


Review

A Recipe to Evolve Complex Life Chemically on Earth

Lei Lei¹ and Zachary Frome Burton^{2,*} ¹ School of Biological Sciences, University of New England, Biddeford, ME 04005, USA; llei@une.edu² Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

* Correspondence: burton@msu.edu

Abstract

Sequences of tRNAs are highly patterned in easily identifiable RNA repeats and RNA inverted repeats (stem–loop–stems). Because of patterning, the multi-step evolution of tRNA can be described in remarkable detail. To evolve life on Earth or another planet or the moon requires the evolution of tRNA or a tRNA-like molecule to act as a genetic adapter. To replace tRNA with an alternate or improved genetic adapter is a remarkably challenging problem, indicating strong chemical selection of tRNA precursors in pre-life. The genetic code, translation systems, and first proteins coevolved with tRNA_{omes} (all of the tRNAs of an organism). Because the tRNA sequence can be separated into component parts, a simple pathway for chemical evolution of life and genetic coding can be described in sufficient detail to allow the assembly of a living entity in laboratories.

Keywords: abiogenesis; astrobiology; tRNA; polyglycine; evolution of the genetic code; origin of life; protocells

1. Introduction

Because features of complex life emerge with tRNA and tRNA_{omes}, the evolution of tRNA is the central and most essential pathway to evolution of a genetic code and complex life supported by coding [1–4]. To generate complex life requires a genetic code, which cannot be evolved except by evolving a genetic adapter. The tRNA molecule arose with specialized features that make tRNA difficult to improve or replace. Life on Earth coevolved with tRNA, tRNA_{omes}, and the genetic code. On another planet or the moon, we suggest that life must evolve by a very similar mechanism, utilizing tRNA or a very similar tRNA-like molecule.

Evolution of tRNA, which occurred about 4.2 billion years ago, has been described in detail [1,3]. The original tRNA molecule was generated from GCG, CGC, and UAGCC repeats and stem–loop–stems (CCGGG_{CU}/NNNAA_{CCCCG} and UAGCCUAGCCUAGCCUA; N indicates sequences scrambled in coding; _ separates sequence features such as stem–loop–stems; / indicates a U-turn) [1,3]. The pattern is obvious from an analysis of typical tRNA diagrams and sequence logos. The original tRNA molecule evolved from ligation of three 31 nt minihelices of mostly known sequences (GCGGCCGG_{UAGCCUAGCCUAGCCUA_CCGCCGC} and GCGGCCGG_{CCGGG_{CU}/NNNAA_CCCGG_CCGCCGC}). In this paper, we substitute the glycine anticodon GCC for NNN. We explain this by simplifying and possibly correcting the assignment below. Ligation of 31 nt minihelices was followed by internal deletions of 9 nt within ligated acceptor stems (CCGCCGC_{GCGGCCGG}). The only sequence ambiguities in the pathway (indicated by N) are the bases that have been altered in coding to form tRNA_{omes}. It is possible that the T loop of tRNA was generated



Academic Editor: Zihua Hu

Received: 11 August 2025

Revised: 14 September 2025

Accepted: 16 September 2025

Published: 25 September 2025

Citation: Lei, L.; Burton, Z.F. A Recipe to Evolve Complex Life Chemically on Earth. *Genes* **2025**, *16*, 1136. <https://doi.org/10.3390/genes16101136>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

from the complement of the anticodon loop minihelix sequence because the complementary sequence is almost identical to the sequence given [1]. ACCA-Gly ligated at the 3'-end allowed the initial tRNA molecule to be utilized to synthesize polyglycine. We suggest that, before the advent of sequence-dependent proteins, polyglycine was a major initial chemical driving force supporting the evolution of living systems. The original print of the tRNA sequence, therefore, was highly patterned and ordered, and shows how the molecule was generated in pre-life. Life evolved around tRNA and the tRNA anticodon loop, explaining why tRNA sequences in living organisms are so highly conserved from pre-life, while interacting systems have become more innovative.

tRNA, tRNA^{Aomes}, aminoacyl-tRNA synthetases, the first proteins (proteins coevolved with the genetic code), the genetic code, ribosomes, and the first cells coevolved. If the tRNA molecule were not generated early in the process, the complex coevolution could not have advanced. There may be very few or no alternative routes than tRNA to the evolution of life on Earth or on other celestial bodies.

To our knowledge, a recent review that integrates the tRNA sequence, structure, and evolution has not been published. Here, we attempt to construct a straightforward description of the tRNA sequence, structure, and evolution that can be used as a guide to recreate many of the core steps in the emergence of complex life on Earth. Life can be defined in various ways. Here, we refer to complex life as manifested in the first microbial cells, supported by a genetic code. The evolution of tRNA and tRNA^{Aomes} was a prerequisite to the origin of microbial life.

2. Materials and Methods

tRNA sequences were obtained from the genomic tRNA database (gtRNAdb) [5] and the tRNA Gene DataBase Curated by Experts (tRNADB-CE) [6–8]. RCSB (Research Collaboratory for Structural Bioinformatics) Protein Data Bank [9–11] files were imaged using ChimeraX (version: 1.10.1 (24 July 2025)) [12–14].

3. Type II tRNA

Type II tRNA has a longer V arm compared to a type I tRNA V loop [1,2]. Otherwise, type II and type I tRNAs are homologous over their entire lengths. The type II V arm was initially 14 nt, generated by ligation of a 3'-acceptor stem (7 nt; initially CCGCCGC) to a 5'-acceptor stem (7 nt; initially GCGGCGG). The type I V loop was processed from an initial type II V arm by a 9 nt deletion (eliminating GCGGCGG). The initial sequence of a type I V loop was CCGCC. The type II V arm evolved to form a stem–loop–stem with a characteristic trajectory from the body of the tRNA. Typically, the cognate aminoacyl-tRNA synthetase (AARS; i.e., LeuRS-IA and SerRS-IIA) recognizes the type II V arm by its sequence and/or its trajectory. In Archaea, type II tRNAs (tRNA^{Leu} (5 anticodons) and tRNA^{Ser} (4 anticodons)) utilize the type II V arm rather than the anticodon loop as a major determinant for accurate amino acid addition [2]. The type I V loop was selected to form contacts to other tRNA residues. Type I tRNAs generally utilize the anticodon loop as a major determinant for cognate tRNA recognition (type I tRNA^{Ala} is an exception) and not the shorter type I V loop.

Figure 1 shows a type II tRNA^{Leu} (CAA) from the ancient Archaeon *Pyrococcus horikoshii*, colored according to the three 31 nt minihelix tRNA evolution theorem [1,3]. Sequences with a common color are homologous to one another. We suggest that the *P. horikoshii* tRNA^{Leu} (CAA) [15,16] is very similar to a tRNA^{Leu} from LUCA (the last universal common (cellular) ancestor) [17,18] because its sequence is very close to the primordial sequence. To the right of the image is a 2-dimensional tRNA schematic diagram with consistent coloring. Historic numbering of tRNAs breaks down in the D loop because of

deletions, and in the type II V arm and type I V loop, because type II V arm and type I V loop sequences were misaligned and because of indels (insertions and deletions). Here, we use D loop numbering D_1 to D_{17} . For the type I V loop, we number V_1 to V_5 . For the type II V arm, we number V_1 to V_n (V arm of n bases). In the absence of indels, V_1 to V_5 align for type I V loops and type II V arms. The type I V loop was processed from an early version of the type II V arm by a 9 nt internal deletion within ligated 3'- and 5'-acceptor stems (CCGCCGC_GCGGCGG was shortened to CCGCC with GC_GCGGCGG deleted).

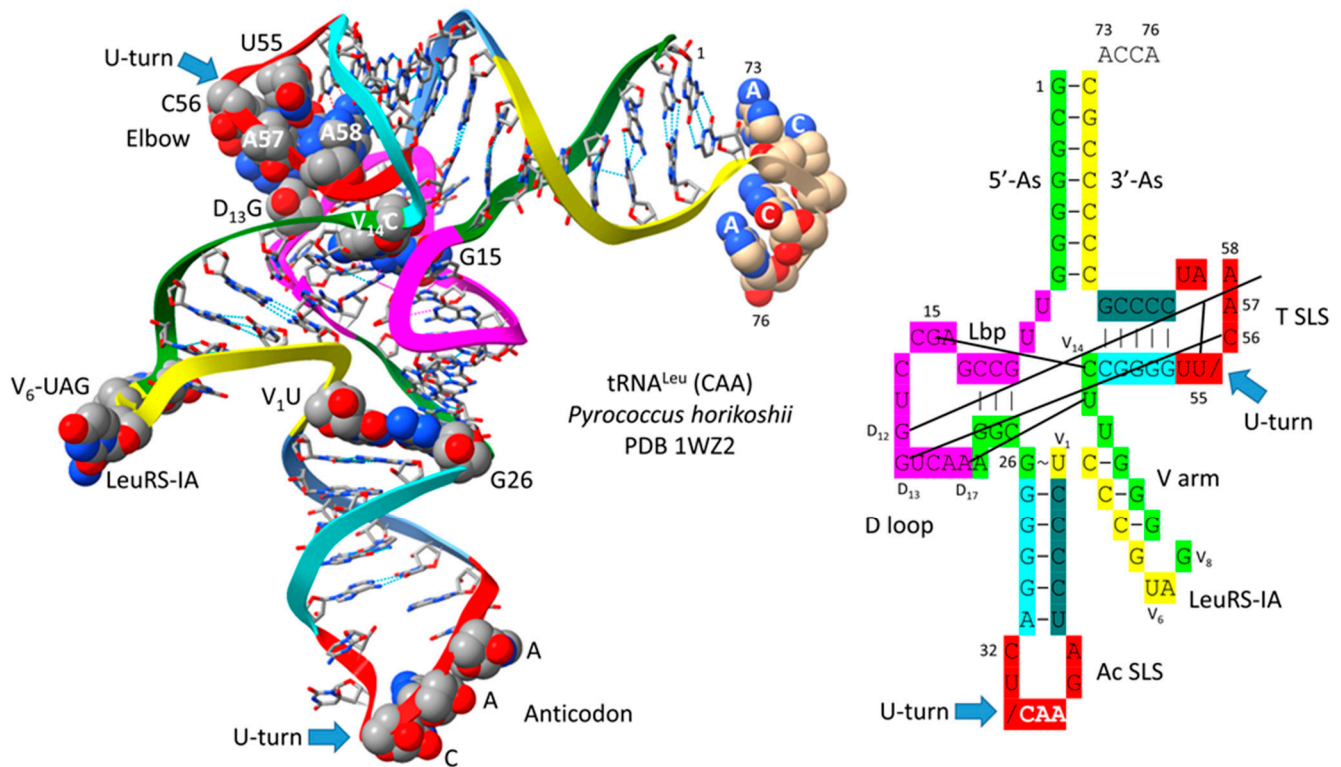


Figure 1. Type II tRNA. A tRNA^{Leu} (CAA) from *P. horikoshii* (an ancient Archaeon) is shown (PDB 1WZ2) [15,16]. The image is colored according to the three 31 nt minihelix theorem for the evolution of tRNA. Consistent colors indicate homologous sequences: green: 5'-acceptor stem (5'-As), 5'-As* (5'-acceptor stem remnant homologous to 5'-As tRNA-3-7), and V_8 - V_{14} of the type II V arm; magenta: the D loop 17 nt minihelix core; yellow: V_1 - V_7 of the V arm and the 3'-acceptor stem (3'-As); cyan: anticodon and T 5'-stem; red: anticodon loop and T loop; cornflower blue: anticodon and T 3'-stem. 3'-ACCA is uncolored. / indicates a U-turn in the RNA backbone. SLS indicates stem-loop-stem. Lbp indicates the Levitt reverse Watson-Crick base pair between tRNA-15G (D_8 G) and V_{14} C. LeuRS-IA binds V_6 -UAG- V_8 as a determinant for cognate charging of tRNA-76A 2'-O of the ribose ring with leucine.

In Figure 1, color coding corresponds to internal tRNA homologies. The 17 nt anticodon and the 17 nt T stem-loop-stems are homologs. Because the complementary sequence of the anticodon stem-loop-stem is almost the same as the forward sequence, the T stem-loop-stem may be derived from the complement of the anticodon stem-loop-stem rather than the direct anticodon stem-loop-stem sequence [1]. The 5'-As* sequence (tRNA-22 to tRNA-26; green) is homologous to tRNA-3 to tRNA-7 of the 5'-As (As for acceptor stem).

The tRNA^{Leu} (CAA) image in Figure 1 is from a co-crystal with the LeuRS-IA AARS charging enzyme [15,16]. The 3'-ACCA is bent down into the LeuRS-IA aminoacylating active site for the addition of leucine (the “hairpin” conformation). LeuRS-IA binding causes some unwinding of the tRNA^{Leu} anticodon loop, although LeuRS-IA does not

bind the anticodon loop directly. Because leucine is in a six-codon box in the standard genetic code and utilizes five tRNA^{Leu}, LeuRS-IA binds the type II V arm instead of the five different anticodon loops as a determinant for cognate tRNA^{Leu} charging.

Figure 2 shows a comparison of type II V arms and type I V loops [1,2]. Type II V arms were derived from ligation of a 7 nt 3'-acceptor stem (CCGCCGC) to a 7 nt 5'-acceptor stem (GCGGCGG) [1,2]. Such a 14 nt sequence could pair along its entire length. The 14 nt V arm sequence evolved to the tRNA^{Leu} and tRNA^{Ser} type II V arms by forming distinct stem-loop-stems with different cognate trajectories from the tRNA body. V₁U interacts with tRNA-26G, forming a single hydrogen bond. The V arm is utilized as a determinant for cognate Leu and Ser charging at the tRNA 3'-CCA end. The trajectory of the V arm from the tRNA body is determined from the number of unpaired bases between the 3'-V arm stem and the Levitt base CV_n (for a V arm of n nucleotides). To specify cognate charging, the trajectory of the V arms is typically distinct for the set of tRNA^{Leu} (five tRNA^{Leu}) and the set of tRNA^{Ser} (four tRNA^{Ser}) and is a determinant for cognate recognition [2].

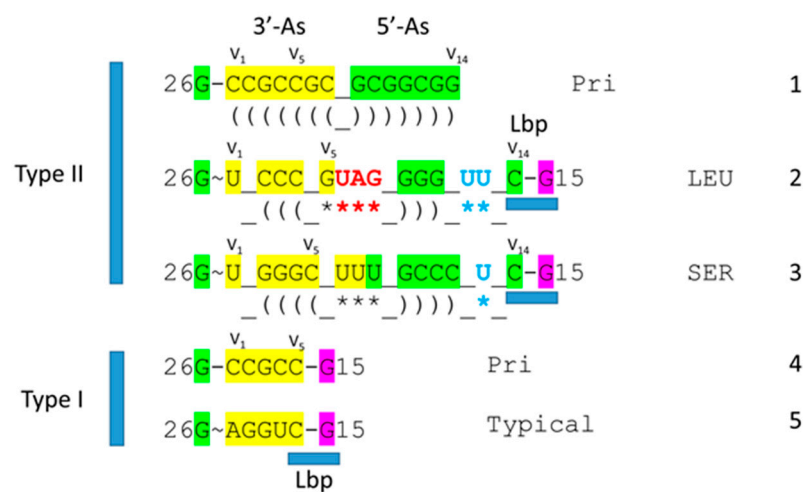


Figure 2. Evolution of type II V arms and type I V loops at the base of the genetic code. Pri indicates the type II (line 1) and type I (line 4) primordial (pre-life) sequences. The type I V loop was processed from the primordial type II V arm sequence. Examples of archaeal tRNA^{Leu} (line 2) and tRNA^{Ser} (line 3) V arm sequences are shown. A typical and common type I V loop sequence from Archaea is shown (line 5). tRNA^{Leu} type II V arm V₆-UAG-V₈ (red) binds LeuRS-IA as a determinant for cognate tRNA^{Leu} charging [15,16]. Cyan bases determine the trajectory of the type II V arm from the tRNA body. _ separates stems and loops. (and) (parentheses) indicate stems. * indicates loops. Colors as in other figures.

Leucine and serine are in six-codon boxes in the standard genetic code. The type II V arm is utilized as a major determinant for cognate leucine and serine charging of tRNAs. In Archaea, the type II V arm is only utilized by tRNA^{Leu} (five tRNA^{Leu}) and tRNA^{Ser} (four tRNA^{Ser}). In Bacteria, tRNA^{Tyr}, tRNA^{Leu}, and tRNA^{Ser} utilize type II V arms [2]. In Archaea, almost all tRNA^{Leu} V arms are 14 nt in length, which is the primordial length. Ligating a 7 nt 3'-acceptor stem to a 7 nt 5'-acceptor stem generates a 14 nt sequence (initially CCGCCGC_GCGGCGG) (Figure 2, line 1). The type II tRNA^{Leu} and tRNA^{Ser} V arms evolved to form stem-loop-stems utilized as determinants by LeuRS-IA and SerRS-IIA for cognate amino acid charging at tRNA-76A. The type I V loop was processed from the primordial type II V arm sequence (initially CCGCCGC_GCGGCGG processed to CCGCC) (line 4).

Type II tRNAs and type I tRNAs are homologous over their entire lengths except for the 9 nt deleted segment in the type I tRNA V loop region. Most contacts in type II tRNAs and type I tRNAs are the same. Some core interactions are noted in Figure 1. The “elbow”

to expose 35-CCA-37 as a determinant for cognate tRNA^{Gly} (CCC) charging with glycine at 3'-ACCA (at the 3'-O of the ribose ring of 76A). From the primordial sequence, 4 nt was deleted from the 17 nt D loop core, as indicated in the schematic. Also, 1 nt (V₂ or V₃) was deleted within the type I V loop. V₃G forms a triplex interaction with D stem residues D₃G and 25U, as indicated in the schematic.

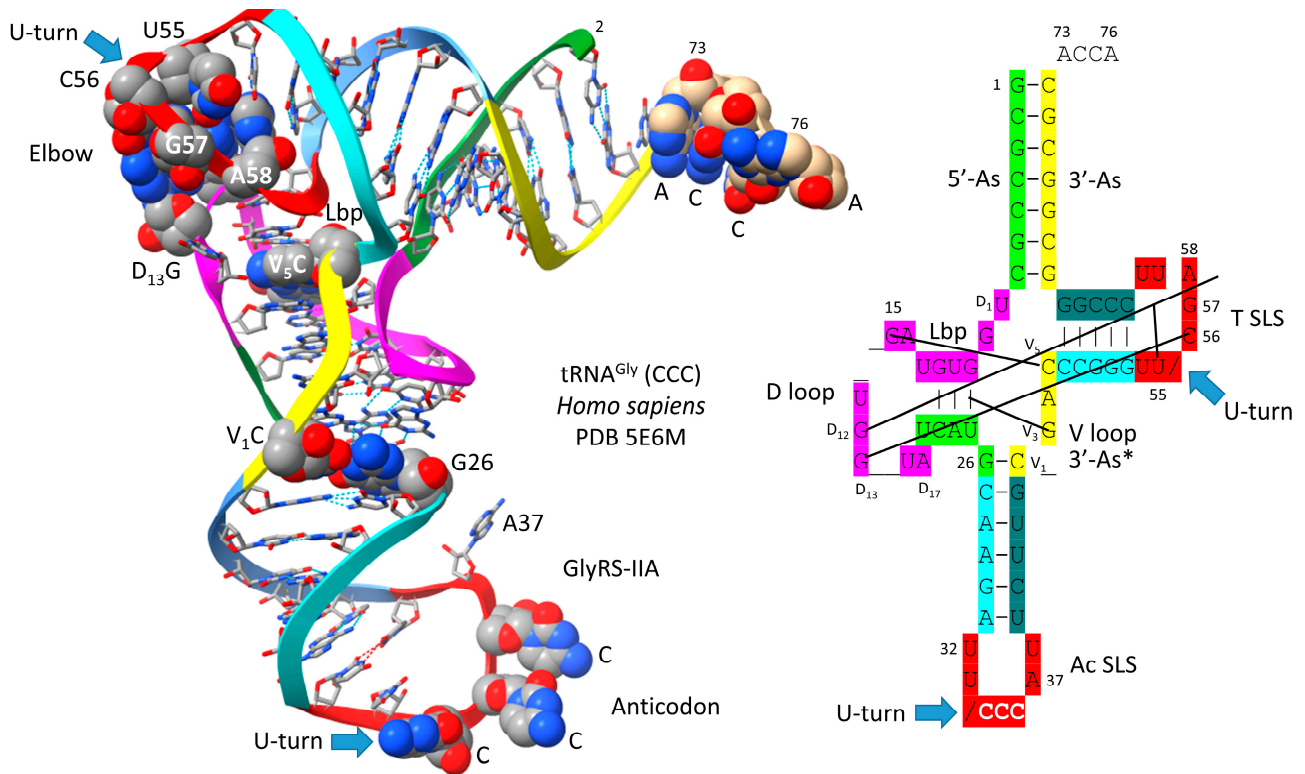


Figure 4. Human tRNA^{Gly} (CCC) from a co-crystal with GlyRS-IIA [20]. D₁₂G is mostly obscured in the image on the left. Anticodon loop residues 35-CCA-37 are unwound to interact with GlyRS-IIA as a determinant for cognate tRNA^{Gly} (GCC, UCC, CCC) charging with glycine. 3'-As* is homologous to tRNA-66 to 70. Colors as in other figures.

In Figure 5, schematic diagrams of a primordial (Pri) type I tRNA and an archaeal *Pyrococcus furiosus* tRNA^{Gly} (GCC) are shown. We have assigned a GCC anticodon to tRNA^{Pri}, as we explain below. The sequences are very similar after ~4.2 billion years, making tRNA^{Gly} derived from an ancient Archaeon a living fossil of the inception of life. It follows that the first tRNAs on Earth were utilized to synthesize polyglycine. In an Archaeon, GlyRS-IIA appears to be the most ancient AARS enzyme. Chemical selection for polyglycine is suggested to have driven the evolution of the first cells. tRNA^{Gly} appears to be the first tRNA from which other tRNAs were derived [21]. In an ancient Archaeon, tRNA^{Gly} is the most similar tRNA to tRNA^{Pri}.

In Figure 6, evolutionary conservation of the first type II and type I tRNAs is summarized. tRNA^{Leu} (CAA) from *P. horikoshii* is shown in line 1 (see also Figure 1). The primordial type II tRNA sequence from which tRNA^{Leu} and tRNA^{Ser} were derived is shown in line 2. A tRNA^{Gly} (GCC) from *P. furiosus*, which is an ancient Archaeon, is shown in line 4 (Figure 5; right panel).

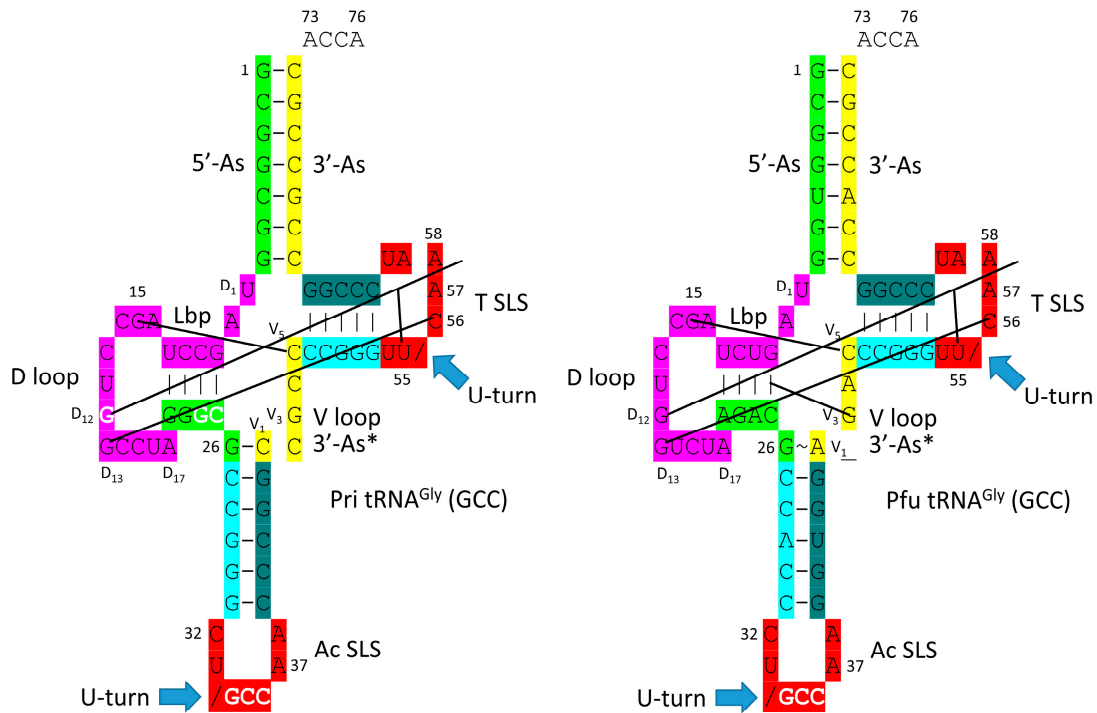


Figure 5. Comparison of tRNA^{Pri} (Pri for primordial) and tRNA^{Gly} (GCC) of *P. furiosus* (Pfu). We suggest that tRNA^{Gly} was the first tRNA, which evolved initially to synthesize polyglycine.

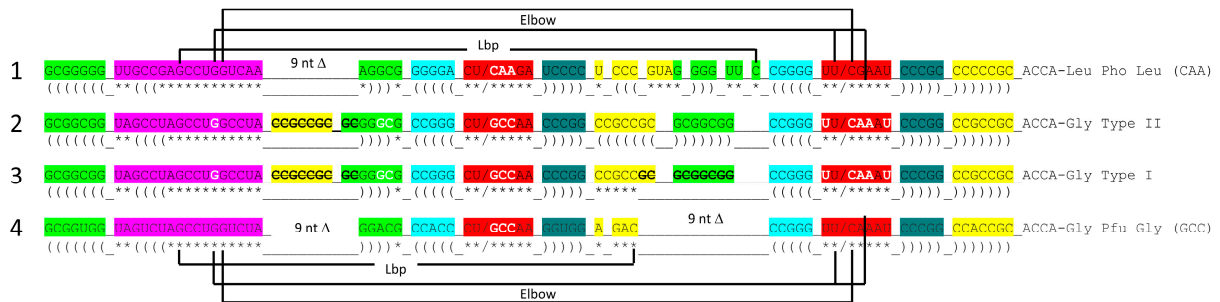


Figure 6. Generation of type II and type I tRNAs at the base of code evolution. Pho: *P. horikoshii* (see Figure 1); Pfu: *P. furiosus* (Figure 4). Levitt base pair (Lbp) and elbow interactions are indicated. Deleted bases are bold with strikethrough. Colors as in other figures.

6. Breaking tRNA^{Pri} into Its Component Parts

To generate complex life supported by a genetic code in a laboratory requires the directed evolution of tRNA^{Pri}. Because tRNA^{Pri} evolved from RNA repeats (GCG, CGC, and UAGCC) and inverted repeats (CCGGG_CU/NNNAA_CCCGG), this goal appears tractable. The only slight deviation from order in tRNA^{Pri} is 3'-ACCA-Gly, which is a short adapter molecule that was attached to many RNAs during the chemical evolution of life [1]. tRNA^{Pri}, therefore, can be broken into its separate parts. Those components can then be separated and combined, and processes can be inferred for transitions that resulted in tRNA^{Pri} evolution. Because tRNA^{Pri} was so highly ordered, its evolution pathway was defined, and intermediates in tRNA evolution were identified. Reproducing transitions between these components would be a major contribution to understanding the evolution of life.

Figure 7 shows the tRNA^{Pri} components that require synthesis. 5'-acceptor stems evolved from GCG repeats. 3'-acceptor stems evolved from complementary CGC repeats. We infer that a chemical mechanism evolved to generate GCG and complementary CGC

repeats on pre-life Earth (Figure 7, lines 1–3). The conservation of GCG and CGC repeats in tRNAs implies a complementary replication mechanism on pre-life Earth. Because processive 5' to 3' ribozyme complementary replication has been somewhat difficult to reproduce in laboratories, perhaps ligation on a template was the mechanism utilized for initial complementary replication [22–26].

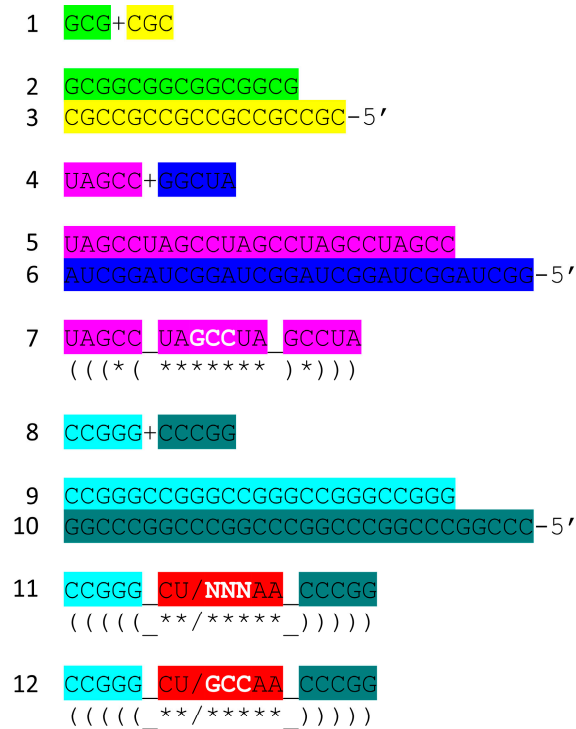


Figure 7. tRNA^{Pri} evolved from RNA repeats and inverted repeats. Colors are as in other figures. See the text for details. We infer that GGCUA repeats (blue) were present as the complement of UAGCC repeats (from which the 17 nt D loop minihelix core of tRNA was derived) (lines 4–7). Lines 8–12 indicate sequences that must be generated to obtain the anticodon and T 17 nt stem–loop–stems. At this time, we do not have a clear idea of how the anticodon and T 7 nt U-turn loops were first generated. It is clear, however, why the 7 nt U-turn loop was selected (see below). Colors as in other figures.

More complex RNA repeats were also synthesized. The 17 nt D loop minihelix core was based on a UAGCC repeat (initially UAGCCUAGCCUAGCCUA) (Figure 7, lines 4–7). Because GCG and CGC repeats are complementary, we infer that GGCUA repeats (blue) were also synthesized on pre-life Earth (lines 4 and 6). We note that a 17 nt UAGCC repeat can fold into a stem–loop–stem that presents a GCC anticodon (line 7). We infer that this molecule could attach 3'-ACCA-Gly for use in polyglycine synthesis using an extended GCG repeat (line 2) as a template. Because the GGCUA repeat (line 6) was not preserved in tRNA sequences, this sequence appears to be extinct now. We infer that many RNA repeats and inverted repeats were present on pre-life Earth, and perhaps only those repeats preserved in tRNA sequences survived the transition to Darwinian selection with the evolution of the first cells.

In addition to the 17 nt UAGCCUAGCCUAGCCUA stem–loop–stem (Figure 7, line 7), an essential 17 nt stem–loop–stem with a 7 nt U-turn loop evolved (Figure 7, lines 8–12) (CCGGG_CU/NNNAA_CCCGG or CCGGG_CU/GCCAA_CCCGG). We argue that the 7 nt U-turn loop (i.e., CU/NNNAA or CU/GCCAA) was the most important innovation in pre-life chemical evolution on Earth. Without this specialized and ribozyme nuclease-resistant loop, tRNA could not have evolved as the genetic adapter. Without a genetic

The greater flexibility of the D loop 17 nt minihelix core, compared to the stiffness of the anticodon stem–loop–stem minihelix, allows for tRNA folding. Ligation of three anticodon stem–loop–stem minihelices (Figure 10, line 1) would be expected to be processed to three anticodon stem–loop–stem 31 nt minihelices (line 2) because of their more stable folding (compare to Figure 9, lines 5–7). These claims can be tested computationally and by experimentation. In evolution, many alternate adapter folds and sequences may have been tested against the pathway that produced type II and type I tRNAs (Figure 9, lines 5–7). The mechanisms that were chemically selected were the fastest mechanisms that resulted in the most successful adapter molecule.

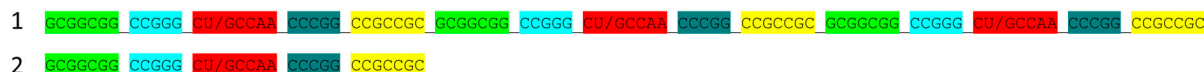


Figure 10. Three ligated anticodon stem–loop–stem 31 nt minihelices are expected to fold and be processed to three 31 nt anticodon stem–loop–stem minihelices, emphasizing the need for the 5′-D loop 31 nt minihelix to generate tRNA (Figure 9; lines 5–7). Colors as in other figures.

Life as we know it on Earth evolved chemically using the RNA adapter tRNA in an aqueous environment. We know of no other chemistries than aqueous chemistry and RNA chemistry that would have been likely to evolve as enabling a genetic adapter as tRNA. We can imagine a sequence substitution for the 17 nt D loop minihelix core, but that substitution likely could not be a 17 nt anticodon stem–loop–stem (compare Figures 9 and 10). tRNA was generated from RNA repeats and inverted repeats that, apparently, were generated accurately on pre-life Earth. Stem–loop–stems chemically evolved, perhaps to cap linear RNAs for accurate complementary replication via ligation (a ribozyme ligase) or to initiate accurate processive replication. Replacing the 7 nt U-turn loop within the anticodon and T stem–loop–stems also appears problematic. The 7 nt U-turn loop is a compact loop that projects a 3 nt anticodon. The 7 nt U-turn loop, furthermore, is expected to have resisted attack by ribozyme nucleases on pre-life Earth. The tight tRNA anticodon loop (see Figure 3), therefore, appears to have been chemically selected versus competing loops. Also, 31 nt minihelices have longer stems than tRNAs (compare Figure 9, lines 1–4 with lines 5–7). Folding into the more complex tRNA, therefore, may have had advantages compared to minihelices. Apparently, tRNAs were easier to melt and replicate on pre-life Earth than minihelices.

10. The Anticodon Loop as Essential Intellectual Property to Evolve Life on Earth

In Figure 11, we show the anticodon loop of *S. cerevisiae* tRNA^{Phe} (GAA) (see also Figure 3) [19]. We argue that the compact 7 nt U-turn anticodon loop was a necessary intellectual property to evolve life on planet Earth. Any attempt to substitute the loop with an RNA loop of another length or alternate sequence would probably be unsuccessful in evolving a code. The U-turn is a U-shaped turn in the anticodon loop backbone. A U-turn loop was necessary to form the tight and compact loop to resist ribozyme endonucleases in the pre-life world. The U-turn also projects three nucleotides to form the anticodon. Initially, both tRNA-34 and tRNA-36 may have had wobble positions [1,3,29]. Wobbling at tRNA-36 was suppressed, in part, by modification of tRNA-37. In the tRNA^{Phe} (GAA) shown, tRNA-37 is modified to YYG (wybutosine; a G modification). At the base of code evolution, reading anticodon tRNA-36A also required a tRNA-37G modification (originally, 37m¹G). To read tRNA-36U required a tRNA-37A modification (originally, 37t⁶A). With unmodified wobble U, tRNA-34U reads mRNA codon 3A, 3G, 3C, and 3U. This is referred to as “superwobbling” and is utilized in mitochondria in 4-codon boxes to shrink the size of

the organelle tRNA^{Aome} [29–31]. To read tRNA-34U in the standard code, therefore, U must be modified to restrict its reading to mRNA wobble 3A and 3G. A first protein, termed elongator Elp3, evolved along with the genetic code to support the use of tRNA-34U and to restrict its reading. tRNA-34A was not utilized at the base of code evolution. To generate the first cells, numerous first proteins coevolved with tRNA^{Aomes} and the genetic code. Because 2'-O-me-C32 and 38A interact (a reverse Hoogsteen interaction), these bases stack with the anticodon stem. Wobbling at tRNA-36 was suppressed, but wobbling at tRNA-34 could not be suppressed in the same way. For one thing, tRNA-33U is on the other side of the U-turn from the anticodon, so modification of tRNA-33U would not alter the reading of tRNA-34. The anticodon loop has specialized properties, modifications, and characteristics that could not easily be substituted by an alternate RNA loop.

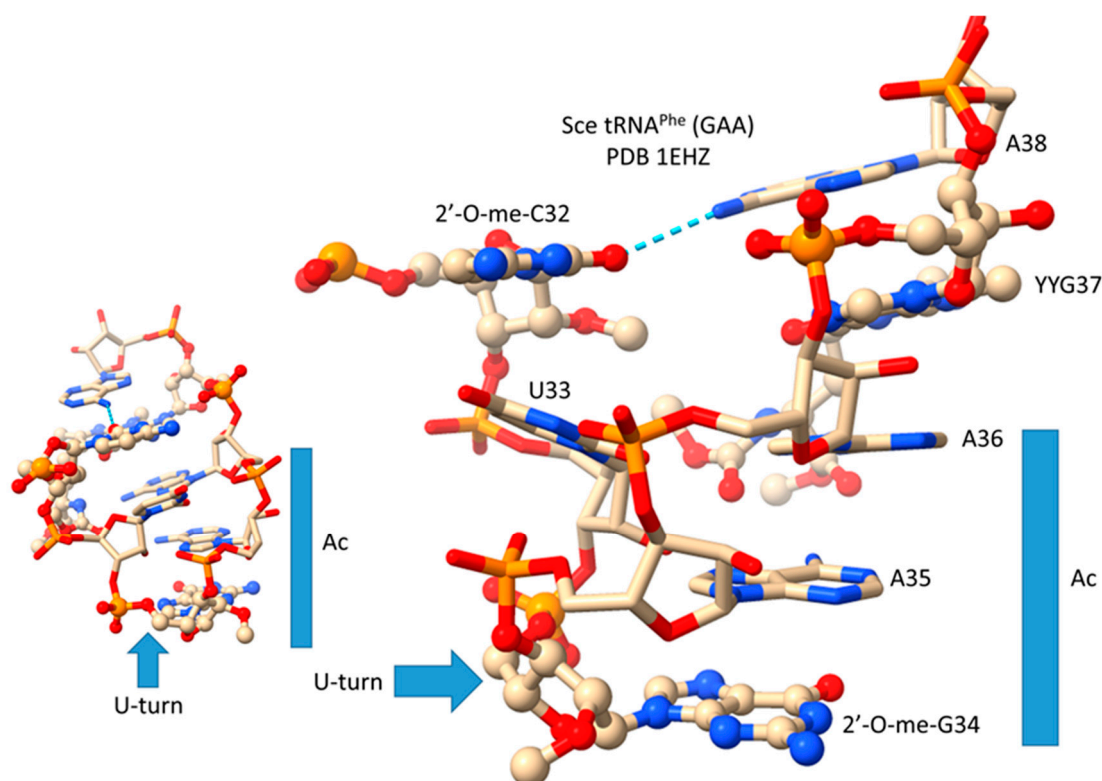


Figure 11. The tRNA anticodon loop. Unmodified bases are in stick representation. Modified bases are in ball and stick representation. Sce for *S. cerevisiae*. YYG for wybutosine (a G modification). Ac for anticodon. Colors: beige: C; blue: N; red: O; orange: P. The view on the left indicates the U-shaped geometry of the U-turn.

11. Determinants on tRNA for Cognate Aminoacyl-tRNA Synthetase Recognition

tRNA^{Aomes} coevolved with the first proteins, aminoacyl-tRNA synthetases (AARSs), which charge tRNAs with their cognate amino acids. AARS enzymes are of two classes: class I and class II. Class II AARSs (i.e., GlyRS-IIA) appear to be more ancient than class I AARSs (i.e., ValRS-IA and LeuRS-IA). Surprisingly, class II and class I AARS enzymes, which have incompatible folds, are homologs by sequence [1,4]. Apparently, class I AARS enzymes were derived from class II AARS by the addition of an N-terminal segment that redirects the fold of class I AARS. We recognize that these data are inconsistent with other models for AARS evolution (i.e., the “urzyme” hypothesis) [32–34].

Most tRNAs are type I. Because only a limited set of V arm trajectories are allowed, only a small number of tRNAs can be type II [2]. In Archaea, only tRNA^{Leu} and tRNA^{Ser}

are type II. Leucine and serine are in six-codon boxes in the standard genetic code. Having five tRNA^{Leu} and four tRNA^{Ser} presented a problem for cognate tRNA charging utilizing the anticodon loop as an AARS determinant, as is utilized for most tRNAs. tRNA^{Leu} and tRNA^{Ser}, therefore, present distinct type II V arms as a major determinant for accurate charging by LeuRS-IA (Figure 1) and SerRS-IIA [2]. Arginine is also within a six-codon sector of the code, but tRNA^{Arg} is a type I tRNA (five tRNAs). ArgRS-IA substantially unwinds the anticodon loop to expose additional bases for cognate tRNA^{Arg} recognition [35]. GlyRS-IIA unwinds the tRNA^{Gly} anticodon loop (three tRNAs) to expose tRNA-35-CCA-37 for recognition (Figure 4) [20]. Apparently, two strategies (type II V arm (Leu and Ser) and unwound anticodon loop (Arg)) were necessary to support three amino acids in six-codon boxes at the base of code evolution [1,2]. Additional determinants are also utilized for cognate amino acid charging including the following: (1) tRNA-73 is the discriminator base, which can be A, G, U or C (initially A); (2) the acceptor stem; (3) the anticodon loop (for all amino acids except alanine, leucine, and serine); (4) the type II V arm (for leucine and serine in Archaea); and (5) the elbow [36,37].

12. Dirty Polyglycine and Emulsification at the Origin of Life

When we consider polyglycine to emulsify pre-cell chemistries, we consider “dirty” polyglycine [1]. Polyglycine, therefore, was part of a background of complementary chemistries. Polyglycine can be modified in many ways on pre-life Earth to increase its length, cross-linking, hydrophilicity, and charge. Many such modifications would potentially render polyglycine a better emulsifier of pre-life chemistry. We suggest that polyglycine be tested for its potential reactivity on pre-life Earth and for its promotion of protocell to cell transitions.

13. Ribozymes and RNA in Pre-Life

Objections to an RNA world include the possible instability of RNA and some limited capacities of ribozymes to catalyze necessary reactions. RNA that is modified at the 2'-O of the ribose ring, however, is as stable as DNA to base hydrolysis [27]. For instance, 2'-O-methyl single-stranded RNAs and ribozymes are stabilized. RNA modifications must be more ancient than the genetic code because multiple tRNA modifications (i.e., Elp3 modification of 34U, 37m¹G, 37t⁶A, 34C→agmatidine, 2'-O-meC) were necessary to generate the code [1,3]. Here, we hypothesize a complex mod-RNA–amino acid–protein–metabolism world [27] (mod-RNA for modified RNA). Our view is supported by analysis of tRNA, tRNA^{ome}, and genetic code evolution.

14. The Three 31 nt Minihelix tRNA Evolution Theorem

There are no theorems (proven models) in biology, but the three 31 nt minihelix tRNA evolution theorem is very close to a proven model. If there is a rational objection to the theorem, we are not aware of it. Other tRNA evolution models have been proposed, but none can be correct [38–52]. No convergent or accretion model can be correct because, at the origin of tRNA^{omes}, all tRNAs are homologous along their entire length. Only a divergent model can be adequate for tRNA evolution. tRNAs evolved from a 93 nt precursor that was processed differentially to generate type II and type I tRNAs. All tRNAs in the tRNA^{sphere} radiated from these forms. No other model can account for internal tRNA homologies. No other model can account for the RNA 3 nt (GCG and CGC) and 5 nt (UAGCC) repeats in tRNA or the conserved inverted repeats (initially ~CCGGG_CU/GCCAA_CCCGG; anticodon and T stem–loop–stems). It appears that tRNA is the most strongly conserved sequence for the pre-life to complex life transition. As such, the tRNA sequence provides a powerful gateway to understand the transition on Earth to complex life supported by

coding. The three 31 nt tRNA evolution theorem is strongly supported by statistical analyses [21,53,54], and most of its features can readily be confirmed by inspection of the conserved sequence. We support universal acceptance of the three 31 nt minihelix tRNA evolution theorem. Without this acceptance, it is unclear that the pre-life to life transition with genetic coding can be understood.

15. Discussion

tRNA appears to be the most highly conserved sequence from pre-life. rRNA is also highly conserved [55], but not as highly conserved as tRNA. Remarkably, the original print of the tRNA sequence has been elucidated, with the exception of those bases scrambled for coding. tRNA was generated from RNA repeats and inverted repeats of known sequences. ACCA-Gly was the primitive adapter molecule. tRNA appears to have evolved initially to synthesize polyglycine. We hypothesize that polyglycine was selected chemically in pre-life for at least two reasons: (1) polyglycine emulsified pre-cellular components to enhance pre-life chemistry; and (2) polyglycine helped form the first protocells and cells [1]. After polyglycine, the genetic code evolved to synthesize GADV polymers [56,57]. At an 8-amino acid stage, the code may have been GADVLSER [1,4,29]. Amino acid-linked chemistry can generate D→N, E→Q and S→C to generate an 11-amino acid stage of code evolution [1,4,29,58]. Suppression of tRNA-36 wobbling allowed the code to expand to 20 amino acids and stops. Fidelity mechanisms froze the code. Because the type II tRNA V arm is a determinant for cognate tRNA charging, only a small set of type II tRNAs can be utilized (in Archaea, tRNA^{Leu} and tRNA^{Ser}) [2]. We suggest that, originally, tRNAs were a mixture of type I and type II that were sorted later in evolution, with most tRNAs selected to be type I.

We have taken a top-down, sequence-based approach to the origin of life. A bottom-up approach would be to reproduce pre-life chemistry in a laboratory [59]. When top-down strategies meet bottom-up approaches, an adequate understanding of the pre-life to life transition should emerge. We were surprised at how powerful the top-down strategy proved to be. We were surprised that tRNA sequences were so highly ordered and relayed such a coherent history of the pre-life to life transition on Earth.

tRNA evolution, structure, and function unify biochemistry, coding, and genetics. Incorporating the evolution of tRNA and its relation to the evolution of coding into instruction will improve science education. For tRNA databases, we advocate for a version of our presentation shown here to further advance the core importance of tRNA and coding evolution at the inception of biology.

Author Contributions: Z.F.B. and L.L. wrote the paper and prepared the figures. All authors have read and agreed to the published version of the manuscript.

Funding: This project received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AARS	Aminoacyl-tRNA synthetase (ligase)
LUCA	Last Universal Common (Cellular) Ancestor
Elp3	Elongator Protein 3
Indel	Insertions and Deletions
Ac	Anticodon
As	Acceptor stem
SLS	Stem-loop-stem

References

1. Lei, L.; Burton, Z.F. Chemical Evolution of Life on Earth. *Genes* **2025**, *16*, 220. [[CrossRef](#)]
2. Lei, L.; Burton, Z.F. Origin of Type II tRNA Variable Loops, Aminoacyl-tRNA Synthetase Allostery from Distal Determinants, and Diversification of Life. *DNA* **2024**, *4*, 252–275. [[CrossRef](#)]
3. Lei, L.; Burton, Z.F. The 3 31 Nucleotide Minihelix tRNA Evolution Theorem and the Origin of Life. *Life* **2023**, *13*, 2224. [[CrossRef](#)]
4. Lei, L.; Burton, Z.F. Evolution of the genetic code. *Transcription* **2021**, *12*, 28–53. [[CrossRef](#)]
5. 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. [[CrossRef](#)]
6. Abe, T.; Inokuchi, H.; Yamada, Y.; Muto, A.; Iwasaki, Y.; Ikemura, T. tRNADB-CE: tRNA gene database well-timed in the era of big sequence data. *Front. Genet.* **2014**, *5*, 114. [[CrossRef](#)]
7. Abe, T.; Ikemura, T.; Sugahara, J.; Kanai, A.; Ohara, Y.; Uehara, H.; Kinouchi, M.; Kanaya, S.; Yamada, Y.; Muto, A.; et al. tRNADB-CE 2011: tRNA gene database curated manually by experts. *Nucleic Acids Res.* **2011**, *39*, D210–D213. [[CrossRef](#)]
8. 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**, *37*, D163–D168. [[CrossRef](#)] [[PubMed](#)]
9. Burley, S.K.; Piehl, D.W.; Vallat, B.; Zardecki, C. RCSB Protein Data Bank: Supporting research and education worldwide through explorations of experimentally determined and computationally predicted atomic level 3D biostructures. *IUCr* **2024**, *11*, 279–286. [[CrossRef](#)]
10. Bittrich, S.; Bhikadiya, C.; Bi, C.; Chao, H.; Duarte, J.M.; Dutta, S.; Fayazi, M.; Henry, J.; Khokhriakov, I.; Lowe, R.; et al. RCSB Protein Data Bank: Efficient Searching and Simultaneous Access to One Million Computed Structure Models Alongside the PDB Structures Enabled by Architectural Advances. *J. Mol. Biol.* **2023**, *435*, 167994. [[CrossRef](#)]
11. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chao, H.; Chen, L.; Craig, P.A.; Crichlow, G.V.; Dalenberg, K.; Duarte, J.M.; et al. RCSB Protein Data Bank (RCSB.org): Delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* **2023**, *51*, D488–D508. [[CrossRef](#)]
12. Meng, E.C.; Goddard, T.D.; Pettersen, E.F.; Couch, G.S.; Pearson, Z.J.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **2023**, *32*, e4792. [[CrossRef](#)]
13. 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](#)]
14. Goddard, T.D.; Huang, C.C.; Meng, E.C.; Pettersen, E.F.; Couch, G.S.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **2018**, *27*, 14–25. [[CrossRef](#)]
15. Fukunaga, R.; Yokoyama, S. Aminoacylation complex structures of leucyl-tRNA synthetase and tRNA^{Leu} reveal two modes of discriminator-base recognition. *Nat. Struct. Mol. Biol.* **2005**, *12*, 915–922. [[CrossRef](#)]
16. 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](#)]
17. 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](#)]
18. 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](#)]
19. Shi, H.; Moore, P.B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: A classic structure revisited. *RNA* **2000**, *6*, 1091–1105. [[CrossRef](#)] [[PubMed](#)]
20. Qin, X.; Deng, X.; Chen, L.; Xie, W. Crystal Structure of the Wild-Type Human GlyRS Bound with tRNA(Gly) in a Productive Conformation. *J. Mol. Biol.* **2016**, *428*, 3603–3614. [[CrossRef](#)]
21. 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](#)]

22. Paul, N.; Joyce, G.F. A self-replicating ligase ribozyme. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12733–12740. [[CrossRef](#)] [[PubMed](#)]
23. Kakoti, A.; Joyce, G.F. RNA Polymerase Ribozyme That Recognizes the Template-Primer Complex through Tertiary Interactions. *Biochemistry* **2023**, *62*, 1916–1928. [[CrossRef](#)] [[PubMed](#)]
24. 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](#)]
25. 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](#)]
26. Horning, D.P.; Joyce, G.F. Amplification of RNA by an RNA polymerase ribozyme. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 9786–9791. [[CrossRef](#)]
27. 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](#)]
28. Burton, Z.F. The 3-Minihelix tRNA Evolution Theorem. *J. Mol. Evol.* **2020**, *88*, 234–242. [[CrossRef](#)] [[PubMed](#)]
29. 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](#)]
30. 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](#)]
31. Rogalski, M.; Karcher, D.; Bock, R. Superwobbling facilitates translation with reduced tRNA sets. *Nat. Struct. Mol. Biol.* **2008**, *15*, 192–198. [[CrossRef](#)] [[PubMed](#)]
32. Tang, G.Q.; Hu, H.; Douglas, J.; Carter, C.W., Jr. Primordial aminoacyl-tRNA synthetases preferred minihelices to full-length tRNA. *Nucleic Acids Res.* **2024**, *52*, 7096–7111. [[CrossRef](#)]
33. Martinez-Rodriguez, L.; Erdogan, O.; Jimenez-Rodriguez, M.; Gonzalez-Rivera, K.; Williams, T.; Li, L.; Weinreb, V.; Collier, M.; Chandrasekaran, S.N.; Ambroggio, X.; et al. Functional Class I and II Amino Acid-activating Enzymes Can Be Coded by Opposite Strands of the Same Gene. *J. Biol. Chem.* **2015**, *290*, 19710–19725. [[CrossRef](#)]
34. Carter, C.W., Jr.; Li, L.; Weinreb, V.; Collier, M.; Gonzalez-Rivera, K.; Jimenez-Rodriguez, M.; Erdogan, O.; Kuhlman, B.; Ambroggio, X.; Williams, T.; et al. The Rodin-Ohno hypothesis that two enzyme superfamilies descended from one ancestral gene: An unlikely scenario for the origins of translation that will not be dismissed. *Biol. Direct.* **2014**, *9*, 11. [[CrossRef](#)]
35. Stephen, P.; Ye, S.; Zhou, M.; Song, J.; Zhang, R.; Wang, E.D.; Giege, R.; Lin, S.X. Structure of Escherichia coli Arginyl-tRNA Synthetase in Complex with tRNA(Arg): Pivotal Role of the D-loop. *J. Mol. Biol.* **2018**, *430*, 1590–1606. [[CrossRef](#)]
36. Giege, R.; Eriani, G. The tRNA identity landscape for aminoacylation and beyond. *Nucleic Acids Res.* **2023**, *51*, 1528–1570. [[CrossRef](#)]
37. Tawfik, D.S.; Gruic-Sovulj, I. How evolution shapes enzyme selectivity—lessons from aminoacyl-tRNA synthetases and other amino acid utilizing enzymes. *FEBS J.* **2020**, *287*, 1284–1305. [[CrossRef](#)] [[PubMed](#)]
38. Demongeot, J.; Seligmann, H. Evolution of small and large ribosomal RNAs from accretion of tRNA subelements. *Biosystems* **2022**, *222*, 104796. [[CrossRef](#)]
39. Demongeot, J.; Seligmann, H. Codon assignment evolvability in theoretical minimal RNA rings. *Gene* **2021**, *769*, 145208. [[CrossRef](#)] [[PubMed](#)]
40. Di Giulio, M. An RNA Ring was Not the Progenitor of the tRNA Molecule. *J. Mol. Evol.* **2020**, *88*, 228–233. [[CrossRef](#)]
41. Demongeot, J.; Seligmann, H. Theoretical minimal RNA rings mimic molecular evolution before tRNA-mediated translation: Codon-amino acid affinities increase from early to late RNA rings. *C. R. Biol.* **2020**, *343*, 111–122. [[CrossRef](#)]
42. 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](#)]
43. Demongeot, J.; Seligmann, H. The primordial tRNA acceptor stem code from theoretical minimal RNA ring clusters. *BMC Genet.* **2020**, *21*, 7. [[CrossRef](#)]
44. 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](#)] [[PubMed](#)]
45. Di Giulio, M. A polyphyletic model for the origin of tRNAs has more support than a monophyletic model. *J. Theor. Biol.* **2013**, *318*, 124–128. [[CrossRef](#)] [[PubMed](#)]
46. Di Giulio, M. The ‘recently’ split transfer RNA genes may be close to merging the two halves of the tRNA rather than having just separated them. *J. Theor. Biol.* **2012**, *310*, 1–2. [[CrossRef](#)] [[PubMed](#)]
47. Di Giulio, M. The origin of the tRNA molecule: Independent data favor a specific model of its evolution. *Biochimie* **2012**, *94*, 1464–1466. [[CrossRef](#)]
48. 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](#)]
49. Di Giulio, M. Transfer RNA genes in pieces are an ancestral character. *EMBO Rep.* **2008**, *9*, 820; Reply in *EMBO Rep.* **2008**, *9*, 820–821. [[CrossRef](#)]

50. Widmann, J.; Di Giulio, M.; Yarus, M.; Knight, R. tRNA creation by hairpin duplication. *J. Mol. Evol.* **2005**, *61*, 524–530. [[CrossRef](#)]
51. Di Giulio, M. The origin of the tRNA molecule: Implications for the origin of protein synthesis. *J. Theor. Biol.* **2004**, *226*, 89–93. [[CrossRef](#)] [[PubMed](#)]
52. Caetano-Anolles, G. Ancestral Insertions and Expansions of rRNA do not Support an Origin of the Ribosome in Its Peptidyl Transferase Center. *J. Mol. Evol.* **2015**, *80*, 162–165. [[CrossRef](#)] [[PubMed](#)]
53. 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](#)]
54. 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](#)]
55. 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](#)] [[PubMed](#)]
56. Ikehara, K. Why Were [GADV]-amino Acids and GNC Codons Selected and How Was GNC Primeval Genetic Code Established? *Genes* **2023**, *14*, 375. [[CrossRef](#)]
57. Ikehara, K. [GADV]-protein world hypothesis on the origin of life. *Orig. Life Evol. Biosph.* **2014**, *44*, 299–302. [[CrossRef](#)]
58. Wehbi, S.; Wheeler, A.; Morel, B.; Manepalli, N.; Minh, B.Q.; Lauretta, D.S.; Masel, J. Order of amino acid recruitment into the genetic code resolved by last universal common ancestor’s protein domains. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2410311121. [[CrossRef](#)]
59. 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](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.