Review

The 3 31 Nucleotide Minihelix tRNA Evolution Theorem and the Origin of Life

Lei Lei [1] and Zachary Frome Burton [2],*

1   School of Biological Sciences, University of New England, Biddeford, ME 04005, USA; llei@une.edu
2   Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA
*   Correspondence: burton@msu.edu

Abstract: There are no theorems (proven theories) in the biological sciences. We propose that the 3 31 nt minihelix tRNA evolution theorem be universally accepted as one. The 3 31 nt minihelix theorem completely describes the evolution of type I and type II tRNAs from ordered precursors (RNA repeats and inverted repeats). Despite the diversification of tRNAome sequences, statistical tests overwhelmingly support the theorem. Furthermore, the theorem relates the dominant pathway for the origin of life on Earth, specifically, how tRNAomes and the genetic code may have coevolved. Alternate models for tRNA evolution (i.e., 2 minihelix, convergent and accretion models) are falsified. In the context of the pre-life world, tRNA was a molecule that, via mutation, could modify anticodon sequences and teach itself to code. Based on the tRNA sequence, we relate the clearest history to date of the chemical evolution of life. From analysis of tRNA evolution, ribozyme-mediated RNA ligation was a primary driving force in the evolution of complexity during the pre-life-to-life transition. TRNA formed the core for the evolution of living systems on Earth.

Keywords: accretion model; convergent evolution model; divergent evolution model; folding of first proteins; genetic code evolution; origin of life; pseudosymmetry; ribozyme/primitive catalyst RNA ligation; type I and type II tRNA evolution

1. Evolution of TRNA

A number of models have been advanced to describe tRNA evolution. We have advanced the 3 31 nt minihelix theorem [1–3]. To support minihelix replication in pre-life, 3 31 nt minihelices were joined by ligation. The D loop 31 nt minihelix had the sequence GCGGCGG_UAGCCUAGCCUAGCCUA_CCGCCGC (the _ separates distinct sequence features). The D loop minihelix is a 7 nt GCG repeat (5′-acceptor stem) linked to a 17 nt UAGCC repeat (D loop minihelix core) linked to a 7 nt CGC repeat (3′-acceptor stem). The anticodon loop and the T loop 31 nt minihelices were probably initially identical, with the sequence GCGGCGG_CCGGG_CU/???AA_CCCGG_CCGCCGC (/ indicates a U-turn; ? indicates A, G, C or U (the pre-life base remains unknown)). This is a 7 nt GCG repeat (5′-acceptor stem) linked to a 17 nt stem-loop-stem (CCGGG_CU/???AA_CCCGG) linked to a 7 nt CGC repeat (3′-acceptor stem). The only pre-life sequence ambiguities are in the 7 nt CU/???AA loops, not in the stems. After LUCA (the last universal common (cellular) ancestor), the dominant anticodon loop sequence was CU/BNNAA or CU/BNNGA (B indicates G, C or U, not A; N indicates A, G, C or U), and the dominant T loop sequence was UU/CAAAU. Moreover, 7 nt loop sequences for the anticodon loop and the T loop were separately selected in evolution because of their different placements within tRNA. The anticodon loops were selected to generate the distinct anticodons required for coding. The T loop sequence was selected to form the tRNA elbow at which the D loop and the T loop interact to form the L-shaped tRNA fold [4]. T loop C3 (tRNA-56) forms a bent Watson–Crick base pair to a D loop G (G19 in standard tRNA numbering; for historical reasons, standard tRNA numbering breaks down in the D loop and V loop). D loop G18

intercalates between T loop A4 (tRNA-57) and A5 (tRNA-58), lifting T loop A6 (tRNA-59) and U7 (tRNA-60) out of the loop. G18 was initially an A, as part of the third UAGCC repeat, but was mutated to G to support D loop–T loop interactions at the elbow. This G to A transition in the D loop is one of the very few systematic sequence changes to support the tRNA fold versus the parental 31 nt minihelix folds. In summary, tRNA evolved in pre-life from ligation of 3 31 nt minihelices followed by internal RNA processing (described below). The 31 nt minihelices that were utilized were of two distinct sequences (D loop and anticodon/T loop). Pre-life tRNA sequences were comprised of RNA repeats and inverted repeats. Post-LUCA, the dominant tRNA anticodon and T loop sequences are known because these sequences have been conserved for ~4 billion years in living organisms. Post-LUCA tRNA sequences can be recovered as typical tRNA sequences from a tRNA database [5]. The 3 31 nt minihelix tRNA evolution theorem, therefore, can readily be confirmed from typical tRNA diagrams of ancient Archaea (i.e., Pyrococcus furiosis, Sulfolobus solfataricus, Aeropyrum pernix, Staphylothermus marinus).

Alternate models for tRNA evolution are of the following types: (1) convergent models [6–17]; (2) accretion models [6–18]; (3) 2 minihelix models [6,11,12,14,16,17]; and (4) highly theoretical models [7,8,10,19,20]. First of all, no convergent model can rationally apply to tRNA evolution. To evolve, tRNA requires a divergent evolution model, in which tRNAomes (all of the tRNAs of an organism) diverged from pre-life type I and type II tRNAs by repeated duplications and re-purposing events. In a convergent model, multiple small tRNA fragments must converge on the same homologous sequence, conformation and form. For tRNA, such a convergent evolution model is untenable, and all convergent models are accretion models because bases must be added and/or subtracted to gain the final tRNA form. Because pre-life tRNAs were generated from completely ordered sequences (RNA repeats and inverted repeats), no accretion or convergent model can be descriptive [2,3,21–25]. Accretion models have reasonably been applied to describe later stages of rRNA evolution [26–28]. For tRNA, by contrast to rRNA, accretion models are convergent models that demand convergence of multiple pre-tRNAs (tRNA fragments) to similarly structured and ordered RNA sequences. For initial tRNA evolution, no accretion model is reasonable.

Multiple 2 minihelix models have been advanced for tRNA evolution. All of these models are convergent and accretion models. One disqualifying objection to 2 minihelix models is that 2 minihelix models are inconsistent with the homology of the anticodon stem-loop-stem and the T stem-loop-stem [3]. Because the 17 nt anticodon and the 17 nt T stem-loop-stems are clearly homologous (i.e., by inspection), no 2 minihelix model can possibly be correct. In a 2 minihelix model, the anticodon and T stem-loop-stems would be required to converge on a homologous conformation (i.e., compact 7 nt U-turn loop) and homologous sequence.

The uroboros (hoop snake with mouth grasping tail) model for tRNA evolution was computer-generated and remains highly theoretical [9,10]. In the uroborous vision, numerous 22 nt covalently closed tRNA rings converged and accreted to homologous and ordered tRNA forms. The uroboros model cannot be correct for tRNA evolution. Some authors employ convergent and accretion tRNA evolution models without realizing their obvious mistake [29–34].

## 2. The Evolutionary History of Life on Earth

Top-down strategies to describe pre-life evolution have the potential advantage of identifying dominant, successful pathways because only routes that survived can be detected [1,2]. The 3 31 nt minihelix theorem relates a top-down, sequence-based argument for two stages of pre-life evolution: (1) polymer world and (2) 31 nt minihelix world. The central focus of the top-down argument is that highly ordered tRNA sequences have been conserved from pre-life to life and maintained since LUCA. Bottom-up strategies would be attempts to create life or a potential pre-life chemical state in vitro [35–41]. Bottom-up strategies result in novel chemistry but may represent evolutionary dead-

ends. When top-down and bottom-up descriptions can be linked, understanding of the pre-life-to-life transition is significantly enriched. Because tRNA sequences are ordered, the 3 31 nt minihelix tRNA evolution theorem describes hundreds of millions of years of pre-life chemical evolution.

The history of the evolution of life on Earth embedded in the tRNA sequence is shown in Figures 1–4. Remarkably, tRNA evolution reduces to a simple sequence puzzle that anyone who can read a four-letter code can solve or verify. Figures 1 and 2 explain type II tRNA evolution [42]. Figures 3 and 4 explain type I tRNA evolution, which is a very similar process [3,21,23–25]. Both type II and type I tRNAs evolved from the same 93 nt tRNA precursor. Type II tRNA evolution converted the 93 nt precursor, which was formed by ligation of 3 31 nt minihelices with two distinct 17 nt core sequences, into type II tRNA through a single internal 9 nt deletion (Figure 1). Generation of a pre-life type II tRNA required only a single internal 9 nt deletion within ligated 3′- and 5′-acceptor stems, as shown (red arrows). In Figure 1, 7 nt of a 3′-acceptor stem (yellow) and 2 nt of a 5′-acceptor stem (green) were deleted, leaving a 5 nt fragment of a 5′-acceptor stem (initially GGCGG; green). The internal deletion fused a magenta segment (a 17 nt UAGCC repeat) to a green segment (a 5′-acceptor stem fragment; 5 nt). The initial type II tRNA body (lacking the 4 nt ACCA adapter) was 93 − 9 = 84 nt (so the total type II tRNA length was initially 88 nt).
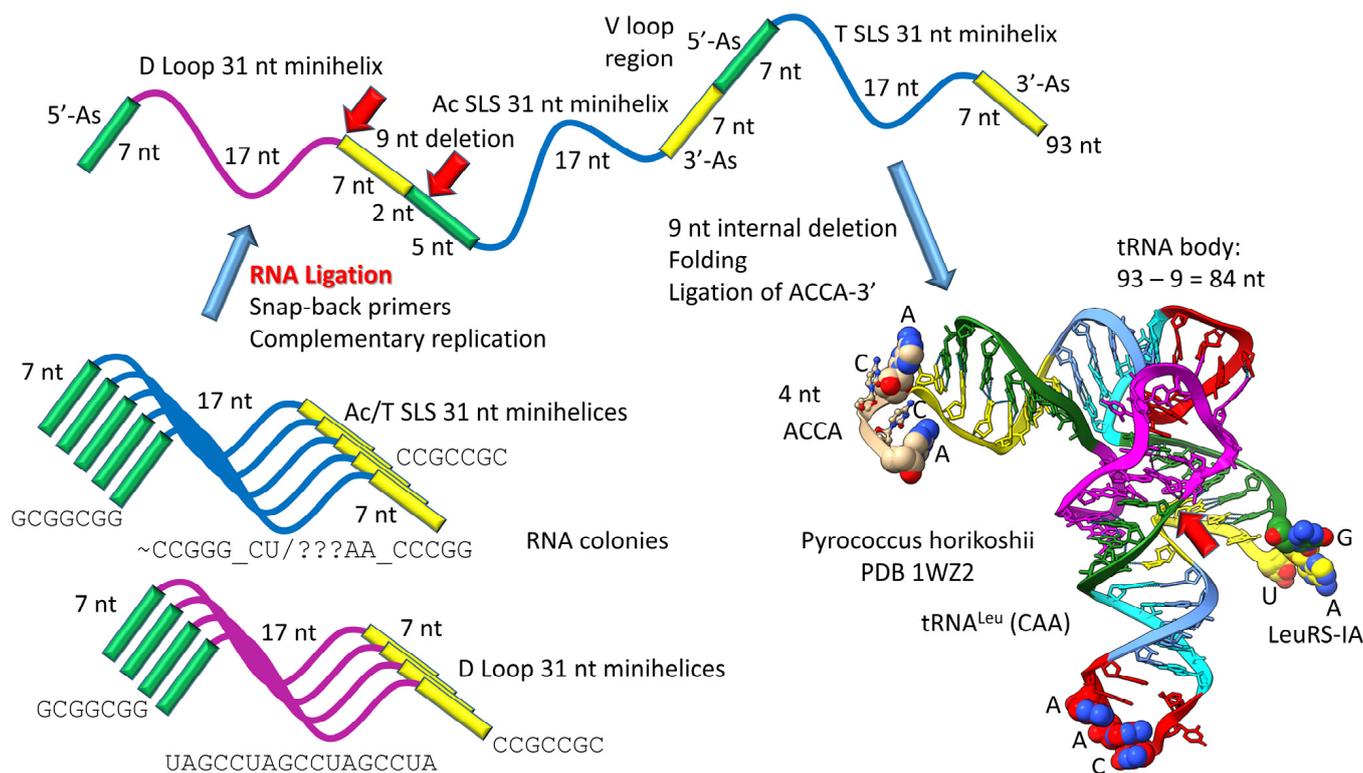

Figure 1. Type II tRNA evolved via RNA ligation and a 9 nt internal deletion within ligated 3′- and 5′-acceptor stems [42]. Also, 3 31 nt minihelices (one D loop minihelix (magenta 17 nt core) and two anticodon stem-loop-stem minihelices (blue 17 nt core)) were fused by ligation for minihelix replication. The 93 nt tRNA precursor was processed by an internal 9 nt deletion (see below) within fused 3′-acceptor (yellow) and 5′-acceptor (green) stems. In the type II tRNA structure, the red arrow indicates the fusion of the magenta segment (17 nt D loop minihelix core; UAGCC repeat) and the green segment (5′-acceptor stem fragment; initially GGCGG). Abbreviations: SLS) stem-loop-stem; Ac) anticodon. Molecular graphics were created using ChimeraX [43].
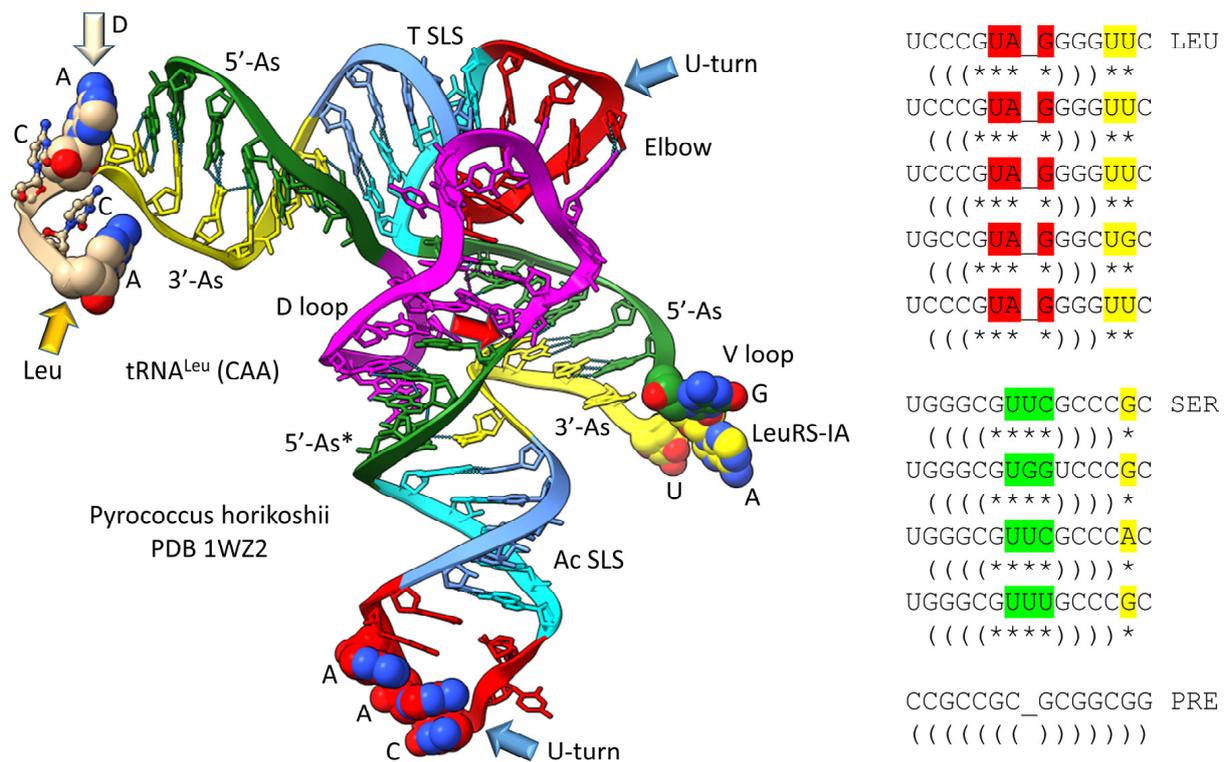
Figure 2. Type II tRNA resulted from failure to process a 14 nt V loop (initially a 7 nt 3′-acceptor stem ligated to a 7 nt 5′-acceptor stem) rather than by accretion. Colors: green) 5′-acceptor stem and 5′-acceptor stem fragment; magenta) 17 nt D loop core; cyan) 5′-anticodon and T stem; red) anticodon and T loops; cornflower blue) 3′-anticodon and T stem; and yellow) 3′-acceptor stem. Arrow colors: blue) U-turns; red) processing site in evolution; light yellow) discriminator base (D); and gold) site of amino acid placement. The structure is of an unmodified Pyrococcus horikoshii tRNA^{Leu} in complex with LeuRS-IA. At the right of the figure are tRNA^{Leu} and tRNA^{Ser} V loops from Pyrococcus furiosis, an ancient Archaeon. Colors: red) V loop UAG that binds LeuRS-IA in tRNA^{Leu} recognition in P. furiosis [44]; yellow) unpaired bases just 5′ of the Levitt base; and green) tRNA^{Ser} bases at the 3′-end of the V loop. PRE indicates an initial pre-life sequence. Parentheses indicate stems; * indicates unpaired bases.

The original pre-life type II tRNA V loop (V for variable) was a 7 nt 3′-acceptor stem (yellow) ligated to a 7 nt 5′-acceptor stem (green) (initially, CCGCCGC_GCGGCGG) [42]. Because the pre-life V loop sequence was self-complementary along its entire length, the V stem-loop-stem evolved to sequences such as those found in archaeal tRNA^{Leu} and tRNA^{Ser}. V loops for tRNA^{Leu} and tRNA^{Ser} from Pyrococcus furiosis were selected to be distinct (i.e., for separate discrimination by LeuRS-IA and SerRS-IIA) and are shown in the right panel of Figure 2, compared to the initial pre-life V loop [5]. V loops for tRNA^{Leu} are numbered V_1 to V_{14} (14 nt was the primordial length). UV_1 forms a wobble pair with G26 (standard tRNA numbering). CV_{14} forms a reverse Watson–Crick pair to G15 (the Levitt base pair; see below). Two unpaired bases are found at V_{12} and V_{13} (UU or UG). The number of unpaired bases just 5′ to the Levitt base pair determines the trajectory of the V stem-loop-stem from the tRNA body. V_6-UAG-V_8 binds the tRNA^{Leu} charging enzyme LeuRS-IA (leucine aminoacyl–tRNA synthetase; structural class and subclass IA) [44]. None of the corresponding tRNA^{Ser} V loop sequences are sufficiently similar to UAG (UUC, UGG, UUU; green shading), so LeuRS-IA will not bind to a tRNA^{Ser} V loop, even if the loop could be accessed by LeuRS-IA. For P. furiosis, the length of the tRNA^{Ser} V loop expanded from 14 to 15 nt. In tRNA^{Ser}, only 1 nt is unpaired just 5′ of the Levitt base pair. The tRNA^{Ser} V stem-loop-stem, therefore, has a different trajectory from the tRNA^{Ser} body compared to tRNA^{Leu}. SerRS-IIA binds the V stem [45], not the V loop, so SerRS-IIA is dependent

on the trajectory of the V stem-loop-stem from the tRNA[Ser] body compared to the distinct trajectory of the V stem-loop-stem of tRNA[Leu]. In summary, type II tRNA evolution is described to the last nucleotide by the 3 31 nt minihelix theorem. The longer V loop in type II tRNAs was initially generated by a failure to process ligated 3'- and 5'-acceptor stems, rather than by insertion of bases (accretion).
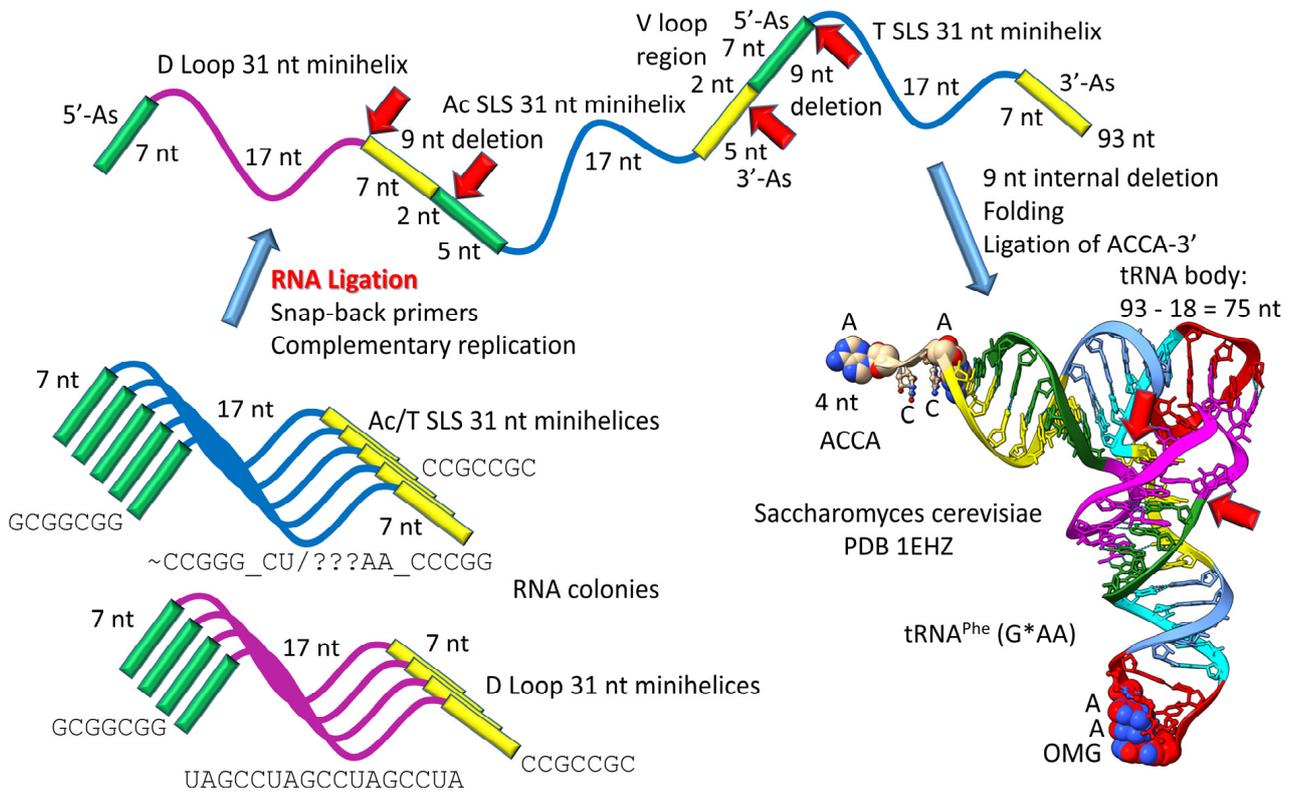
Figure 3. Evolution of type I tRNA via RNA ligation and two related, internal 9 nt deletions. Colors and arrow colors are as in Figures 1 and 2. G* (OMG) is 2'-O-methyl-G. Also, 9 nt internal deletions generate a magenta-green fusion and a yellow-cyan fusion (red arrows).

Most tRNAs are type I, with a shorter V loop (i.e., 5 nt) compared to type II tRNAs (initially 14 nt). Notably, the 3 31 nt minihelix theorem describes type I tRNA evolution to the last nucleotide (Figures 3 and 4). In Figure 3, type I tRNA evolution is shown from the same 93 nt tRNA precursor. In this case, two closely related 9 nt deletions occurred within ligated 3'- and 5'-acceptor stems. The more 5' internal 9 nt deletion was identical to the processing event that generated type II tRNAs (Figure 1). The more 3' 9 nt deletion removed 7 nt of the 5'-acceptor stem (green) and 2 nt of the 3'-acceptor stem (yellow), leaving 5 nt of the 3'-acceptor stem (initially CCGCC; yellow). Thus, in type I tRNA evolution, CCGCC (yellow) was linked to CCGGG (cyan). In type I tRNA, the more 5' and 3' 9 nt deletions within ligated acceptor stems are closely related. If the deletions occurred on complementary RNA strands, the more 3' 9 nt deletion was identical to the more 5' 9 nt deletion. Below, we suggest a ribozyme/primitive catalyst processing mechanism to describe the internal 9 nt deletions. In summary, type II and type I tRNAs evolved by a common mechanism that involved ligation of the same 3 31 nt minihelices to form the same 93 nt tRNA precursor followed by internal processing and deletion of complementary 9 nt segments. The original type I tRNA was 93 − 18 = 75 nt (tRNA body) plus the 4 nt ACCA adapter sequence (so 79 nt total). It is likely that the first tRNAs were mixtures of type II and type I tRNAs.
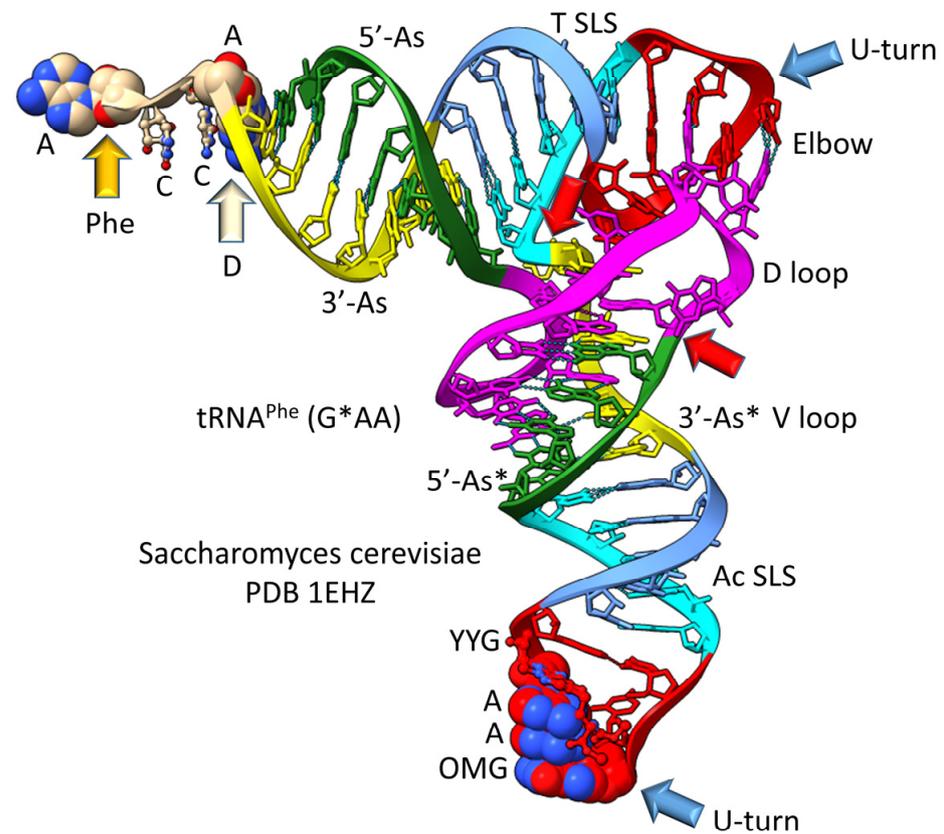
Figure 4. Type I tRNA. Colors and arrow colors are as in Figures 1 and 2. G* (OMG) is 2′-O-methyl-G. YYG is Wy-butosine [46]. The V loop (3′-As*; yellow) is fused to the cyan (5′-T stem), in slight contrast to type II tRNA processing (Figures 1 and 2).

3. The 3 31 nt Minihelix Theorem

The 3 31 nt minihelix theorem is shown in Figure 5 as a linear sequence. The inset in the figure shows a mechanism to generate the 5′- (type I and type II tRNAs) and 3′- (type I tRNA only) acceptor stem fragments found in tRNAs. The proposed mechanism involved a ribozyme/primitive catalyst endonuclease to cleave RNAs at stem-loop boundaries followed by RNA ligation. RNA stem-loop-stems cannot maintain a 2 nt loop because a 2 nt loop is too constrained, so 4 nt loops were the expected substrates for processing (as shown).

Embedded in tRNA sequences is evidence for the minihelix world and, before that, the polymer world [1,2,21]. Surprisingly, precursor tRNA sequences from pre-life were highly ordered: RNA repeats and inverted repeats [3,24,25,42]. Initially, this was a shock to us because we thought tRNA was generated by a chaotic process. Inspection of tRNA sequences, however, shows that pre-life tRNA precursors were ordered. Because tRNAs were generated from ordered sequences, pre-life tRNA precursors were arranged as RNA repeats and inverted repeats, as shown. Analysis of tRNA sequences, therefore, reveals a history of pre-life worlds on Earth. It is not reasonable to consider that ordered tRNA precursors were generated by any chaotic process. We stress that these recovered sequences from living organisms are fossils from the pre-life-to-life transition on Earth ~4 billion years ago.
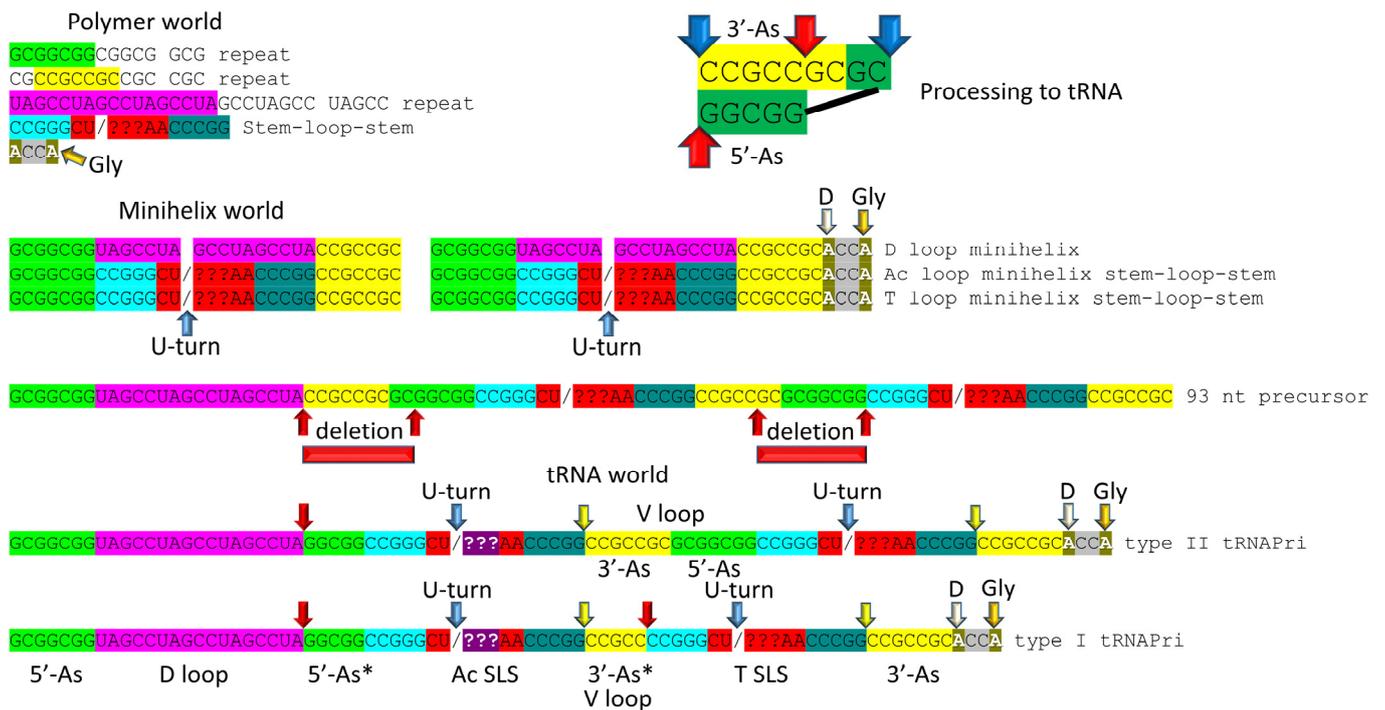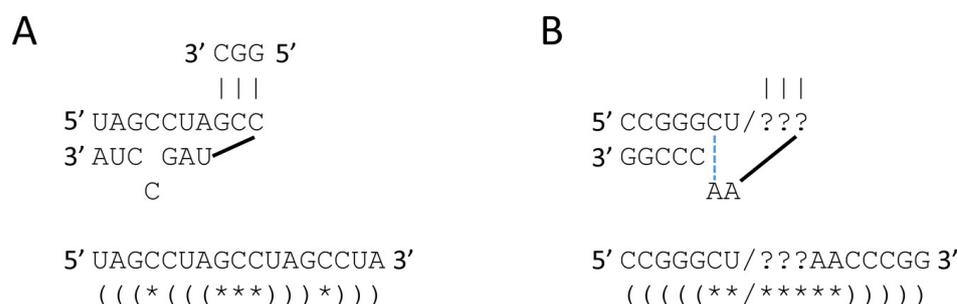
Figure 5. The 3 31 nt minihelix theorem, and evolution of tRNA world from polymer world and 31 nt minihelix world. The inset describes the 9 nt deletions to generate tRNAs: the more 5′ processing event involves deletion between the blue arrows; the more 3′ processing event (type I tRNA only) involves deletion between the red arrows. Internal deletions were at stem-loop junctions. Colors and arrows are consistent with previous figures. Yellow arrows mark the cornflower blue-yellow junction, indicating the degree of order in tRNA assembly.

4. Polymer and 31 nt Minihelix Worlds

In pre-life, the polymer world, minihelix world and tRNA world were overlapping and more complex than can now be described just from the tRNA sequence. The surviving history was limited to tRNA sequences. This is a limitation of top-down and sequence-based analyses of pre-life. The polymer world included GCG repeats, CGC repeats and UAGCC repeats, which are sequences found in tRNAs. We posit that the polymer world included ACCA as the most primitive RNA-amino acid adapter. Attaching glycine at the 3′-end, ACCA becomes ACCA-Gly, which can be used for polyglycine synthesis. A GCG repeat includes multiple CGGC sequences, which can anneal with ACCA-Gly for polyglycine synthesis. Hydration–dehydration cycles may have supported polyglycine synthesis with multiple ACCA-Gly immobilized in proximity [38,47]. Phase separations supported by water, supercritical CO_2 and lipid within the Earth's crust under high pressure may also have contributed to polypeptide polymerization reactions [39–41].

Figure 6 shows some of the features expected in polymer and 31 nt minihelix worlds. Figure 6A shows that the D loop minihelix core is self-complementary and can present anticodon GCC to pair codon GGC. GGC is found in the GCG repeat identified in the polymer world from analysis of tRNA sequences (Figure 5). Figure 6B shows the anticodon and T stem-loop-stem sequences. The codon recognized would depend on the ??? sequence (i.e., anticodon GCC would recognize codon GGC). If ACCA-Gly were ligated at the RNA 3′-ends, these sequences could be utilized for polyglycine synthesis. In pre-life, the stem-loop-stem sequence CCGGG_CU/???AA_CCCGG indicates that some form of complementary replication was present in the polymer world. Also, a stem-loop-stem can function as a primer for complementary replication. We posit that the anticodon stem-loop-stem was the most central intellectual property in pre-life on Earth. The 7 nt loop sequence (CU/???AA) allowed tRNA to learn to code.

A

```
              3' CGG 5'
                 |||
      5' UAGCCUAGCC
      3' AUC  GAU
           C
```

```
5' UAGCCUAGCCUAGCCUA 3'
   (((*(((***)))*)))
```

B

```
                    |||
      5' CCGGGCU/???
      3' GGCCC
                  AA
```

```
5' CCGGGCU/???AACCCGG 3'
   (((((**/*****)))))
```

Figure 6. Features of polymer world. (**A**) The D loop minihelix core could function as a primitive translational adapter to recognize codon GGC. (**B**) The anticodon and T stem-loop-stems could function as a translational adapter. The dotted blue line indicates a Hoogsteen A–C pair that stabilizes the U-turn loop. Ligation of 3′-ACCA-Gly converted these sequences into primitive translational adapters in the pre-life world. Parentheses indicate paired bases. Asterisks indicate loop bases.

We posit that the minihelix world evolved and was selected alongside the polymer world as an improved means to generate polyglycine [37,48]. When we invoke polyglycine or GADV world, we do not mean to imply that the synthesis of these sequences was particularly accurate. Other available amino acids could have been incorporated, as well. Translational fidelity coevolved with the genetic code only as sequence-dependent proteins became more strongly protected by selection. Judging from the tRNA sequence, the minihelix world was the clear precursor of the tRNA world because tRNA was evolved by ligation of 3 31 nt minihelices of almost completely known sequence (Figures 1–5). To generate the minihelix world from the polymer world required a small number of ribozymes/primitive catalysts, most or all of which have been generated by scientists in vitro. Similarly, a small set of ribozymes/primitive catalysts would be sufficient to convert the minihelix world into the tRNA world. Most likely, these conversions could be reproduced in vitro.

Organisms have lived in the tRNA world for about 4 billion years. The advantage of the tRNA world over the polymer and minihelix world was, initially, that tRNA evolved and was selected as an improved means to synthesize polyglycine as a component of proto-cells [1,2]. From the tRNA–polyglycine world, the tRNA–GADV world emerged [36,49–52]. From an 8 amino acid stage (i.e., GADVLSER), the genetic code and sequence-dependent proteins emerged [1,2,21]. Thus, no chicken and egg problem need be invoked in the evolution of the genetic code because the system did not need the foresight that it would eventually encode sequence-dependent proteins. Polyglycine and GADV polypeptides were selected for coacervate functions (i.e., supporting membrane-less organelles in proto-cells) [1,2,53]. With added complexity, the genetic code and sequence-dependent proteins evolved and were selected and coevolved with tRNAomes [21,23]. Once the genetic code evolved, living systems began.

5. Arguments for and against the 3 31 nt Minihelix Theorem

For reasons that we do not understand, the 3 31 nt minihelix theorem, which is fully supported by the tRNA sequence, has not gained universal acceptance. The 3 31 nt minihelix theorem is important because it relates the history of the pre-life-to-life transition on Earth. We understand that there are competing tRNA evolution models, but alternate models are falsified. The competing models are all chaotic, convergent, accretion and 2 minihelix models. None of these models can possibly be reasonable.

Statistics strongly favor the 3 31 nt minihelix theorem (Table 1) [24]. Remarkably, every feature of the theorem was predicted and is supported by statistical analysis. As the analysis was performed, a *p*-value of 0.001 indicated a similar sequence with a 1 in 1000 chance of being due to random chance. So, a *p*-value of 0.001 indicates homology. A *p*-value approaching 1 indicates sequences are not homologous. A *p*-value of <0.05 would indicate similarity and probable homology. First of all, as expected, the 5′-acceptor stem is

apparently homologous to the complement of the 3′-acceptor stem with which it pairs. Also, as expected, the 3′-acceptor stem is homologous to the complement of the 5′-acceptor stem. The demonstrated complementarity of the 5′- and 3′-acceptor stems appears to partly verify the statistical test. The 5′-acceptor stem fragment tests as homologous to the 5′-acceptor stem, positions 3–7. The 3′-acceptor stem fragment (type I V loop) tests as homologous to the 3′-acceptor stem, positions 66–70, as expected from the theorem. The 17 nt anticodon stem-loop-stem tests as homologous to the 17 nt T stem-loop-stem, as anticipated from the theorem. As expected, neither the 17 nt anticodon stem-loop-stem nor the 17 nt T stem-loop-stem test is homologous to the 17 nt D loop core (initially, a UAGCC repeat). Both the type II tRNA^{Leu} and tRNA^{Ser} V loops test as homologous to a 3′-acceptor stem ligated to a 5′-acceptor stem (Table 1) [42]. In summary, every aspect and prediction of the 3 31 nt minihelix theorem is reinforced by statistical tests. The theorem completely describes the evolution of type II and type I tRNAs to the last nucleotide.

Table 1. Internal homologies within archaeal tRNAs. C indicates a sequence complement. In type II V loops, n indicates the total length of the V loop. In type II tRNAs, V_1 was selected to be U to form a $G26{\sim}UV_1$ wobble pair. V_n was selected to be C to form the $G15{-}CV_n$ reverse Watson–Crick Levitt base pair. Similar statistics were obtained for Bacterial tRNAs [24,42].

Sequence #1	Sequence #2	Length (nt)	*p*-Value
5′-As	3′-As-C	7	0.001
3′-As	5′-As-C	7	0.001
5′-As*	5′-As (3 to 7)	5	0.001
3′-As* (V loop)	3′-As (66 to 70)	5	0.001
Ac SLS	T SLS	17	0.001
D loop	Ac loop	17	0.979
D loop	T loop	17	~1
3′-As and 5′-As (66 to 71 and 2 to 7)	V-loop Leu V_2 to V_7 and V_{n-7} to V_{n-1}	12	0.001
3′-As and 5′-As (66 to 71 and 2 to 7)	V-loop Ser V_2 to V_7 and V_{n-7} to V_{n-1}	12	0.001

Typical tDNA sequences from the tRNA database are shown in Figure 7 for *Pyrococcus furiosis* (Figure 7A) and a large collection of Archaea (Figure 7B) [5]. The greener image in Figure 7A indicates a stronger consensus, indicating that some Archaea are more derived from LUCA compared to *P. furiosis*. We consider the ancient Archaeon *P. furiosis* as a reasonable model organism for LUCA. Strangely, Di Giulio has argued that the 17 nt anticodon and 17 nt T stem-loop-stems, each with a compact 7 nt U-turn loop, with the U-turn positioned between loop positions 2 and 3, are not homologs [11]. Di Giulio's assertion is not credible. The anticodon stem-loop-stem and the T stem-loop-stem are homologous by inspection and by statistical test (Figure 7; Table 1). According to Di Giulio's tRNA evolution model, the 17 nt anticodon stem-loop-stem and the 17 nt T stem-loop-stem must take on apparent homologous sequences and common structures by convergent evolution [12,14,54].

One might argue that acceptor stems are not based on a GCG (5′-acceptor stem) and complementary CGC (3′-acceptor stem) repeat, although inspection and statistical analyses support the sequence repeats [24]. The typical *P. furiosis* tRNA sequence gives the 5′-acceptor stem sequence as GCGGCGG (a perfect GCG repeat) and the 3′-acceptor stem sequence as CCGC??C (G pairs with both C and U; ? indicates that the typical base was not scored) (Figure 7A). For a larger collection of Archaea in the tRNA database, the typical 5′-acceptor stem sequence is GC?GCGG and the typical 3′-acceptor stem sequence is CCG??GC (Figure 7B). Acceptor stems diverge from a perfect GCG repeat for distinct recognition by cognate AARS enzymes. In *P. furiosis*, the tRNA^{Ser} D loop core begins with two perfect UAGCC repeats UAGCCUAGCC. A tRNA^{Gly} 17 nt D loop core sequence is UAGUCUAGCCUGGUCUA, which is a very close match to a UAGCC repeat [5]. The *P. furiosis* tRNA^{Gly} sequence is the closest to tRNA^{Pri} (the pre-life, primordial tRNA sequence; Figure 5). The closest homology of tRNA^{Pri} and tRNA^{Gly} is consistent with life

evolving from a tRNA–polyglycine world [23]. Because acceptor stems and their fragments in tRNA evolved from GCG and CGC repeats and the 17 nt D loop core evolved from a UAGCC repeat, the pre-life world generated RNA repeats, and, at least in some cases, their complement (GCG and CGC repeats are complementary). Anticodon and T stem-loop-stems are snap-back primers and self-complementary at the stems. One might try to argue that the RNA repeats and homologous inverted repeats conserved in tRNAs for ~4 billion years were all caused by convergent evolution, but such an argument would be untenable. For the highly skeptical, at the very least, the 3 31 nt minihelix theorem is a remarkably good model for tRNA evolution from ~4 billion years ago. Our position is that tRNA sequences embed an unambiguous history of the most central process in the pre-life-to-life transition on Earth.
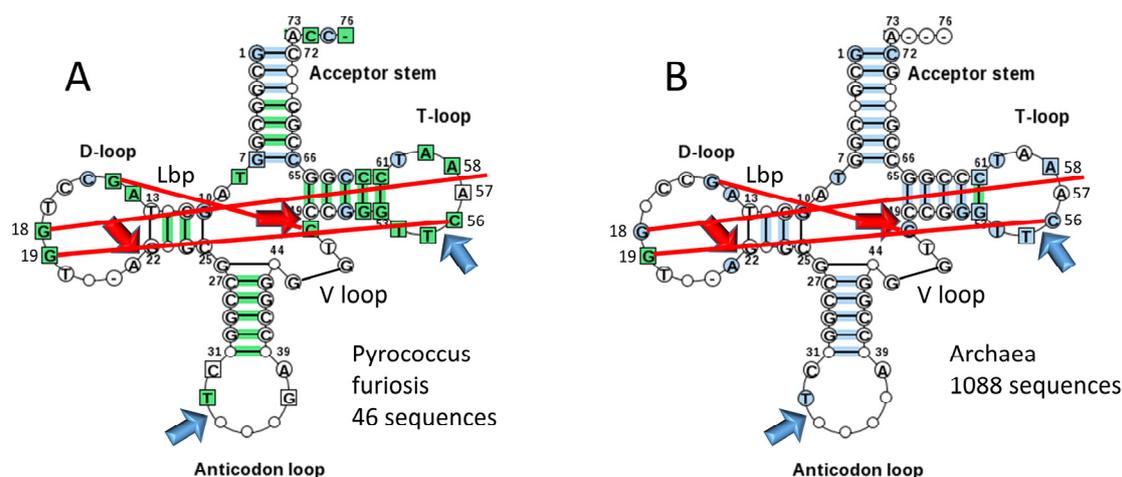
Figure 7. Typical tDNA diagrams for Pyrococcus furiosis (an ancient Archaeon) (**A**) and Archaea (**B**) [5]. Arrow colors: red) processing positions for evolution of type I tRNAs; and blue) U-turns. Red lines indicate: (1) the Levitt reverse Watson–Crick base pair (G15-CV$_5$) (Lbp for Levitt base pair); (2) intercalation of G18 between A57 and A58 ("elbow" contact); and (3) Watson–Crick interaction of G19 and C56 ("elbow" contact) [4].

6. Evolution of the Genetic Code

An appreciation of tRNA evolution appears to demand an anticodon-centric view of genetic code evolution [1,2,21]. We posit that the genetic code evolved around the tRNA anticodon according to how the anticodon was read on the coevolving ribosome. Wobbling at tRNA-34 evolved as the ribosome "learned" to read the anticodon. We posit that tRNA-34 and tRNA-36 were originally both wobble positions but that, as the code and ribosome coevolved, wobbling was suppressed at tRNA-36 by chemical modifications at tRNA-37 and by the closing of the ribosome 16S rRNA in order to tighten recognition of the tRNA–mRNA anticodon–codon interaction [55–57]. A comprehensive model for genetic code evolution has been published in which the placements and distributions of all amino acids in the code are rationalized according to principles of genetics and molecular biology [1,2,21]. Consistent with a polyglycine–tRNA world [37,48], glycine occupies the most favored anticodon (BCC). Consistent with a GADV–tRNA world [36,49–52,58], GADV occupies the most favored row of the code, row 4 (BNC). The model explains the evolution of related amino acids within code columns (i.e., column 1 (BAN) encodes closely related amino acids Val, Ile, Met and Leu; ValRS-IA, IleRS-IA, MetRS-IA and LeuRS-IA are closely related AARS enzymes). Features of the genetic code model include explanations for the coevolution of amino acid metabolism, amino acid chemistry, homologous AARS enzymes, tRNAomes and stop codons.

During code evolution, previously encoded amino acids initially occupied larger segments of the code. For instance, we posit that glycine initially filled the genetic code using all available anticodons to encode glycine. To accommodate the entry of newly

encoded amino acids, previously encoded amino acids retreated, retaining the most favored available anticodon positions. Favored positions in the code relate to favored anticodons according to clear preference rules. For tRNA-34 wobble, A is not utilized in Archaea [5]. Wobble tRNA-34G is favored. Also, at the base of code evolution, wobble tRNA-34C and tRNA-34U are approximately equivalent, limiting the creation of 1 codon sectors (i.e., for Met and Trp; see below). For tRNA-35 and tRNA-36, the anticodon preference rules are C > G > U >> A. So, glycine occupies the most favored anticodon (BCC). GADV occupies the most favored row 4 (BNC). Phenylalanine, tyrosine and tryptophan occupy disfavored row 1 (BNA) and are late additions to the code. Stop codons occupy disfavored row 1. The model explains why unmodified A is not utilized at a wobble position (i.e., unmodified tRNA-34A is not utilized). A rational model is proposed for serine jumping from column 2 to column 4 of the genetic code (see also [59]). Essentially, all aspects of genetic code evolution are potentially explained.

Moreover, the 1 codon sector encoding Met and Trp is explained as a somewhat special case. In Archaea, tRNA^{Met} (CAU) is utilized with minor or no wobble tRNA-34C modification to recognize Met mRNA codon AUG but not Ile codon AUA. Interestingly, tRNA^{Ile} (CAU) is also utilized but with tRNA-34C modified to agmatidine to recognize Ile codon AUA but not Met codon AUG. In the case of the tryptophan 1 codon sector, tRNA^{Trp} (CCA) shares a 2 codon box with a stop codon UGA. Stop codons are recognized by binding of protein release factors to mRNA codons [60], so no tRNA (UCA) ambiguity conflicts with the tRNA^{Trp} (CCA) reading of codon UGG.

The 6 codon sectors encoding leucine, serine and arginine have been explained. We posit that leucine, serine and arginine probably entered the genetic code at about the same time (GADVLSER world). These amino acids occupied larger segments of the code and then retreated to their current positions as new amino acids were added. Serine jumped from column 2 to column 4 of the code, as we have described. Arginine may have entered the code initially as ornithine, which was then converted to arginine through the evolution of tRNA-linked reactions. Because 6 codon sectors have previously been addressed in detail [1,2,21], the entire discussion is not reproduced here.

Other views of genetic code evolution have been reported [61–74]. We object to codon-centric models with a complexity of 64 codons to describe the initial establishment of the code. The genetic code evolved around the tRNA anticodon and its reading on the coevolving ribosome. Wobbling at tRNA-34 limits the wobble position to purine–pyrimidine discrimination, limiting the complexity of the genetic code to 2 × 4 × 4 = 32 assignments (not 4 × 4 × 4 = 64 assignments, as in mRNA). The standard genetic code gained closure at 20 amino acids plus stops (21 assignments). Late in the process of code evolution, translational fidelity limited further expansions of the code.

7. Recorded History of the Pre-Life-to-Life Transition

Sequences of tRNAs relate a history of the pre-life-to-life transition (Figures 1–7). To replicate RNAs and minihelices in the pre-life world required the ligation of RNAs. The emergence of tRNAs shows that minihelices were ligated together (Figures 1 and 3). The importance of RNA ligation in establishing pre-life genetic innovation and complexity cannot be overstated. Ligation of related and unrelated RNAs led to pre-life chemical and combinatorial complexity, which led to life. Ligation explains how tRNA evolved from 3 31 nt minihelices (Figures 1 and 3). In order to replicate 31 nt minihelices, RNAs were ligated together, replicated and then processed to generate new minihelices. Because of side reactions of ribozymes/primitive catalysts, in the process of replicating existing RNAs, novel and more complex RNAs, such as tRNAs, were generated. The ribozymes/primitive catalysts required for these processes include the following: RNA ligases [75–79], RNA helicases/chaperones [80], complementary RNA template-dependent RNA replicase [81–87], and RNA endonucleases [88,89]. Very clearly, to support the polymer world required a ribozyme/primitive catalyst to generate and replicate RNA repeats.

Scientists have encountered some difficulties generating a ribozyme RNA template-dependent RNA replicase [87,90]. Also, to our knowledge, no one has generated a ribozyme/primitive catalyst to generate RNA repeats. These issues may be related. We suggest that the search for ribozymes/primitive catalysts to generate RNA repeats and complementary RNA covalent assemblies be broadened to determine alternate routes to support these activities. Experiments must now be performed to reproduce the polymer and minihelix world. The RNA reactants and products are known with reasonable certainty (Figure 5). Only specific ribozymes/primitive catalysts to support the transitions are lacking. Most of these ribozymes have been generated or approximated in vitro.
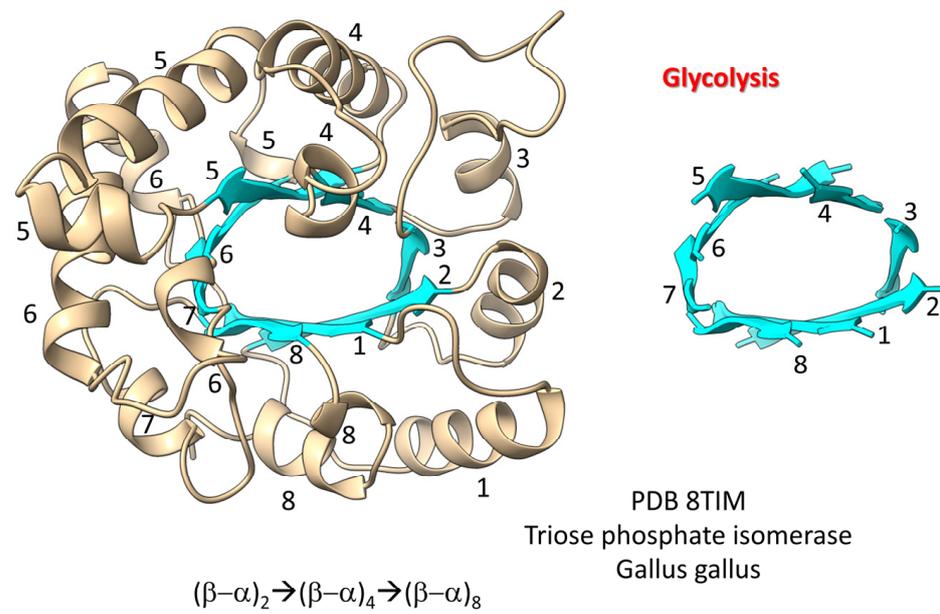
We posit that complementary stems of 5 nt and 7 nt, which are found in minihelices and tRNAs, were selected because shorter stems were unstable and longer RNA duplexes, generally, were more difficult and slower to unwind using ribozyme/primitive catalyst helicases/chaperones. We posit that the tRNA U-turn loop (Figures 1–6) was selected because it is a tight loop that resists attack by ribozyme nucleases. If a 7 nt anticodon U-turn loop has loop positions 1C and 7A (Figures 6B and 7), a Hoogsteen H-bond forms that supports U-turn geometry and loop stability [46]. Once tRNA evolved, a genetic code became inevitable. Given the pre-life chemical milieu, tRNA was a molecule that could "teach" itself ("learn") to code [1,2,21].

Carell and colleagues have indicated that modification of the 2′-O of RNA (i.e., 2′-O-methyl) may have stabilized RNA to OH^- hydrolysis and also may have modified ribozyme activities [38]. RNA modified at the 2′ position is expected to be a precursor to DNA. We posit that many RNA modifications existed in the pre-life world, and RNA modification enzymes coevolved with the code [91]. For instance, the evolution of the genetic code probably required modifications of tRNA-34U, tRNA-37A and tRNA-37G, at a minimum. Probably tRNA-34U methylation-based modifications were necessary to suppress "superwobbling" or 4-way wobbling in which tRNA wobble U reads mRNA wobble A, G, C and U [91–93]. Therefore, tRNA-34U methylation-based modifications were necessary to generate 2 codon sectors in the genetic code (i.e., column 3 of the genetic code). It appears that tRNA-37A and tRNA-37G modifications were necessary to read tRNA-36U and tRNA-36A anticodons, respectively.

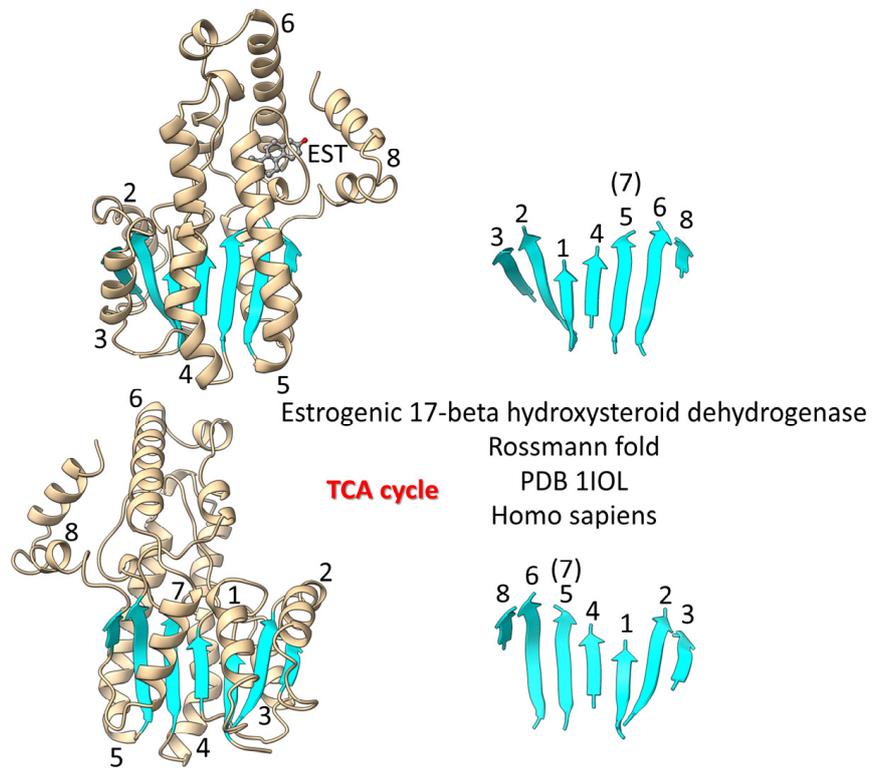## 8. RNA Ligation, Protein Folding and Protein Pseudosymmetry

The process by which tRNA was generated in pre-life (described above) was critically dependent on RNA ligation. To replicate minihelices, RNA ligation and endonuclease processing was necessary. To generate tRNAs, RNA ligation and then processing to novel products occurred. We posit that RNA ligation was a primary mechanism for generating chemical complexity and genetic innovation in the pre-life world. Type II and type I tRNAs were clearly generated by a process in which RNA ligation played a critical role.

The first proteins (i.e., ribosomal proteins, AARS enzymes, $(\beta-\alpha)_8$ barrels (i.e., TIM barrels), $(\beta-\alpha)_8$ sheets (Rossmann folds), RNA modification enzymes, RNA and DNA polymerases) coevolved with the genetic code [94]. We posit that from about the 8 amino acid stage of genetic code evolution, sequence-dependent proteins began to coevolve with the emerging code. We posit that RNA ligation was a central feature in the evolution of the first proteins. As examples, we show Figures 8–11. Figure 8 shows a $(\beta-\alpha)_8$ barrel protein (i.e., a TIM barrel protein; TIM for triose phosphate isomerase). Because β-sheets require a partner β-sheet, we posit that $(\beta-\alpha)_8$ barrels were generated by ligation of two $(\beta-\alpha)_2$ RNAs encoding parallel β-sheets to form a $(\beta-\alpha)_4$ RNA encoding all parallel β-sheets. Ligation of two $(\beta-\alpha)_4$ RNAs created a $(\beta-\alpha)_8$ RNA encoding all parallel β-sheets that was translated and folded into a $(\beta-\alpha)_8$ barrel [95]. In pre-life, folding proteins into pseudosymmetrical barrels may have resulted from RNA ligations combining multiple identical RNAs. Translation of RNA repeats generated protein repeats that could fold into barrels. Glycolytic enzymes include $(\beta-\alpha)_8$ barrels, so we have described the evolution of glycolysis in pre-life.

Figure 8. Evolution of $(\beta-\alpha)_8$ barrels by RNA ligation, translation and pseudosymmetrical folding. β-sheets and α-helices are numbered in the figure.

Figure 9. Refolding of a $(\beta-\alpha)_8$ barrel generated a $(\beta-\alpha)_8$ sheet. β7 lost its β-sheet partners in the refolding. Two views are shown. EST is estradiol. Helices are numbered in the left images. β-sheets are numbered in the right images.
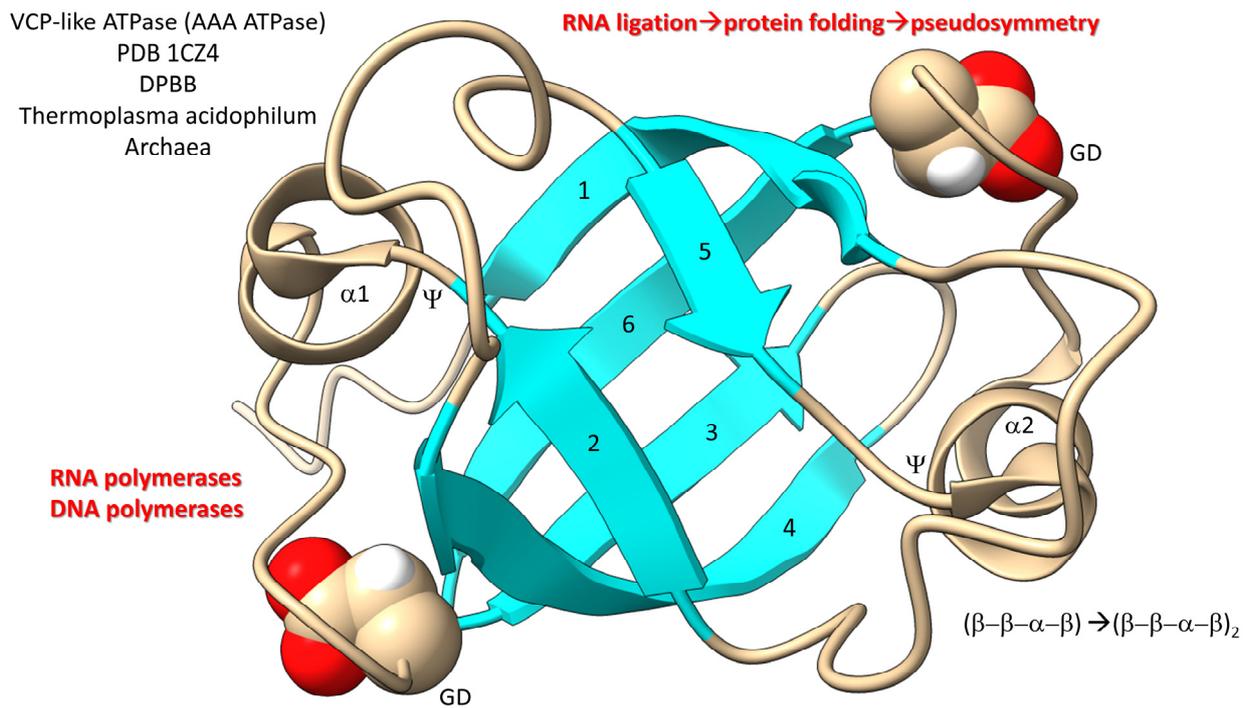
Figure 10. Generation of double-$\Psi-\beta$-barrels in pre-life. Numbers indicate β-sheets.

Expect = 3 x 10^{-13}

```
Query  357  KVVAKPDMKKFGPLFKGDSPKIKAVLDETDATIIKNAFEADGTFKVEVEGKEYELTEDLV  416
            K++AKP++K  GP  +GD+PK+    L E D + IK+  +A+G++ VEV+G+  EL  D +
Sbjct  872  KIIAKPNLKTLGPRLRGDAPKVMKHLTEADGSEIKSILDAEGSYSVEVDGRSIELGVDDI  931

Query  417  SFQ  419 WP_071907205.1 glycine--tRNA ligase [Methanobacterium congolense]
            F+
Sbjct  932  LFE  934 WP_095652122.1 isoleucine--tRNA ligase [Methanobacterium bryantii]
```

Expect = 6 x 10^{-11}

```
Query  359  TVKPNMGILGPRFKGKAAKIANALKALKPEELSGDV----IELTIDGEKITIEKDAVAFE  414
            VKPNM I+GP+F+ +A  I    L ++ P E++  +     I + IDGE I +E ++VA E
Sbjct  782  NVKPNMAIIGPKFRKQAGAIIKTLTSMDPVEVANIISKGNININIDGEDIELEPESVAIE  841

Query  415  K  415 WP_096204125.1 glycine--tRNA ligase [Candidatus Methanoperedens nitroreducens]
            K
Sbjct  842  K  842 WP_011500185.1 valine--tRNA ligase [Methanococcoides burtonii]
```
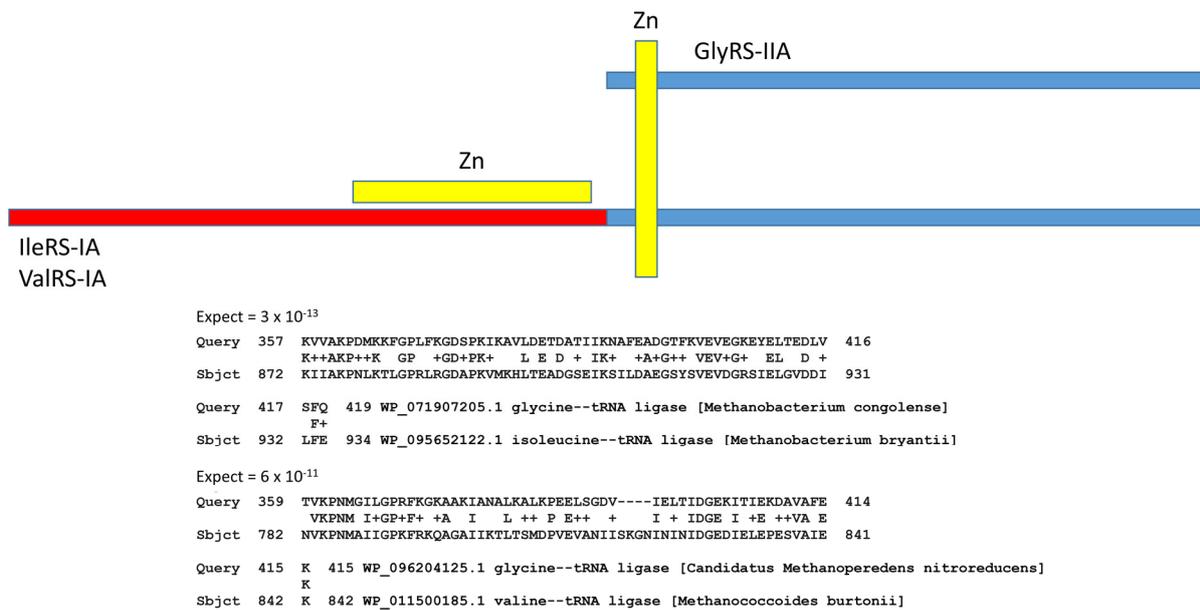
Figure 11. Evolution of AARS enzymes. Class II AARS are simple homologs of class I AARS. The blue segments include homologous sequences, including a Zn finger. The red segment is unique to class I AARS and directs the distinct class I AARS fold. At the bottom of the figure are two alignments demonstrating homology of GlyRS-IIA (class II), IleRS-IA (class I) and ValRS-IA (class I). + indicates amino acid similarity. Expect values indicate homology of sequences.

We posit that during pre-life $(\beta-\alpha)_8$ barrels were rearranged into $(\beta-\alpha)_8$ sheets (Rossmann folds) by protein refolding (Figure 9) [96]. TCA cycle enzymes include Rossmann folds. Thus, significant metabolic capacity coevolved with the genetic code in pre-life via RNA ligations, much as described for the evolution of type II and type I tRNAs.

In Figure 10, a domain of a AAA-ATPase is shown [97–99]. This pseudosymmetrical barrel is a double-$\Psi-\beta$-barrel $(\beta-\beta-\alpha-\beta)_2$. We posit that the double-$\Psi-\beta$-barrel was formed from ligation of two $(\beta-\beta-\alpha-\beta)$ RNAs followed by translation and pseudosymmetrical folding. PolD (an archaeal DNA polymerase) and multi-subunit RNA polymerases are 2 double-$\Psi-\beta$-barrel polymerases. The GD motif in the loop separating $\beta5$ and $\beta6$ was the precursor to the NADFDGD motif that is highly conserved in multi-subunit RNA polymerases [100,101]. We posit that these core life functions were generated in pre-life via RNA ligation, translation and pseudosymmetrical folding. Thus, based on the apparent RNA ligation mechanism for tRNA evolution, glycolysis, the TCA cycle, AAA-ATPases, DNA polymerases and RNA polymerases may have been generated during pre-life [102].

AARS enzymes coevolved with tRNAomes and the genetic code. The evolution of AARS enzymes, however, has been improperly understood [1,2,22]. We have attempted to clarify how AARS enzymes evolved. Despite their different folds, class II and class I AARS enzymes are simple homologs (Figure 11). Notably, GlyRS-IIA is a homolog of IleRS-IA and ValRS-IA. In ancient Archaea, these enzymes often share a Zn finger and significant homology. At the bottom of the figure, we show local alignments of the same segment of GlyRS-IIA with homologous segments of IleRS-IA and ValRS-IA. Initially, class II and class I AARS were not thought to be homologous because these enzymes have distinct folds. We posit that, in pre-life, a primitive ValRS-IA was derived from a primitive GlyRS-IIA by ligation of a distinct RNA encoding the ValRS-IA N-terminal segment, followed by translation and folding. The addition of the N-terminal ValRS-IA extension determined the distinct class I AARS fold, as we have described.

9. Proof of the 3 31 nt Minihelix tRNA Evolution Theorem

The axioms (i.e., assumptions; obvious statements of proposition) on which the proof is based are as follows: (1) an obvious GCG repeat (5′ acceptor stem) is a GCG repeat (Figure 7); (2) a GCG repeat complement, a CGC repeat (3′ acceptor stem) is a CGC repeat; (3) a UAGCC repeat is a UAGCC repeat; and (4) a homologous anticodon (CCGGG_CU/???AA_CCCGG) 17 nt stem-loop-stem and a T (CCGGG_UU/CAAAU_CCCGG) 17 nt stem-loop-stem are homologous stem-loop-stems [1–3,21]. The theorem has made the following predictions: (1) the statistics reported in Table 1 were predicted [3,24,42] and (2) the type II V loop was initially based on the ligation of a 7 nt 3′-acceptor stem and a 7 nt 5′-acceptor stem (Figure 1 and Table 1) [42]. The very few sequence deviations from the theorem are readily justified by principles of tRNA folding. For instance, G19 was selected to be G instead of A (in the third D loop UAGCC repeat) to support D loop–T loop interactions at the elbow of tRNA (Figure 7) [3]. Minor sequence changes in the D stem and V loop occurred to support the tRNA fold versus a minihelix fold (i.e., the Levitt base pair in type II tRNAs). The theorem predicts and justifies all sequences found in tRNAs of Archaea and Bacteria. This is a paper in which definite and publically available conserved tRNA sequence data were used to reach definite conclusions about the origin of life on Earth.

10. Discussion and Conclusions

In the context of pre-life, tRNA was a molecule that could teach itself to code. Because tRNA^{Pri} sequences were highly patterned and their order conserved, the evolution of type II and type I tRNAs is a solved problem. The tRNAomes of LUCA and organisms diverged from type II and type I tRNAs by duplications, mutations and repurposing. For obvious reasons (see above), no chaotic convergent or accretion model can be reasonable to describe the earliest tRNA evolution. To build the genetic code, the evolution of mRNA codons conformed to the evolution of tRNAome anticodons.

The genetic code coevolved with amino acid metabolism, polypeptides, tRNAomes, mRNA, AARS enzymes, first proteins and enzymes, tRNA modification enzymes (some of which depend on relics of $(\beta-\alpha)_8$ barrels) [103] and protocells. The patterns of relatedness within tRNAomes and AARS in ancient Archaea relate to a history of the evolution of the genetic code. We posit that RNA ligation was a major driving force for the evolution of

the complexity of the first RNAs and the folding complexity and pseudosymmetry of the first proteins.

Sequences of tRNAs are among the most highly conserved from pre-life. Sequences of rRNA have been reported to have been derived, in part, from tRNA sequences [29–32]. The first proteins appear to include sequences and structures that partly reflect the pre-life-to-life transition (Figures 8–11). So far as we are aware, tRNA sequences represent the clearest insights into the pre-life-to-life transition on Earth.

To teach the biological sciences, integrate and emphasize genetic coding, evolution, sequence analysis, translation, structure and the central functions of RNA in biology. The history of tRNA chemical evolution, as written into and conserved in living genetic code, relates these fundamental lessons and describes the pre-life-to-life transition from about 4 billion years ago. Note that we are advocating for a tRNA-centered evolution of life on Earth based on analysis of tRNA sequences conserved over about 4 billion years since life began. The RNA world was initially hypothesized for similar reasons because tRNA, mRNA and rRNA form the functional core of molecular biology. The concept was to identify core conserved functions and to evolve around them toward more complex systems. We advocate for centering attention on tRNA. It appears to us that mRNA and rRNA coevolved to accommodate the central functions of tRNAs in the evolution of coding. The evolution of tRNA combines aspects of evolution, the birth of biology on Earth, the origin of life, chemical evolution, sequence pattern recognition, coding and decoding, tRNA modifications, learning and genetic memory and the evolution of the patterning of the genetic code. A tRNA-first approach to the evolution of the genetic code makes sense. Ribosome-first, mRNA-first and genetic code-first approaches are illogical.

Because tRNA evolution was first solved by inspection and as a puzzle, tRNA evolution should be of interest to gamers and puzzlers. Because tRNA evolution is a problem in biological coding, learning and problem-solving, tRNA evolution is a subject for programmers and psychologists. Linguists and mathematicians have shown an interest in the evolution of the genetic code [68–74]. Sentient humans on Earth must know this content.

TRNA forms the functional core of living systems on Earth. The ribosome apparently evolved largely after tRNA. For instance, the ribosome includes a tRNA-shaped channel through the 23S rRNA within the large ribosomal subunit [104]. Also, rRNA is hypothesized to have evolved partly from tRNA fragments [29–32]. The anticodon 7 nt U-turn loop was selected in pre-life to support a 3 nt genetic code [1–3,21,24]. We posit that once tRNA chemically evolved, the evolution of the genetic code became inevitable. Acceptance of these ideas allows pre-life on Earth to be described. The rejection of these ideas is not useful in describing the chemical evolution of life on Earth.

On Earth, tRNA is the genetic adapter. Without a genetic adapter, there can be no genetic code and no life. To re-engineer tRNA to form a different genetic adapter would be problematic. To design a genetic adapter using another chemistry than RNA seems unlikely. To redesign the anticodon loop of tRNA would present a significant challenge (Figure 6B). tRNA evolution presents a remarkable history of the pre-life-to-life transition on Earth. If life is discovered on Enceladus (or elsewhere in the cosmos), we posit it will utilize a similar adapter molecule to tRNA. tRNA and the tRNA anticodon loop are core intellectual properties in the evolution of life.

References

1. Lei, L.; Burton, Z.F. Evolution of the genetic code. *Transcription* **2021**, *12*, 28–53. [CrossRef]
2. Lei, L.; Burton, Z.F. Evolution of Life on Earth: tRNA, Aminoacyl-tRNA Synthetases and the Genetic Code. *Life* **2020**, *10*, 21. [CrossRef]
3. Burton, Z.F. The 3-Minihelix tRNA Evolution Theorem. *J. Mol. Evol.* **2020**, *88*, 234–242. [CrossRef] [PubMed]
4. Zhang, J.; Ferre-D'Amare, A.R. The tRNA Elbow in Structure, Recognition and Evolution. *Life* **2016**, *6*, 3. [CrossRef] [PubMed]
5. Juhling, F.; Morl, M.; Hartmann, R.K.; Sprinzl, M.; Stadler, P.F.; Putz, J. tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res.* **2009**, *37*, D159–D162. [CrossRef] [PubMed]
6. Nagaswamy, U.; Fox, G.E. RNA ligation and the origin of tRNA. *Orig. Life Evol. Biosph.* **2003**, *33*, 199–209. [CrossRef] [PubMed]
7. Demongeot, J.; Seligmann, H. RNA Rings Strengthen Hairpin Accretion Hypotheses for tRNA Evolution: A Reply to Commentaries by Z.F. Burton and M. Di Giulio. *J. Mol. Evol.* **2020**, *88*, 243–252. [CrossRef]
8. Demongeot, J.; Seligmann, H. The primordial tRNA acceptor stem code from theoretical minimal RNA ring clusters. *BMC Genet.* **2020**, *21*, 7. [CrossRef]
9. Di Giulio, M. An RNA Ring was Not the Progenitor of the tRNA Molecule. *J. Mol. Evol.* **2020**, *88*, 228–233. [CrossRef]
10. Demongeot, J.; Seligmann, H. The Uroboros Theory of Life's Origin: 22-Nucleotide Theoretical Minimal RNA Rings Reflect Evolution of Genetic Code and tRNA-rRNA Translation Machineries. *Acta Biotheor.* **2019**, *67*, 273–297. [CrossRef]
11. Di Giulio, M. A comparison between two models for understanding the origin of the tRNA molecule. *J. Theor. Biol.* **2019**, *480*, 99–103. [CrossRef] [PubMed]
12. Di Giulio, M. The origin of the tRNA molecule: Independent data favor a specific model of its evolution. *Biochimie* **2012**, *94*, 1464–1466. [CrossRef]
13. Di Giulio, M. Formal proof that the split genes of tRNAs of Nanoarchaeum equitans are an ancestral character. *J. Mol. Evol.* **2009**, *69*, 505–511. [CrossRef]
14. 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] [PubMed]
15. Di Giulio, M. Permuted tRNA genes of Cyanidioschyzon merolae, the origin of the tRNA molecule and the root of the Eukarya domain. *J. Theor. Biol.* **2008**, *253*, 587–592. [CrossRef]
16. Di Giulio, M. The non-monophyletic origin of the tRNA molecule and the origin of genes only after the evolutionary stage of the last universal common ancestor (LUCA). *J. Theor. Biol.* **2006**, *240*, 343–352. [CrossRef]
17. Di Giulio, M. The origin of the tRNA molecule: Implications for the origin of protein synthesis. *J. Theor. Biol.* **2004**, *226*, 89–93. [CrossRef]
18. Sun, F.J.; Caetano-Anolles, G. The origin and evolution of tRNA inferred from phylogenetic analysis of structure. *J. Mol. Evol.* **2008**, *66*, 21–35. [CrossRef]
19. Demongeot, J.; Seligmann, H. Spontaneous evolution of circular codes in theoretical minimal RNA rings. *Gene* **2019**, *705*, 95–102. [CrossRef] [PubMed]
20. Demongeot, J.; Glade, N.; Moreira, A.; Vial, L. RNA relics and origin of life. *Int. J. Mol. Sci.* **2009**, *10*, 3420–3441. [CrossRef]
21. 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]
22. Pak, D.; Kim, Y.; Burton, Z.F. Aminoacyl-tRNA synthetase evolution and sectoring of the genetic code. *Transcription* **2018**, *9*, 205–224. [CrossRef] [PubMed]
23. Pak, D.; Du, N.; Kim, Y.; Sun, Y.; Burton, Z.F. Rooted tRNAomes and evolution of the genetic code. *Transcription* **2018**, *9*, 137–151. [CrossRef]
24. Pak, D.; Root-Bernstein, R.; Burton, Z.F. tRNA structure and evolution and standardization to the three nucleotide genetic code. *Transcription* **2017**, *8*, 205–219. [CrossRef] [PubMed]
25. Root-Bernstein, R.; Kim, Y.; Sanjay, A.; Burton, Z.F. tRNA evolution from the proto-tRNA minihelix world. *Transcription* **2016**, *7*, 153–163. [CrossRef]
26. Petrov, A.S.; Gulen, B.; Norris, A.M.; Kovacs, N.A.; Bernier, C.R.; Lanier, K.A.; Fox, G.E.; Harvey, S.C.; Wartell, R.M.; Hud, N.V.; et al. History of the ribosome and the origin of translation. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15396–15401. [CrossRef]
27. Petrov, A.S.; Bernier, C.R.; Hsiao, C.; Norris, A.M.; Kovacs, N.A.; Waterbury, C.C.; Stepanov, V.G.; Harvey, S.C.; Fox, G.E.; Wartell, R.M.; et al. Evolution of the ribosome at atomic resolution. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10251–10256. [CrossRef]
28. Petrov, A.S.; Bernier, C.R.; Gulen, B.; Waterbury, C.C.; Hershkovits, E.; Hsiao, C.; Harvey, S.C.; Hud, N.V.; Fox, G.E.; Wartell, R.M.; et al. Secondary structures of rRNAs from all three domains of life. *PLoS ONE* **2014**, *9*, e88222. [CrossRef]
29. de Farias, S.T.; Rego, T.G.; Jose, M.V. Origin of the 16S Ribosomal Molecule from Ancestor tRNAs. *J. Mol. Evol.* **2021**, *89*, 249–256. [CrossRef]
30. de Farias, S.T.; Jose, M.V. Transfer RNA: The molecular demiurge in the origin of biological systems. *Prog. Biophys. Mol. Biol.* **2020**, *153*, 28–34. [CrossRef]
31. de Farias, S.T.; Rego, T.G.; Jose, M.V. tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World. *Life* **2016**, *6*, 15. [CrossRef]
32. Farias, S.T.; Rego, T.G.; Jose, M.V. Origin and evolution of the Peptidyl Transferase Center from proto-tRNAs. *FEBS Open Bio* **2014**, *4*, 175–178. [CrossRef]

33. de Farias, S.T.; do Rego, T.G.; Jose, M.V. Evolution of transfer RNA and the origin of the translation system. *Front. Genet.* **2014**, *5*, 303. [CrossRef]

34. de Farias, S.T. Suggested phylogeny of tRNAs based on the construction of ancestral sequences. *J. Theor. Biol.* **2013**, *335*, 245–248. [CrossRef] [PubMed]

35. Greenwald, J.; Kwiatkowski, W.; Riek, R. Peptide Amyloids in the Origin of Life. *J. Mol. Biol.* **2018**, *430*, 3735–3750. [CrossRef]

36. Ikehara, K. Evolutionary Steps in the Emergence of Life Deduced from the Bottom-Up Approach and GADV Hypothesis (Top-Down Approach). *Life* **2016**, *6*, 6. [CrossRef] [PubMed]

37. Bernhardt, H.S.; Patrick, W.M. Genetic code evolution started with the incorporation of glycine, followed by other small hydrophilic amino acids. *J. Mol. Evol.* **2014**, *78*, 307–309. [CrossRef]

38. Muller, F.; Escobar, L.; Xu, F.; Wegrzyn, E.; Nainyte, 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] [PubMed]

39. Mayer, C.; Schreiber, U.; Davila, M.J.; Schmitz, O.J.; Bronja, A.; Meyer, M.; Klein, J.; Meckelmann, S.W. Molecular Evolution in a Peptide-Vesicle System. *Life* **2018**, *8*, 16. [CrossRef] [PubMed]

40. Mayer, C.; Schreiber, U.; Davila, M.J. Periodic Vesicle Formation in Tectonic Fault Zones--an Ideal Scenario for Molecular Evolution. *Orig. Life Evol. Biosph.* **2015**, *45*, 139–148. [CrossRef] [PubMed]

41. Schreiber, U.; Locker-Grutjen, O.; Mayer, C. Hypothesis: Origin of life in deep-reaching tectonic faults. *Orig. Life Evol. Biosph.* **2012**, *42*, 47–54. [CrossRef] [PubMed]

42. Kim, Y.; Kowiatek, B.; Opron, K.; Burton, Z.F. Type-II tRNAs and Evolution of Translation Systems and the Genetic Code. *Int. J. Mol. Sci.* **2018**, *19*, 3275. [CrossRef] [PubMed]

43. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Meng, E.C.; Couch, G.S.; Croll, T.I.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30*, 70–82. [CrossRef] [PubMed]

44. Fukunaga, R.; Yokoyama, S. Crystal structure of leucyl-tRNA synthetase from the archaeon Pyrococcus horikoshii reveals a novel editing domain orientation. *J. Mol. Biol.* **2005**, *346*, 57–71. [CrossRef]

45. Biou, V.; Yaremchuk, A.; Tukalo, M.; Cusack, S. The 2.9 A crystal structure of T. thermophilus seryl-tRNA synthetase complexed with tRNA(Ser). *Science* **1994**, *263*, 1404–1410. [CrossRef] [PubMed]

46. Shi, H.; Moore, P.B. The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: A classic structure revisited. *RNA* **2000**, *6*, 1091–1105. [CrossRef]

47. Rodriguez-Garcia, M.; Surman, A.J.; Cooper, G.J.T.; Suarez-Marina, I.; Hosni, Z.; Lee, M.P.; Cronin, L. Formation of oligopeptides in high yield under simple programmable conditions. *Nat. Commun.* **2015**, *6*, 8385. [CrossRef]

48. Bernhardt, H.S.; Tate, W.P. Evidence from glycine transfer RNA of a frozen accident at the dawn of the genetic code. *Biol. Direct* **2008**, *3*, 53. [CrossRef]

49. Ikehara, K. [GADV]-protein world hypothesis on the origin of life. *Orig. Life Evol. Biosph.* **2014**, *44*, 299–302. [CrossRef]

50. Ikehara, K. Pseudo-replication of [GADV]-proteins and origin of life. *Int. J. Mol. Sci.* **2009**, *10*, 1525–1537. [CrossRef]

51. Oba, T.; Fukushima, J.; Maruyama, M.; Iwamoto, R.; Ikehara, K. Catalytic activities of [GADV]-peptides. Formation and establishment of [GADV]-protein world for the emergence of life. *Orig. Life Evol. Biosph.* **2005**, *35*, 447–460. [CrossRef]

52. Ikehara, K. Possible steps to the emergence of life: The [GADV]-protein world hypothesis. *Chem. Rec.* **2005**, *5*, 107–118. [CrossRef]

53. Hansma, H.G. Better than Membranes at the Origin of Life? *Life* **2017**, *7*, 28. [CrossRef] [PubMed]

54. Branciamore, S.; Di Giulio, M. The presence in tRNA molecule sequences of the double hairpin, an evolutionary stage through which the origin of this molecule is thought to have passed. *J. Mol. Evol.* **2011**, *72*, 352–363. [CrossRef]

55. Djumagulov, M.; Demeshkina, N.; Jenner, L.; Rozov, A.; Yusupov, M.; Yusupova, G. Accuracy mechanism of eukaryotic ribosome translocation. *Nature* **2021**, *600*, 543–546. [CrossRef] [PubMed]

56. Rozov, A.; Wolff, P.; Grosjean, H.; Yusupov, M.; Yusupova, G.; Westhof, E. Tautomeric G*U pairs within the molecular ribosomal grip and fidelity of decoding in bacteria. *Nucleic Acids Res.* **2018**, *46*, 7425–7435. [CrossRef]

57. Rozov, A.; Demeshkina, N.; Westhof, E.; Yusupov, M.; Yusupova, G. New Structural Insights into Translational Miscoding. *Trends Biochem. Sci.* **2016**, *41*, 798–814. [CrossRef] [PubMed]

58. Ikehara, K. Why Were [GADV]-amino Acids and GNC Codons Selected and How Was GNC Primeval Genetic Code Established? *Genes* **2023**, *14*, 375. [CrossRef]

59. Inouye, M.; Takino, R.; Ishida, Y.; Inouye, K. Evolution of the genetic code; Evidence from serine codon use disparity in Escherichia coli. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 28572–28575. [CrossRef]

60. Burroughs, A.M.; Aravind, L. The Origin and Evolution of Release Factors: Implications for Translation Termination, Ribosome Rescue, and Quality Control Pathways. *Int. J. Mol. Sci.* **2019**, *20*, 1981. [CrossRef] [PubMed]

61. Koonin, E.V.; Novozhilov, A.S. Origin and Evolution of the Universal Genetic Code. *Annu. Rev. Genet.* **2017**, *51*, 45–62. [CrossRef] [PubMed]

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

63. Yarus, M. Fitting the standard genetic code into its triplet table. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2021103118. [CrossRef] [PubMed]

64. Yarus, M. Evolution of the Standard Genetic Code. *J. Mol. Evol.* **2021**, *89*, 19–44. [CrossRef]

65. Yarus, M. Optimal Evolution of the Standard Genetic Code. *J. Mol. Evol.* **2021**, *89*, 45–49. [CrossRef]

66. Yarus, M. Crick Wobble and Superwobble in Standard Genetic Code Evolution. *J. Mol. Evol.* **2021**, *89*, 50–61. [CrossRef]
67. 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]
68. Negadi, T. Revealing the Genetic Code Symmetries through Computations Involving Fibonacci-like Sequences and Their Properties. *Computation* **2023**, *11*, 154. [CrossRef]
69. Rosandic, M.; Paar, V. The Evolution of Life Is a Road Paved with the DNA Quadruplet Symmetry and the Supersymmetry Genetic Code. *Int. J. Mol. Sci.* **2023**, *24*, 12029. [CrossRef]
70. Rosandic, M.; Paar, V. Standard Genetic Code vs. Supersymmetry Genetic Code—Alphabetical table vs. physicochemical table. *Biosystems* **2022**, *218*, 104695. [CrossRef]
71. Rosandic, M.; Paar, V. Codon sextets with leading role of serine create "ideal" symmetry classification scheme of the genetic code. *Gene* **2014**, *543*, 45–52. [CrossRef]
72. Barbieri, M. Evolution of the genetic code: The ambiguity-reduction theory. *Biosystems* **2019**, *185*, 104024. [CrossRef] [PubMed]
73. Zolyan, S. On the minimal elements of the genetic code and their semiotic functions (degeneracy, complementarity, wobbling). *Biosystems* **2023**, *231*, 104962. [CrossRef] [PubMed]
74. Zolyan, S. On the context-sensitive grammar of the genetic code. *Biosystems* **2021**, *208*, 104497. [CrossRef]
75. DasGupta, S.; Zhang, S.; Szostak, J.W. Molecular Crowding Facilitates Ribozyme-Catalyzed RNA Assembly. *ACS Cent. Sci.* **2023**, *9*, 1670–1678. [CrossRef] [PubMed]
76. Kasuga, M.; Mutsuro-Aoki, H.; Ando, T.; Tamura, K. Molecular Anatomy of the Class I Ligase Ribozyme for Elucidation of the Activity-Generating Unit. *Biology* **2023**, *12*, 1012. [CrossRef]
77. Nomura, Y.; Yokobayashi, Y. RNA ligase ribozymes with a small catalytic core. *Sci. Rep.* **2023**, *13*, 8584. [CrossRef]
78. DasGupta, S.; Zhang, S.J.; Smela, M.P.; Szostak, J.W. RNA-Catalyzed RNA Ligation within Prebiotically Plausible Model Protocells. *Chemistry* **2023**, *29*, e202301376. [CrossRef]
79. Walton, T.; DasGupta, S.; Duzdevich, D.; Oh, S.S.; Szostak, J.W. In vitro selection of ribozyme ligases that use prebiotically plausible 2-aminoimidazole-activated substrates. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 5741–5748. [CrossRef]
80. Lei, Z.-C.; Yin, X.; Wang, X.; Ke, G.; Cao, X.; Fan, C.; Yang, C.J.; Liang, H.; Tian, Z.-Q. RNA can function as molecular chaperone for RNA folding. *Giant* **2020**, *1*, 100008. [CrossRef]
81. Tjhung, K.F.; Sczepanski, J.T.; Murtfeldt, E.R.; Joyce, G.F. RNA-Catalyzed Cross-Chiral Polymerization of RNA. *J. Am. Chem. Soc.* **2020**, *142*, 15331–15339. [CrossRef] [PubMed]
82. Tjhung, K.F.; Shokhirev, M.N.; Horning, D.P.; Joyce, G.F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 2906–2913. [CrossRef]
83. Horning, D.P.; Bala, S.; Chaput, J.C.; Joyce, G.F. RNA-Catalyzed Polymerization of Deoxyribose, Threose, and Arabinose Nucleic Acids. *ACS Synth. Biol.* **2019**, *8*, 955–961. [CrossRef]
84. Samanta, B.; Horning, D.P.; Joyce, G.F. 3′-End labeling of nucleic acids by a polymerase ribozyme. *Nucleic Acids Res.* **2018**, *46*, e103. [CrossRef]
85. Samanta, B.; Joyce, G.F. A reverse transcriptase ribozyme. *eLife* **2017**, *6*, 31153. [CrossRef]
86. Horning, D.P.; Joyce, G.F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9786–9791. [CrossRef] [PubMed]
87. McGinness, K.E.; Joyce, G.F. In search of an RNA replicase ribozyme. *Chem. Biol.* **2003**, *10*, 5–14. [CrossRef]
88. Zhao, N.N.; Liu, Y.Z.; Zhang, L.; Liu, W.; Zou, X.; Xu, Q.; Zhang, C.Y. Construction of Multiple DNAzymes Driven by Single Base Elongation and Ligation for Single-Molecule Monitoring of FTO in Cancer Tissues. *Anal. Chem.* **2023**, *95*, 12974–12981. [CrossRef] [PubMed]
89. Su, J.; Sun, C.; Du, J.; Xing, X.; Wang, F.; Dong, H. RNA-Cleaving DNAzyme-Based Amplification Strategies for Biosensing and Therapy. *Adv. Healthc. Mater.* **2023**, *12*, e2300367. [CrossRef]
90. Pavlinova, P.; Lambert, C.N.; Malaterre, C.; Nghe, P. Abiogenesis through gradual evolution of autocatalysis into template-based replication. *FEBS Lett.* **2023**, *597*, 344–379. [CrossRef]
91. Lei, L.; Burton, Z.F. "Superwobbling" and tRNA-34 Wobble and tRNA-37 Anticodon Loop Modifications in Evolution and Devolution of the Genetic Code. *Life* **2022**, *12*, 252. [CrossRef]
92. Alkatib, S.; Scharff, L.B.; Rogalski, M.; Fleischmann, T.T.; Matthes, A.; Seeger, S.; Schottler, M.A.; Ruf, S.; Bock, R. The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet.* **2012**, *8*, e1003076. [CrossRef]
93. Rogalski, M.; Karcher, D.; Bock, R. Superwobbling facilitates translation with reduced tRNA sets. *Nat. Struct. Mol. Biol.* **2008**, *15*, 192–198. [CrossRef] [PubMed]
94. Weiss, M.C.; Preiner, M.; Xavier, J.C.; Zimorski, V.; Martin, W.F. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genet.* **2018**, *14*, e1007518. [CrossRef]
95. Romero-Romero, S.; Kordes, S.; Michel, F.; Hocker, B. Evolution, folding, and design of TIM barrels and related proteins. *Curr. Opin. Struct. Biol.* **2021**, *68*, 94–104. [CrossRef] [PubMed]
96. Medvedev, K.E.; Kinch, L.N.; Dustin Schaeffer, R.; Pei, J.; Grishin, N.V. A Fifth of the Protein World: Rossmann-like Proteins as an Evolutionarily Successful Structural unit. *J. Mol. Biol.* **2021**, *433*, 166788. [CrossRef] [PubMed]
97. Alva, V.; Dunin-Horkawicz, S.; Habeck, M.; Coles, M.; Lupas, A.N. The GD box: A widespread noncontiguous supersecondary structural element. *Protein Sci.* **2009**, *18*, 1961–1966. [CrossRef] [PubMed]

98. Alva, V.; Koretke, K.K.; Coles, M.; Lupas, A.N. Cradle-loop barrels and the concept of metafolds in protein classification by natural descent. *Curr. Opin. Struct. Biol.* **2008**, *18*, 358–365. [CrossRef]

99. Coles, M.; Hulko, M.; Djuranovic, S.; Truffault, V.; Koretke, K.; Martin, J.; Lupas, A.N. Common evolutionary origin of swapped-hairpin and double-psi beta barrels. *Structure* **2006**, *14*, 1489–1498. [CrossRef]

100. Lei, L.; Burton, Z.F. Early Evolution of Transcription Systems and Divergence of Archaea and Bacteria. *Front. Mol. Biosci.* **2021**, *8*, 651134. [CrossRef]

101. Koonin, E.V.; Krupovic, M.; Ishino, S.; Ishino, Y. The replication machinery of LUCA: Common origin of DNA replication and transcription. *BMC Biol.* **2020**, *18*, 61. [CrossRef] [PubMed]

102. Weiss, M.C.; Sousa, F.L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W.F. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **2016**, *1*, 16116. [CrossRef]

103. Schwalm, E.L.; Grove, T.L.; Booker, S.J.; Boal, A.K. Crystallographic capture of a radical S-adenosylmethionine enzyme in the act of modifying tRNA. *Science* **2016**, *352*, 309–312. [CrossRef] [PubMed]

104. Opron, K.; Burton, Z.F. Ribosome Structure, Function, and Early Evolution. *Int. J. Mol. Sci.* **2018**, *20*, 40. [CrossRef] [PubMed]