The God Molecule

The God Molecule:

Evolution of tRNA and the Genetic Code

By

Zachary F. Burton and Lei Lei

Cambridge Scholars Publishing

The God Molecule: Evolution of tRNA and the Genetic Code

By Zachary F. Burton and Lei Lei

This book first published 2024

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-0364-1012-4
ISBN (Ebook): 978-1-0364-1013-1

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

PREFACE

ABOUT THIS BOOK

Transfer RNA (tRNA) is the God Molecule because the story of genesis of life is told in tRNA sequences, and because, without tRNA, no complex life could have evolved on Earth. Surprisingly, tRNA evolved from orderly RNA repeats and inverted repeats of known sequence, and the record of evolution was preserved in the tRNAomes of living organisms. Because genetic code can be read, the history of genesis is revealed. Pre-life evolved from a strange Polymer World to a highly ordered Minihelix World to tRNA world, and the record of this history is embedded in tRNA sequences in living organisms. We organisms have lived in a tRNA world for about 4 billion years on Earth. Without a genetic adapter as good or better than tRNA, complex life could not evolve on Earth or another planet.

Analyses of tRNAs and tRNAomes (all the tRNAs of an organism), therefore, provide a rich description of the pre-life to life transition. Furthermore, the genetic code coevolved with tRNAomes, ribosomes and aminoacyl-tRNA synthetases (the enzymes that charge tRNAs with their cognate amino acids). Once the genetic code was established, the sequence-dependent synthesis of proteins became possible, and living cells emerged. This later story is told in the highly structured genetic code and the coevolution of aminoacyl-tRNA synthetases.

These stories are beautiful (many color plates). Because of its ordered sequence, the evolution of tRNA was solved as a puzzle, much easier than a cross-word puzzle. Life evolved from strange genetic building blocks: RNA repeats and inverted repeats. If the evolution of life had been chaotic, as the authors and others initially assumed, such a compelling description of the evolution of life on Earth could not have been elucidated. Because puzzle solutions are provided, the reader need not be concerned about any difficulties encountered working out puzzles.

The standard genetic code is shown in Figure 0-1, as a codon table. The information in this representation is not wrong, but it is limited to describe the evolution of life on Earth. The complexity of the genetic code expressed in this way is 4x4x4=64 codons (61 sense codons encoding 20 amino acids and 3 stop codons). So, this representation is overly complex

and overly degenerate. Code degeneracy means that there are too many synonymous codons encoding too few amino acids. Synonymous codons are those that encode the same amino acid. So, both UUU and UUC codons encode phenylalanine (Phe). Remarkably, UCU, UCC, UCA, UCG, AGU and AGC codons encode serine (Ser).

Life on Earth, however, evolved around the transfer RNA (tRNA) anticodon, and the evolution of the mRNA codon was secondary to the evolution of the tRNA anticodon. Because code degeneracy is a limited function of the tRNA anticodon reading on the ribosome, the maximum complexity of the genetic code that includes anticodon data is 2x4x4=32 assignments. The genetic code evolved, however, to encode 21 assignments, which corresponds to 20 amino acids and stops. To express this information in a more informative way, we wrote this book. If the genetic code is properly considered, the evolution of life on Earth becomes a rational albeit detailed story.

	U	C	A	G		2nd
U	Phe	Ser	Tyr	Cys	U	
U	Phe	Ser	Tyr	Cys	C	
U	Leu	Ser	STOP	Trp	A	
U	Leu	Ser	STOP	STOP	G	
C	Leu	Pro	His	Arg	U	
C	Leu	Pro	His	Arg	C	
C	Leu	Pro	Gln	Arg	A	
C	Leu	Pro	Gln	Arg	G	
A	Ile	Thr	Asn	Ser	U	
A	Ile	Thr	Asn	Ser	C	
A	Ile	Thr	Lys	Arg	A	
A	Met	Thr	Lys	Arg	G	
G	Val	Ala	Asp	Gly	U	
G	Val	Ala	Asp	Gly	C	
G	Val	Ala	Glu	Gly	A	
G	Val	Ala	Glu	Gly	G	
1st					3rd	

Figure 0-1. The standard genetic code expressed as a codon table. 1st, 2nd and 3rd indicate codon positions, read 5' to 3', as is standard. Sets of three bases (i.e., U, C, A and G; 3 bases = 1 codon) encode 1 amino acid, as indicated (i.e., UUU and UUC encode phenylalanine (Phe)). The information in this figure is not incorrect, but this description is limited to understand evolution of life on Earth.

Transfer RNA (tRNA) is a relic of pre-life and the pre-life to life transition on Earth about 4 billion years ago (Lei and Burton 2023).

Embedded in the tRNA sequence are two recently discovered stages of pre-life, “Polymer World” and “Minihelix World”.

In biology, consider systems to be evolved such that there is a reason for any observation and that the reason is evolution and natural selection. Evolution is the driving force creating differences among molecules, systems and organisms. So, consider that there is a reason for observations and that evolution can provide the explanation.

Years ago, when we started looking at tRNA evolution, we thought solving the problem was an impossible task. We thought that tRNA evolved by a chaotic process and that ~4 billion years of evolution would have overwritten any original print of pre-life, such as the primordial tRNA sequence. We were wrong, and we were not initially very clever. TRNA evolution is a very straightforward problem, very similar to solving a puzzle, like a cross-word. At least one of the authors of this book hates working out cross-word puzzles. They give him frustration and a pounding headache. Despite some headaches, he was much happier working out a puzzle partly of his own creation but mostly of biology’s and the Earth’s creation. The evolution of tRNA is a better puzzle than one found in a newspaper.

TRNA is the most consequential primordial biological molecule (Figure 0-2). The most important feature of tRNA is the anticodon stem-loop-stem (Figure 0-3). The essential feature of the tRNA 7 nt (nt for nucleotide) anticodon loop is the U-turn, which is a U-shaped turn in the RNA backbone. This feature allows for the presentation of a 3 nt anticodon that can interact in antiparallel orientation (5’→3’) with a 3 nt codon in messenger RNA (mRNA). This is the basis for the 3 nt genetic code that supports complex life on Earth. TRNA is a living molecular fossil from the birth of genetics and cells.

One way to look at this book is from the point of view of a collector of antiquities. The tRNA sequence evolved by a largely known chemical process ~4 billion years ago, before the first living organisms and cells on Earth. The reader is enjoined to collect the tRNA sequence and structure as one might collect a priceless antique, far beyond value. How many of your friends possess antiquities or family heirlooms that are ~4 billion years old? For context, the Earth is 4.6 billion years old. Because we all have bodies that encode tRNAs, the authors guess we all possess such wonders of evolution and creation.

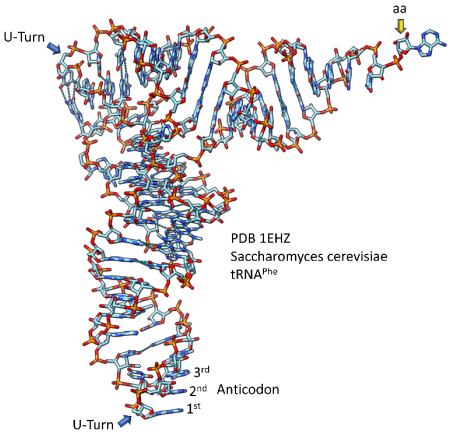

Figure 0-2. The God Molecule. TRNA^{Phe} from Saccharomyces cerevisiae (brewer's yeast). U-turns, amino acid (aa) placement and the anticodon (GAA) are labeled. Colors: cyan) carbon; red) oxygen; blue) nitrogen; and orange) phosphorous. Hydrogens are not shown. The RCSB (Research Collaboratory for Structural Bioinformatics) protein data bank identifier is shown (PDB 1EHZ) (Shi and Moore 2000). TRNA is the most important intellectual property in evolution of life on Earth. Molecular graphics were done using ChimeraX (Pettersen *et al*. 2021). 1st, 2nd and 3rd indicate anticodon bases.

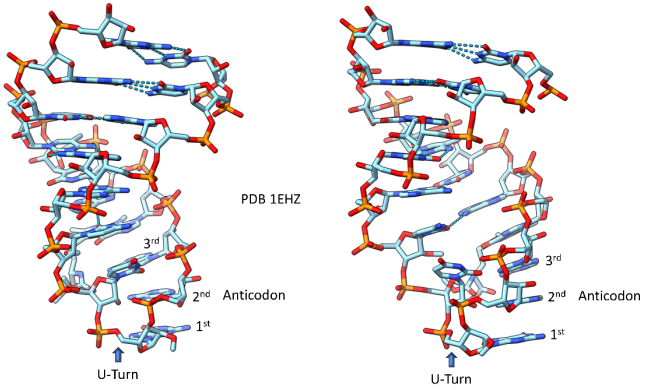

Figure 0-3. The anticodon stem-loop-stem of tRNA^{Phe} (two views). The U-turn is a U-shaped turn in the RNA backbone. The 7 nt U-turn loop in the tRNA anticodon stem-loop-stem is the most important intellectual property in the evolution of life on Earth. 1st, 2nd and 3rd indicate anticodon bases.

As we describe, tRNA evolved from highly regular sequences: RNA repeats and inverted repeats (stem-loop-stems). These ancient, ordered sequences were maintained and can be identified in tRNAs in living organisms today. If the tRNA sequence was not ordered, the root sequence of a primordial (pre-life) tRNA ($tRNA^{Pri}$) could not have been determined.

Because $tRNA^{Pri}$ was made up of recognizable RNA repeats and inverted repeats, it must have been possible to synthesize repeat sequences and inverted repeats in a pre-life world. This makes tRNA sequence a written history of pre-life chemical evolution. The history was recorded in genetic code (A, G, C and U), which is straightforward to read. So, humans are not necessary to write history, because biology on Earth wrote history in genetic code. Humans, however, may be capable to read this natural biological history of Earth by reading genetic code. Genetic coding in tRNA presents a written history of pre-life on Earth before the evolution of the genetic code. The written history in the tRNA sequence predates the capacity to fully encode proteins and predates cells and intact genomes. Before we did this work, if you had told the authors that the tRNA sequence preserves a written natural biological history of life before cells, we would have thought you insane, so we feel your scepticism.

So, tRNA is a historical record of pre-life (Lei and Burton 2023). TRNA is also the most central feature of evolution of the genetic code. TRNA's centrality makes tRNA the God Molecule. Without tRNA, there is no life on Earth, because there is no genetic code. Without a code, there are no encoded proteins. If there is another approach to evolve complex life chemically, we would like to know what it might be. We have no reasonable ideas other than tRNA. When God is discussed in this book, God is a metaphor for the chemical evolution of life on Earth. Personally, the authors remain somewhat agnostic with regard to the existence of God.

If you are enthralled by the search for life on another planet, in a galaxy, far, far away, you are not alone in your quest. When you meet an alien, find out the sequence of their genetic adapter. If it is tRNA, like our tRNA, consider panspermia. If it is different to our tRNA, is it RNA? If it is different to our tRNA and not made of RNA, how, in God's name, was it evolved? We are very curious. It is a big challenge to think of a replacement for tRNA as a genetic adapter. We do not believe complex life could possibly evolve without a genetic adapter as good (or better) than tRNA.

Also, there is a kinetic aspect to the evolution of life on Earth and evolution of the genetic adapter. With tRNA and here on Earth, the fastest technology to evolve an adequate genetic adapter won. TRNA evolved by ligation of three 31 nt minihelices (Lei and Burton 2023). Two of these minihelices were initially identical. So, RNA ligation was important in

evolution of tRNA, as we will describe (Chapter 1). In the pre-life world, RNA ligation was supported by a ribozyme RNA ligase that generated complex RNA molecules that included RNA repeats. When translation evolved, RNA repeats were translated into protein repeats. Protein repeats allow pseudosymmetry to generate protein barrels, which were key features of protein evolution, as previously described in the book *Evolution Since Coding* (Burton).

The idea of RNA ligation in pre-life is simple. In pre-life, RNAs concentrated and were replicated chemically in colonies. RNAs were replicated by ligation, priming, copying and processing, mostly by ribozymes. Many complex RNAs were generated that sometimes included RNA repeats, synthesized by ligating identical RNAs together. TRNAs include a record of RNA ligations in the creation of Polymer World, Minihelix World and tRNA World. Once translation systems evolved, ubiquitous RNA repeats were translated into protein repeats. Protein repeats can fold into pseudosymmetrical barrels that have stability, solubility and structural closure. The chemical evolution of a small collection of enduring pseudosymmetrical protein barrels describes the evolution of glycolysis (sugar metabolism), the tricarboxylic acid (TCA) cycle, RNA synthesis and DNA synthesis. One of the primary mechanisms for the pre-life chemical evolution of living systems was RNA ligation, protein repeat pseudosymmetry to form barrels and sometimes pseudosymmetry breaking to form sheets (i.e., Rossmann folds). Generating complex RNAs by ligation during RNA copying amplified the repertoire of chemical complexity in the pre-life world. The authors of this book consider RNA ligation and protein pseudosymmetry to be fundamental concepts to make sense of the rapid chemical evolution in pre-life. RNA ligation and protein pseudosymmetry were major accelerants in the evolution of life on Earth.

So, at the core, the meaning of life is tRNA. Without tRNA, there is no life and no genetic code. If there are no genetically encoded proteins, there is no complexity to organisms or a basis for complex genetics. TRNA specifically supported a 3 nt (nt for nucleotide) genetic code. There is no complex life with a 2 nt genetic code unless the number of encoding base pairs is expanded, and this would cause other problems, such as error catastrophe caused by improper base pairing. A 4 nt genetic code, probably, could not evolve and, also, would not be reasonably supported by tRNA that, specifically, has an anticodon loop conformation to support a genetic code with a 3 nt register. So, the tRNA structure and tRNA conformation place limits on the complexity of evolved life on Earth.

The story of tRNA evolution is also a lesson in biological artificial intelligence and “machine” learning, because emerging life taught itself

(chemically evolved) to code. TRNA evolved to tRNAomes (all of the tRNAs of an organism) encoding many amino acids. The ribosome and associated factors coevolved to generate a functional translation system. The components of translation systems taught themselves through coevolution to code and to code with sufficient accuracy to support emerging cellular systems. Most of the systems that convert nucleotide sequence (RNA and DNA) to amino acid sequence (proteins) via a genetic adapter (tRNAs) coevolved before the evolution of the first cells in a pre-life world. The process of protein synthesis is referred to as "translation" because a language written in RNA sequence is translated via a tRNA adapter to a protein polymer sequence of amino acids with a (generally) higher order catalytic and structural function.

In teaching, start with the hook and then attempt to inspire. Ultimately, people learn what they teach themselves and not necessarily what they hear in lectures or read in a book. All of the headaches and pain in learning tRNA evolution was ours, we think. You can now teach yourself this material with much more facility.

The reader, therefore, is enjoined to *know thine genetic adapter*, tRNA (Figures 0-2 and 0-3). We hope this short preface is a sufficient tease to encourage further reading. The reader is enjoined to attack tRNA using this book, websites, sequence, molecular graphics (i.e., ChimeraX) and a computer. TRNA is a good problem. It is a problem that is solved, so the solution is known and can be explained. To an extent, the contents of this book are similar to solving a puzzle with the puzzle solution in front of you. Cheating is ok. Cheating is encouraged. Cheat on. If you can solve tRNA evolution without the help of the authors in a reasonable span of time, compared to us, you are very clever.

References

Lei, L., and Z. F. Burton. 2023. "The 3 31 Nucleotide Minihelix tRNA Evolution Theorem and the Origin of Life." *Life (Basel)* 13 (11). doi: 10.3390/life13112224.

Pettersen, E. F., T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, and T. E. Ferrin. 2021. "UCSF ChimeraX: Structure visualization for researchers, educators, and developers." *Protein Sci* 30 (1):70-82. doi: 10.1002/pro.3943.

Shi, H., and P. B. Moore. 2000. "The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited." *RNA* 6 (8):1091-105. doi: 10.1017/s1355838200000364.

CHAPTER 1

RNA LIGATION AND PROTEIN PSEUDOSYMMETRY

Abstract

RNA ligation, RNA repeats, protein repeats and protein pseudosymmetry were major accelerants to rapid chemical evolution that gave rise to life on Earth. In this introductory chapter, the authors describe how type II tRNA, type I tRNA, $(\beta-\alpha)_8$ protein barrels, refolded $(\beta-\alpha)_8$ protein sheets (Rossmann folds), cradle-loop barrels (i.e., double-Ψ−β-barrels) and AARS enzymes may have originated via RNA ligations, RNA repeats, protein repeats, protein pseudosymmetry and, in the case of Rossmann folds, protein pseudosymmetry breaking, during pre-life.

Previous Publication

The work and figures in this chapter have previously been published in an open access journal and are republished here in a similar form with permission (Lei and Burton 2023).

The pre-life world was highly ordered, and this was a surprise. The most enduring molecules of pre-life were generated by ordered processes. Often, the fastest processes to generate molecules with structural integrity, chemical value and closure was the process that succeeded in producing molecules that endured, making pre-life chemical evolution and competition into something of a race. TRNA was generated from highly ordered molecules: RNA repeats and inverted repeats. Similar RNAs were ligated together (i.e., by ribozyme ligases) to generate the first type I and type II tRNAs. Many of the first proteins that coevolved with the genetic code were generated by ordered processes: i.e., ligation of similar or identical RNAs to form RNA repeats that were subsequently translated into proteins with recurring motifs and structural closure. The authors of this book initially assumed that pre-life chemical evolution was chaotic, but we were wrong.

The pre-life world was complex. The authors imagine a pre-life world dominated by RNA, with ubiquitous RNA-amino acid, RNA-peptide and RNA-chemical linkages (Lei and Burton 2023, 2021b, 2020, Kim, Opron, and Burton 2019, Muller et al. 2022, Chatterjee and Yadav 2022). A primitive but complex metabolism must have existed. Many RNA, polypeptide and carbohydrate polymers were present. Protocells evolved that would become the first true cells at LUCA (the last universal common (cellular) ancestor). In our conception, LUCA represents the first fully enveloped cells with intact DNA genomes. Remarkably, some of the pre-life world need not be purely imagined because relics of pre-life have survived throughout ~4 billion years of organismal evolution to present day (Weiss et al. 2018, Weiss et al. 2016). One purpose of this book is to describe some of the living molecular fossils that were conserved from pre-life to present day. People consider a RNA World, a RNA-Protein World, a Metabolism World, etc. The authors advocate for a complex pre-life world that, nonetheless, can be described in some detail and with authority.

A schematic recipe for evolution of life on Earth is shown in Figure 1-1. How tRNA evolved is a solved problem, because the history for chemical evolution of tRNA was recorded and preserved in tRNA sequences (Lei and Burton 2023). Type I tRNA evolved by ligation of three 31 nt minihelices of almost completely known sequence. The tRNA precursor molecule was 3x31=93 nt. Type I tRNA was generated by two internal 9 nt deletions. 93-18=75 nt. Addition of 3'-ACCA by ligation completes type I tRNA (79 nt). Type II tRNA evolved from the same 93 nt tRNA precursor molecule. A single 9 nt internal deletion generated the first type II tRNA (93-9=84 nt). Addition of 3'-ACCA generated an 88 nt type II tRNA. Because tRNA was a molecule that on pre-life Earth could "teach" itself to code, chemical evolution of tRNA led inexorably to evolution of the genetic code and life on Earth.

Figure 1-2 shows the concept of RNA ligation creating RNA repeats applied to type II tRNA evolution. The sequence of tRNA shows that both type II and type I tRNAs were generated via RNA repeats. In the process, three 31 nt minihelices were ligated together. The most 5' 31 nt minihelix became the D loop of type II tRNA. The next 31 nt minihelix became the anticodon stem-loop-stem. The third 31 nt minihelix that formed the T stem-loop-stem was initially identical to the anticodon stem-loop-stem minihelix. The T 17 nt stem-loop-stem remains homologous (genetically related) to the anticodon 17 nt stem-loop-stem, today (Lei and Burton 2023).

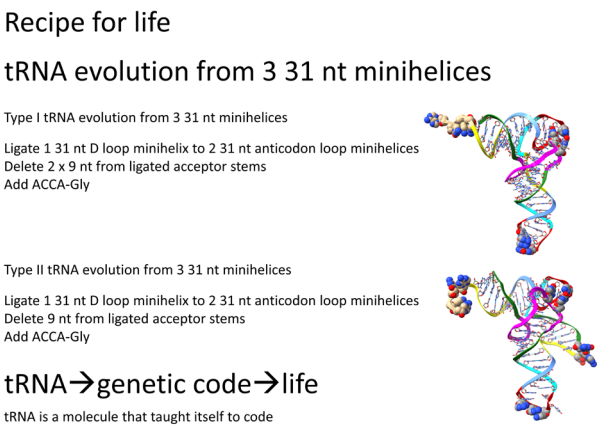

Figure 1-1. Recipe for life. See the text for details.

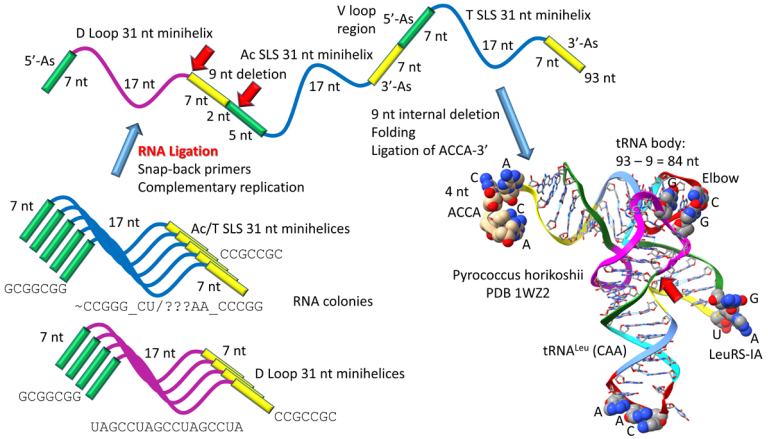

Figure 1-2. Type II tRNA evolved from ligation of RNAs. Colonies of 31 nt minihelices were replicated by ligation, snap-back priming, complementary replication and processing. Type II tRNA was produced by an internal 9 nt deletion within ligated 3'- and 5'-acceptor stems (As) (ligating yellow segments to green segments), as shown. ACCA-3' may have initially been ligated to the 3'-ends of minihelices and type II tRNAs. The type II tRNA was coloured according to the three 31 nt minihelix theorem (see Figure 1-6, below). Red arrows (i.e., where magenta joins green) indicate positions at which the 93 nt tRNA precursor was processed by an internal 9 nt deletion.

In the 31 nt Minihelix World, replication of minihelices required ligation of RNAs. A ribozyme RNA ligase was undoubtedly present (DasGupta, Zhang, and Szostak 2023, Nomura and Yokobayashi 2023, DasGupta et al. 2023). 31 nt minihelices were capped by a 7 nt 5'-acceptor stem (green segments) and a 7 nt 3'-acceptor stem (yellow segments). 31 nt minihelices concentrated in colonies that included identical and similar RNAs. The history of pre-life on Earth from ~4 billion years ago was embedded in the sequences of tRNAs and can easily be extracted by just about anyone. With guidance, the puzzle is not difficult to solve.

Minihelix replication required snap-back primers, which are RNA stem-loop-stems. Because 5'- and 3'-acceptor stems are complementary, any 31 nt minihelix could fulfil the role of a snap-back primer. We assume copying of ligated minihelices to be by primed and complementary replication. The existence of stem-loop-stems strongly indicates complementary replication. So far, scientists have had difficulty generating a highly processive and accurate RNA replicase, but scientists may yet succeed (McGinness and Joyce 2003). Because of the existence of stem-loop-stems in the pre-life world, it appears that a complementary replicase ribozyme must have existed. Replication would be facilitated by a ribozyme helicase or chaparone that scientists have succeeded in synthesizing. The lengths of RNA stems may have been selected based on the capabilities of a ribozyme chaparone to unwind them. Complementary stems in RNAs that were excessively long may have inhibited replication by inhibiting strand separation. Judging from tRNA sequences, there may have been strong chemical selections for complementary CG-rich RNA stems of 5 and 7 nt.

The authors have considered the idea that pre-life complementary replication may have initially involved assembly of complementary segments of RNA on a template RNA followed by ligation by a ribozyme ligase. To our knowledge, no one has established such a ribozyme system for complementary replication in vitro. Such a system might be expected to select for a limited repertoire of short complementary RNA sequences, consistent with the formation of minihelices and tRNAs that include these modules.

ACCA appears to have been among the most ancient adapter molecules to attach amino acids (i.e., ACCA-Gly). We posit that ACCA or ACCA-Gly was attached by ligation to many RNAs in the pre-life world. Attachment of an amino acid would protect ACCA from ribozyme 3'→5' exonucleases. In Polymer World, ACCA-Gly could base-pair to a GCG repeat (i.e., GCGGCG) by Watson-Crick pairing to GG. Bringing ACCA-Gly molecules into proximity on a GCG repeat scaffold is expected to be

sufficient to synthesize polyglycine, which we posit to have been a critical component of early protocells (Lei and Burton 2021b, 2020). We posit that Minihelix World was an improved system to synthesize polyglycine. We posit that, initially, tRNA World evolved as an advanced system to synthesize polyglycine. Someone should develop these systems in a laboratory to test the value of GCG repeats, ACCA-Gly, minihelices and tRNAs in synthesis of polyglycine. Also, the value of polyglycine and other pre-life polymers in manipulations of protocells must be tested. In principle, note that we have solved the "chicken and egg" problem of evolution of translation systems to encode RNA sequence-dependent proteins (Lei and Burton 2023, 2021b, 2020). The first translation systems evolved to synthesize polyglycine as an essential component of protocells. Subsequently, tRNA World evolved to match tRNA anticodons and mRNA codons to encode 20 amino acids and stops with mRNA sequence-dependent synthesis of proteins. TRNA World was a system that, by trial and error, could chemically "teach" itself to code.

Type II tRNAs have a longer V loop (V for variable) than type I tRNAs. Initially, the type II V loop was 14 nt (7 nt + 7 nt; a 7 nt 3'-acceptor stem ligated to a 7 nt 5'-acceptor stem; Figure 1-2). In evolution, longer and shorter lengths of type II V loops were selected. In Archaea, $tRNA^{Leu}$ and $tRNA^{Ser}$ have type II V loops. In Bacteria, $tRNA^{Leu}$, $tRNA^{Ser}$ and $tRNA^{Tyr}$ have type II V loops. So, most tRNAs are type I.

Type I tRNAs evolved by RNA ligation and RNA processing (Figure 1-3). The mechanism is very similar to that for type II tRNAs (Figure 1-2). The only difference is that an additional 9 nt internal deletion occurred in the 14 nt V loop region of the same 93 nt tRNA precursor (Figure 1-3). 7 nt of 5'-acceptor stem (green) and 2 nt of 3'-acceptor stem (yellow) were deleted. The resulting type I V loop was 5 nt (yellow) derived from a 3'-acceptor stem. Type II and type I tRNAs, therefore, were initially identical molecules except for the length of the V loop, which was initially 14 nt for type II tRNAs and 5 nt for type I tRNAs.

Appreciation of RNA ligation and orderly RNA processing provides new insights into the pre-life world. RNA ligations produced novel RNAs, many of significant length and complexity. Numerous novel ribozymes were generated from such complexity. Precursors to the peptidyl transferase center (23S rRNA) and decoding center (16S rRNA) were generated. Selections for RNA stem and loop lengths occurred. CG-rich stems of 7 and 5 nt appear to have been favoured. In evolution of tRNA, a particular loop that was favoured was 7 nt in length with a U-turn between loop positions 2 and 3 and a reverse Hoogsteen interaction between loop positions 1 and 7 (see below). This is the tRNA anticodon

loop that supported evolution of the 3 nt genetic code (Pak, Root-Bernstein, and Burton 2017). It is a tight loop that may have resisted attacks by pre-life ribozyme ribonucleases.

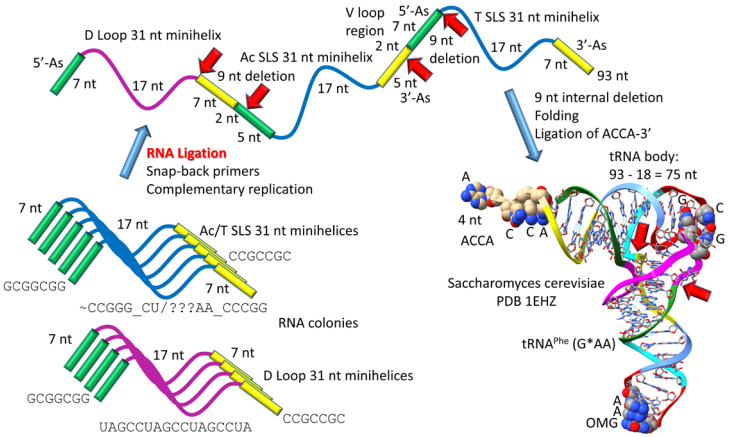

Figure 1-3. Chemical evolution of type I tRNA by RNA ligation and internal RNA processing. Colonies of 31 nt minihelices were replicated by ligation, snap-back priming, complementary replication and processing. Two internal 9 nt deletions produced type I tRNA (red arrows). The 5' 9 nt internal deletion was identical to that which occurred in evolution of type II tRNA (Figure 1-2). Red arrows indicate RNA processing positions.

An unmodified type II $tRNA^{Leu}$ (CAA) from an ancient Archaea is shown in Figure 1-4. In our view, Pyrococcus horikoshii is little different than LUCA, particularly for translation functions (i.e., tRNA, tRNAomes and ribosomes). The image was coloured according to the three 31 nt minihelix theorem. A natural modified type I $tRNA^{Phe}$ (G*AA) from Brewer's Yeast, a Eukaryote, is shown in Figure 1-5. Much of this book is a discussion of these structures and their evolutionary meanings. Structures were coloured according to internal homologies. The relevant difference between type II and type I tRNA structures is the length of the V loop (V for variable). In Figure 1-4, the type II V loop of $tRNA^{Leu}$ (CAA) is the pre-life length of 14 nt corresponding to 7 nt of 3'-acceptor stem (yellow) ligated to 7 nt of 5'-acceptor stem (green). The V loop of $tRNA^{Phe}$ (G*AA) (G* indicates 2'-O-methyl-G), by contrast, is a 5 nt yellow segment derived from a 3'-acceptor stem. The reader is enjoined to analyse these structures in detail and to refer back to these figures as reference in further reading. The best way to approach the analysis is to

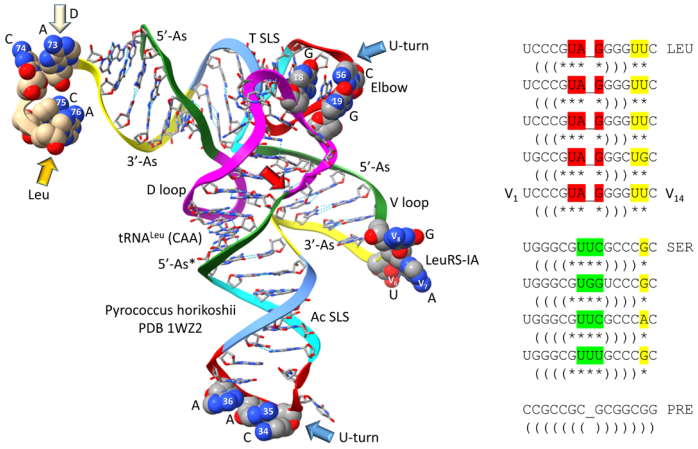

Figure 1-4. A type II tRNA^{Leu} (CAA) from Pyrococcus horikoshii, an ancient Archaeon. As) acceptor stem; As*) acceptor stem fragment; SLS) stem-loop-stem (inverted repeat); Ac) anticodon; D) discriminator. Arrow colors: orange) amino acid; light orange) discriminator; blue) U-turn; red) RNA processing site. The V_6-UAG-V_8 segment (red in right panel) in the V loop is a recognition determinant for tRNA^{Leu} charging with leucine by archaeal LeuRS-IA. The right panel shows P. horikoshii V loop sequences for tRNA^{Leu} and tRNA^{Ser} (Juhling et al. 2009).

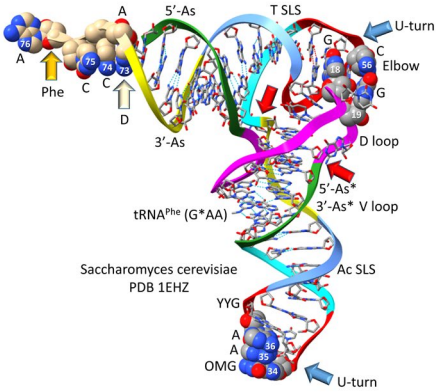

Figure 1-5. A type I tRNA^{Phe} (G*AA) from Saccharomyces cerevisiae (Brewer's Yeast), a Eukaryote. OMG) 2'-O-methyl-G (G*); YYG) wy butosine.

learn to use ChimeraX by running tutorials and to analyse the structures using a personal computer.

The anticodon (Ac) stem-loop-stem and the T stem-loop-stem are homologs (cyan-red-cornflower blue). In evolution of tRNAs from 31 nt minihelices, anticodon and T stem-loop-stem sequences were initially identical. The loop segments (red) have changed somewhat in evolution from pre-life to LUCA. The 7 nt T loop was selected to support "elbow" interactions with the D loop (magenta). For historic reasons, tRNA numbering breaks down within the D loop, because numbering was based on eukaryotic tRNA D loops with 3 deleted bases from $tRNA^{Pri}$. D loop G18 intercalates between T loop bases 4 and 5, lifting loop base 5 to fill the loop and to flip loop bases 6 and 7 out of the loop. D loop G19 forms a bent Watson-Crick base pair with T loop base C3 (tRNA-56). These details will be discussed in more detail below.

The 7 nt anticodon (Ac) loop within the anticodon stem-loop-stem is an important structure. In its closed conformation, this is a tight RNA loop that projects the 3 nt anticodon (Figure 1-5). The wobble base, tRNA-34, which forms both Watson-Crick and wobble interactions to mRNA wobble codon 3, is often modified. OMG is 2'-O-methyl-G34. The tRNA-36 and tRNA-35 anticodon bases are rarely modified. These bases form Watson-Crick interactions with mRNA codon 1 and codon 2. The tRNA-37 base is often modified (i.e., G→wy butosine). The tRNA-37 base is part of the reading of the tRNA-36 base. The tRNA-36 base, we posit, was once a wobble position. In terms of coding, wobble positions only achieve purine versus pyrimidine resolution (2 possibilities), as we describe in more detail below. Both the anticodon loop and the T loop include a U-turn structure after an encoded U base at loop position 2. The U-turn creates the tight closed anticodon loop conformation that is important for projection of the 3 nt anticodon and is also important for coding (Figure 1-5). By contrast, the type II $tRNA^{Leu}$ (CAA) anticodon loop shown in Figure 1-4 is somewhat unwound because of the structure from which the tRNA image was taken.

With the exception of the pre-life adapter ACCA-3', the pre-life $tRNA^{Pri}$ (Pri for primordial) was comprised of RNA repeats and inverted repeats. Green segments were initially GCG repeats. Yellow segments were initially CGC repeats. The magenta segment was initially a 17 nt UAGCC repeat. The anticodon and T stem-loop-stems were initially something close to ~CCGGG_CU/???AA_CCCGG (_ separates stem and loop sequences; / indicates a U-turn; ? indicates that the pre-life sequence may not now be known). For the pre-life sequence, ambiguity is only in the ~CU/???AA (/ indicates a U-turn) 7 nt loop not in the stems. These

issues are discussed in more detail below. Otherwise, the $tRNA^{Pri}$ sequence from pre-life is known with authority, and dominant sequences from LUCA are known with essential certainty. Type II and type I tRNAs, therefore, are living fossils of pre-life from ~4 billion years ago on Earth.

The pre-life discriminator base (D) was initially A. In tRNAs, the discriminator base can be any base, as described in more detail below. The discriminator base is often a determinant for accurate placement of the cognate amino acid at the 3'-end of the tRNA. The two Cs of 3'-CCA (tRNA-74C and -75C) are sites of attachment for tRNAs in the A-site (acceptor site) and P-site (peptidyl site) of the peptidyl transferase centre of the ribosome. 3'-A is the site of amino acid attachment. Cognate amino acids are attached at the 2'- or 3'-OH of the ribose ring of tRNA-76-3'-A. Class I aminoacyl-tRNA synthetases (AARS) utilize the 2'-OH for amino acid attachment. Class II AARS enzymes utilize the 3'-OH for attachment (Tawfik and Gruic-Sovulj 2020).

In Figure 1-3, in Archaea, the UAG motif in the V loop is bound as a LeuRS-IA determinant for accurate attachment of leucine to the $tRNA^{Leu}$ (CAA). The V_6-UAG-V_8 (i.e., Pyrococcus) V loop determinant is specific for archaeal LeuRS-IA. Comparing Archaea and Bacteria, there are significant changes in LeuRS-IA $tRNA^{Leu}$-recognition determinants, discussed below.

In tRNA, a stack of bases runs from the T loop through the anticodon stem. T loop bases 6 and 7 are flipped out of the T loop because of intercalation of G18 of the D loop (an "elbow" contact) between T loop bases 4 and 5. The flipped out T loop bases 6 and 7 stack on the D stem that stacks on the anticodon stem. More discussion of these details will follow. There is much to learn about tRNA structure and evolution.

Embedded in tRNA sequence is a clear record of tRNA evolution from a Polymer World to a Minihelix World to tRNA World. TRNA sequence provides evidence for three 31 nt minihelices (Figure 1-6). Minihelices were chemically evolved and assembled from regular RNA polymers synthesized in Polymer World. Minihelix World included very regular 31 nt minihelices, as shown. In pre-life, the only sequence ambiguity is in the red segments (~CU/???AA). After LUCA, because of different selection pressures, the dominant sequence was slightly different for the anticodon loop (~CU/BNNAA; B is G, C or U (not A); N is any base) and the T loop (UU/CAAAU). All other $tRNA^{Pri}$ sequences are now known with authority and essential certainty. The pre-life tRNA sequence was completely ordered as RNA repeats and RNA inverted repeats (stem-loop-stems). We posit that ACCA-3' was the pre-life adapter to attach an amino acid (i.e., ACCA-Gly). A major purpose of Polymer World was to

synthesize polyglycine as a component of protocells. Minihelix World evolved as an improved means to synthesize polyglycine. Initially, tRNA evolved as a further improvement to synthesize polyglycine. To chemically evolve LUCA, tRNA World evolved from Polyglycine World and taught itself to encode amino acid sequence-dependent proteins.

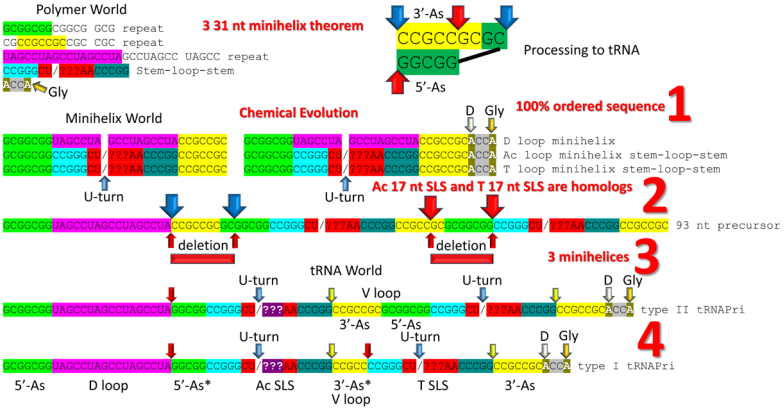

Figure 1-6. Evolution of tRNA from Polymer World to Minihelix World to tRNA World. TRNA sequences were conserved from pre-life and LUCA to present day. The figure represents chemical evolution during pre-life. A genetic code must chemically evolve before evolution of LUCA. Four points of emphasis are indicated (see text). The inset relates a mechanism for processing the tRNA 93 nt precursor at the 5' deletion position (type II and type I tRNAs) (large blue arrows) and the 3' deletion position (type I tRNA only) (large red arrows). Abbreviations: As) acceptor stem; As*) acceptor stem fragment; Ac) anticodon; SLS) stem-loop-stem.

The genetic code evolved around the tRNA anticodon. The number of amino acids that could be encoded was limited by wobbling at the tRNA-34 position. We posit that wobbling of tRNA-36 initially limited the genetic code to 8 amino acids. Suppression of wobbling at tRNA-36 allowed expansion of the genetic code to a maximum complexity of 32 assignments. The genetic code, however, froze at 20 amino acids plus stops because of translational fidelity restraints.

Four points about tRNA evolution must be emphasized. First, except for the ACCA-Gly adapter, tRNA evolved from completely ordered sequences: RNA repeats and RNA inverted repeats (stem-loop-stems). Second, the 17 nt anticodon and 17 nt T stem-loop-stems are homologous (genetically related). Because of homology, the pre-life loop sequence is

no longer known with certainty. Fairly clearly, anticodon and T sequences were initially identical. After ~4 billion years of Darwinian selection in the tRNA molecule, anticodon and T stem-loop-stem sequences diverged slightly but remain clear homologs. Third, tRNA evolved chemically from ligation of three 31 nt minihelices. Others have advocated for 2 minihelix models; however, no 2 minihelix model can possibly be adequate (Di Giulio 2019, 2012, 2009, Nagaswamy and Fox 2003, Lei and Burton 2023). Fourth, the three 31 nt minihelix theorem fully describes type I and type II tRNA evolution to the last nucleotide (Lei and Burton 2023). No competing model can account for both type I and type II tRNA evolution. Advocates for competing tRNA evolution models make fundamental errors in logic and in their analyses that they have failed to recognize (Lei and Burton 2023).

As noted, the genetic code first evolved to synthesize polyglycine (Polyglycine World) (Figure 1-6). Subsequently, the genetic code expanded to synthesize polymers of glycine, alanine, aspartic acid and valine (GADV World) (Ikehara 2023, 2016, 2014, 2009, Oba et al. 2005, Ikehara 2005). We do not believe that polyglycine or GADV synthesis was particularly accurate, so many available amino acids could have been incorporated into early polypeptides. Starting from synthesis of nearly random (i.e., GADV) polymers, a primitive translation system evolved to encode proteins specified by mRNA and tRNA sequences. On planet Earth, chemical evolution "learned" to encode proteins. Despite loud claims by some others, there is no chicken and egg problem in chemical evolution of the genetic code (Lei and Burton 2023). By trial and error, a system lacking foresight taught itself to synthesize polyglycine, then to synthesize more complex amino acid polymers and then to encode sequence-dependent proteins.

Once sequence-dependent proteins evolved, RNA ligation including diverse RNA repeats was translated, generating proteins with repetitive motifs and sequences. Repetitive motifs in proteins folded into barrels that have structural closure and solubility. TRNA evolved by RNA ligations (Figures 1-1 to 1-6). We posit that a large collection of important proteins that trace to LUCA evolved by RNA ligations followed by protein repeats and pseudosymmetrical folding to form barrels (Weiss et al. 2018). Protein barrels translated from RNA repeats can fold to pseudosymmetrical forms (i.e., $(\beta-\alpha)_8$ barrels and cradle loop barrels (i.e., double-Ψ−β-barrels)). Subsequently, $(\beta-\alpha)_8$ barrels rearranged by protein refolding and breaking of pseudosymmetry to form linear $(\beta-\alpha)_8$ sheets (Rossmann folds). We posit that RNA repeats, protein repeats and protein pseudosymmetry were

core contributions to rapid chemical evolution during pre-life (Lei and Burton 2023).

Figure 1-7 shows a pseudosymmetrical $(\beta–\alpha)_8$ barrel, triose phosphate isomerase from chicken. The $(\beta–\alpha)_8$ barrel motif is as old as LUCA (Weiss et al. 2018, Weiss et al. 2016). We posit that the $(\beta–\alpha)_8$ barrel motif evolved from two duplications of a $(\beta–\alpha)_2$ motif (Lei and Burton 2023). A β-sheet requires a β-sheet partner to maintain β-sheet extended chain geometry. Three duplications of a $(\beta–\alpha)$ motif, therefore, is perhaps less likely for $(\beta–\alpha)_8$ barrel evolution, because a $(\beta–\alpha)$ unit would have to dimerize to maintain its β-sheet. We posit that $(\beta–\alpha)_8$ barrel proteins evolved from translation of RNA repeats and pseudosymmetrical folding to gain closure as a barrel and to attain solubility. Essentially, all glycolytic (sugar breakdown) enzymes are $(\beta–\alpha)_8$ barrel proteins.

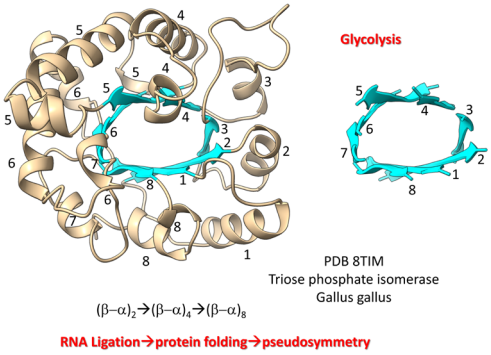

Figure 1-7. In pre-life, $(\beta–\alpha)_8$ barrels may have been generated by translation of RNA repeats and pseudosymmetrical protein folding (Lei and Burton 2023).

We posit that Rossmann fold proteins, which are $(\beta–\alpha)_8$ linear sheets, were refolded from $(\beta–\alpha)_8$ barrel proteins. The triboxylic acid (TCA) cycle (3-carbon sugar oxidation) utilizes many Rossmann fold proteins. Figure 1-8 shows a $(\beta–\alpha)_8$ linear sheet protein. β7 was lost during refolding because it lost contact with its partner β-sheets (β6 and β8 in a $(\beta–\alpha)_8$ barrel).