. Response to selection for seed yield and nitrogen (N2) fixation in common bean (*Phaseolus vulgaris* L.). *Field Crop. Res.* **1999**, *62*, 119–128. [CrossRef]

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

37. 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]

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

39. Weil, G.; Heus, K.; Faraut, T.; Demongeot, J. An archetypal basic code for the primitive genome. *Theor. Comput. Sci.* **2004**, *322*, 313–334. [CrossRef]

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

41. Di Giulio, M. A Comparison among the Models Proposed to Explain the Origin of the tRNA Molecule: A Synthesis. *J. Mol. Evol.* **2009**, *69*, 1–9. [CrossRef]

42. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2016**, *44*, D7–D19. Available online: https://www.ncbi.nlm.nih.gov/nucleotide/ (accessed on 22 April 2020). [CrossRef] [PubMed]

43. Kalvari, I.; Argasinska, J.; Quinones-Olvera, N.; Nawrocki, E.P.; Rivas, E.; Eddy, S.; Bateman, A.; Finn, R.D.; Petrov, A.I. Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **2018**, *46*, D335–D342. Available online: https://rfam.xfam.org/ (accessed on 22 April 2020).

44. Chan, P.P.; Lowe, T.M. GtRNAdb 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **2016**, *44*, D184–D189. Available online: http://gtrnadb.ucsc.edu/GtRNAdb2/genomes/bacteria/Rhod_spha_ATCC_17029 (accessed on 22 April 2020). [CrossRef] [PubMed]

45. Szymanski, M.; Zielezinski, A.; Karlowski, W.M. 5SRNAdb: An information resource for 5S ribosomal RNAs. *Nucleic Acids Res.* **2016**, *44*, D180–D183. Available online: http://www.combio.pl/rrna/alignment/ (accessed on 22 April 2020). [CrossRef] [PubMed]

46. Xayaphoummine, A.; Bucher, T.; Isambert, H. Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.* **2005**, *33*, W605–W610. Available online: http://kinefold.curie.fr/ (accessed on 22 April 2020). [CrossRef]

47. Jühling, F.; Mörl, M.; Hartmann, R.K.; Sprinzl, M.; Stadler, P.F.; Pütz, J. tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* **2009**, *37*, D159–D162. Available online: http://trnadb.bioinf.uni-leipzig.de/DataOutput/ (accessed on 22 April 2020). [CrossRef] [PubMed]

48. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. Available online: https://www.arb-silva.de/ (accessed on 22 April 2020).

49. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol. Biol.* **2019**, *1962*, 1–14. Available online: http://lowelab.ucsc.edu/tRNAscan-SE/ (accessed on 22 April 2020).

50. Griffiths-Jones, S.; Saini, H.K.; van Dongen, S.; Enright, A.J. miRBase. 2008. Available online: http://www.mirbase.org/ (accessed on 22 April 2020).

51. Abe, T.; Ikemura, T.; Ohara, Y.; Uehara, H.; Kinouchi, M.; Kanaya, S.; Yamada, Y.; Muto, A.; Inokuchi, H. tRNADB-CE: tRNA gene database curated manually by experts. *Nucleic Acids Res.* **2009**, D163–D168. Available online: http://trna.ie.niigata-u.ac.jp/cgi-bin/trnadb/index.cgi (accessed on 22 April 2020).

52. Cognat, V.; Pawlak, G.; Duchêne, A.M.; Daujat, M.; Gigant, A.; Salinas, T.; Michaud, M.; Gutmann, B.; Giegé, P.; Gobert, A.; et al. PlantRNA, a database for tRNAs of photosynthetic eukaryotes. *Nucleic Acids Res.* **2013**, *41*, D273–D279. Available online: http://plantrna.ibmp.cnrs.fr/ (accessed on 22 April 2020). [CrossRef]

53. 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]

54. 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] [PubMed]

55. 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] [PubMed]

56. 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] [PubMed]

57. 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]

58. 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]

59. 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]

60. 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]

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

62. 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]

63. 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]

64. 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]

65. 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] [PubMed]

66. 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] [PubMed]

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

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

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

70. 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]

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

72. 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. [CrossRef]

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

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

75. 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] [PubMed]

76. 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] [PubMed]

77. 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] [PubMed]

78. 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]

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

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

81. 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]

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

83. Seligmann, H.; Raoult, D. Unifying view of stem-loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses. *Curr. Opin. Microbiol.* **2016**, *31*, 1–8. [CrossRef]

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

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

86. Maturana, H.R.; Varela, F.J. *Autopoiesis and Cognition: The Realization of the Living*; Reidel: Boston, MA, USA, 1980.

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

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

89. 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.

90. 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]

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

92. 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]

93. 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]

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

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

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

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

98. 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] [PubMed]

99. 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] [PubMed]

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

101. 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*, 1–23. [CrossRef]

102. Demongeot, J.; Seligmann, H. Bias for 3'-dominant codon directional asymmetry in theoretical minimal RNA rings. *J. Comput. Biol.* **2019**, *26*, 1003–1012. [CrossRef]
103. Demongeot, J.; Seligmann, H. Theoretical minimal RNA rings designed according to coding constraints mimick deamination gradients. *Sci. Nat.* **2019**, *106*, 44. [CrossRef]
104. Demongeot, J.; Seligmann, H. Pentamers with non-redundant frames: Bias for natural circular code codons. *J. Mol. Evol.* **2020**, *88*, 194–201. [CrossRef]
105. Demongeot, J.; Seligmann, H. The primordial tRNA acceptor stem code from theoretical minimal RNA ring clusters. *BMC Genet.* **2020**, *21*, 7. [CrossRef]
106. 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]
107. 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*. [CrossRef]
108. Demongeot, J.; Seligmann, H. Theoretical minimal RNA rings maximizing coding information overwhelmingly start with the universal initiation codon AUG. *BioEssays* **2020**, *42*. [CrossRef]