 doi:10.20944/preprints202209.0189.v1

Article

Realignment of tRNA Variable Loops, the 3-31 nt Minihelix Theorem and Consequences for the Origin of Life

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

[1] School of Biological Sciences, University of New England, Biddeford, ME, USA 04005; llei@une.edu
[2] Department of Biochemistry and Molecular Biology, Michigan State University, E. Lansing, MI, USA 48824-1319; burton@msu.edu
* Correspondence: burton@msu.edu; 1-517-881-2243

Simple Summary: TRNA sequences are highly conserved from their primordial root ~4 billion years ago. The 3-31 nucleotide (nt) minihelix transfer RNA (tRNA) evolution theorem describes the main path for the chemical evolution of life. Significantly, once tRNA evolved in pre-life, the genetic code evolved, leading to the evolution of the first cells and organisms. The theorem describes the relationship comparing type I and type II tRNA variable loops and requires a realignment of variable loop sequences. Notably, tRNA sequence maintains a historical record of chemical evolution through the pre-life to life transition.

Abstract: Evolutionary, sequence and structural analyses of tRNA variable (V) loops provides a new understanding. Type I tRNA V loops have a primordial length of 5 nt, and type II tRNA V loops have a primordial length of 14 nt. Sequence-based alignments of type I and type II V loops gave deceptive results. Type II V loops are characterized by the trajectory of the V arm, the size and sequence of the loop and the first and last V loop nucleotides and their known contacts. Depending on the V arm trajectory, type II V stems and loops can be interaction sites for aminoacyl-tRNA synthetases. So, type II V arms for $tRNA^{Leu}$, $tRNA^{Ser}$ and $tRNA^{Tyr}$ (found in Bacteria) were coevolved and selected to distinguish determinants to support cognate tRNA charging. The alignment of type I and type II V loops was adjusted based on tRNA evolution. Results are consistent with Archaea being an older and simpler life form than Bacteria. TRNAome sequences were derived from radiation of an ordered $tRNA^{Pri}$ (Pri for primordial) sequence comprised of known RNA repeats and inverted repeats. V loop and tRNA sequences, therefore, convey a history of the primary successful pathway in the origin of life.

Keywords: aminoacyl-tRNA synthetase; chemical evolution; origin of life; primordial tRNA; RNA inverted repeats; RNA repeats; tRNA; tRNAome; variable loop.

1. Introduction

The 3-31 nt minihelix theorem for the evolution of tRNA made predictions about type I and type II tRNA variable (V) loops [1-5]. The primordial length of a type I V loop is 5 nt, and the root sequence was CCGCC. The primordial length of a type II V loop is 14 nt, and the root sequence was CCGCCGCGCGGCGG (Figure 1). Because the 3-31 nt minihelix model explains all features, sequences and segments of type I and type II tRNAs, the model has been described as a proven theorem [3]. Competing models for tRNA evolution are all accretion models [6-14], in which tRNA segments expand and/or contract by random insertions and/or deletions (indels) to a final form. Because the entire primordial tRNA ($tRNA^{Pri}$) sequence is ordered as discrete segments of RNA repeats and inverted repeats, however, no accretion model with random indels can adequately describe tRNA evolution [3]. The 3-31 nt minihelix theorem, by contrast, is clearly correct in every detail and is strongly supported by statistical evaluations of sequences in living organisms

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2022 doi:10.20944/preprints202209.0189.v1

[5,15]. It was initially surprising that such a strong sequence record of tRNA^{Pri} was maintained under selection for ~4 billion years, but tRNA is the molecule around which translation systems, the genetic code and life evolved. So, hysteresis in tRNA sequence divergence might be expected in ancient organisms. The 3-31 nt minihelix model fully describes evolution of both type I and type II tRNAs. No competing model, no accretion model (with random indels) and no 2 minihelix model can adequately describe early tRNA evolution. The 3-31 nt minihelix model accurately predicted how type II tRNAs evolved and how V loops relate in sequence to tRNA acceptor stems [5]. The 3-31 nt minihelix model explained the origins of type I and type II tRNAs, and analyses of the shorter and longer V loops and their sequences helped to reinforce and confirm the model.

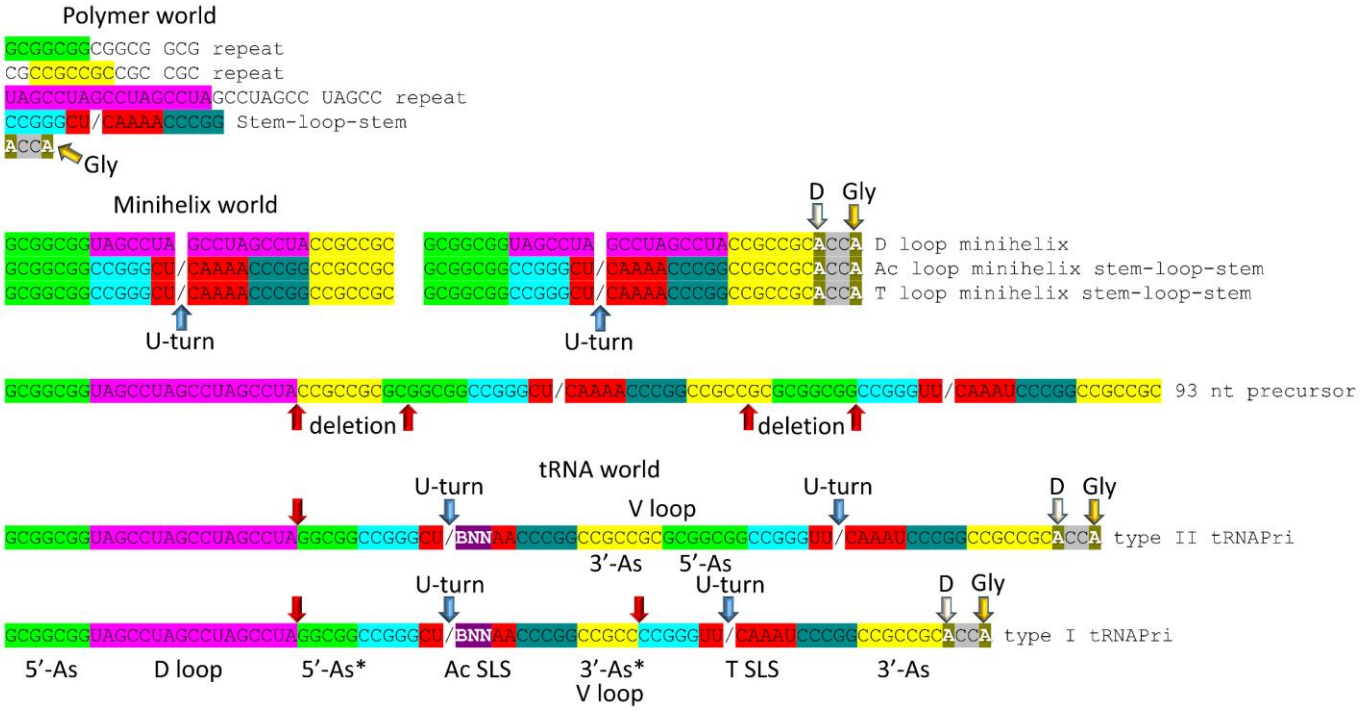


Figure 1. The 3-31 nt minihelix theorem for tRNA evolution. Polymer world evolved to minihelix world that evolved to tRNA world including type I and type II tRNAs that vary in the length of the V loop. Colors: green) 5′-acceptor stems (As) and 5′-acceptor stem remnants (5′-As*); magenta) D loop minihelix 17 nt core; cyan) anticodon (Ac) and T stem-loop-stem (SLS; 17 nt) 5′-stems; red) Ac and T 7 nt loops; purple) the anticodon; cornflower blue) Ac and T 3′-stems; yellow) 3′-As and 3′-As* fragments; gold) discriminator base and aminoacylated 3′ base; gray) 3′-CC for attachment to the ribosome acceptor site and peptidyl transfer center of the ribosome. Arrow colors: red) 9 nt deletion sites; blue) U-turns; light yellow) discriminator base; gold) amino acid attachment. See also Figures 2 and 3. B indicates G, C or U but not A. Wobble A is not utilized in Archaea. N indicates any base (A, G, C or U).

The 3-31 nt minihelix model is shown in Figure 1. The model indicates that tRNA world, which quickly advanced to the genetic code and life, was preceded by an ancient polymer world and, then, 31 nt minihelix world. Because polymer world, the 31 nt minihelix world and the inception of tRNA world are pre-life processes, the 3-31 nt minihelix model describes a primary successful pathway of chemical evolution toward life. A small number (i.e., probably fewer than 20) of ribozymes/primitive catalysts were necessary to convert polymer world to the 31 nt minihelix world [1,2]. TRNA world, in which organisms have lived for about 4 billion years, was a natural consequence of the complementary replication and processing of 31 nt minihelices. Minihelices were ligated (by a ribozyme/primitive catalyst), generating polymers that ended in a snap-back primer for com-

 doi:10.20944/preprints202209.0189.v1

plementary replication (by a ribozyme/primitive catalyst). Any 31 nt minihelix can function as a snap-back primer because of the complementarity of the 5′- and 3′-acceptor stems (Figure 1). The 93 nt tRNA precursor was generated via ligation and then processed by primitive catalysts by two 9 nt deletions or one 9 nt deletion to type I and type II $tRNA^{Pri}$ molecules. It is worth noting that the more 5′ and 3′ 9 nt deletions that generated type I tRNAs were the same processing event on complementary RNA strands. Therefore, only a single deletion mechanism was necessary to process to type I and type II tRNAs, assuming complementary 5′→3′ replication. Considering the similarity of the 9 nt deletions, the need for snap-back primers and the existence of inverted repeats, templated 5′→3′ RNA replication must have been a feature of the pre-life world. TRNAomes (all of the tRNAs of an organism), the genetic code and aminoacyl-tRNA synthetase (AARS) enzymes co-evolved to produce living organisms. Polymer world, 31 nt minihelix world and tRNA world were selected as covalent RNA-amino acid systems for forming RNA-dependent polypeptide chains. Initially, polyglycine was a major product of these systems, although pre-life polypeptide synthesis may not have been highly selective for glycine. Glycine is the simplest amino acid and was probably the most abundant in the pre-life world.

An attempt at a similar study of the evolution of type II V loops was previously published [16]. That study, however, was done without knowledge of root type I and type II tRNA sequences.

2. Materials and Methods

TRNA sequences were collected from the tRNA Database (http://trnadb.bioinf.uni-leipzig.de/) [17]. Typical tRNAs are displayed as tDNA sequences. Archaeal type II V loop sequences were realigned and annotated manually (Supplementary Files #1 and #2). Structures were visualized using UCSF ChimeraX [18]. Sequence logos were generated using WebLogo 3.7.4 (https://weblogo.threeplusone.com/) [19,20]. To confirm the 3-31 nt minihelix theorem requires little effort using the tRNA Database and online tools.

3. Results

Figure 1 shows a model for chemical evolution of tRNA from an ancient polymer world to a 31 nt minihelix world to the current tRNA world sustained on Earth for about 4 billion years. The model adjusts the alignment of type I and type II tRNA V loops, compared to alignments shown in online databases. The model shows that a primordial tRNA ($tRNA^{Pri}$) sequence can be identified with high confidence. The entire sequence of $tRNA^{Pri}$ for type I and type II tRNAs was comprised of repeats and inverted repeats of known lengths and mostly known sequence. According to the theorem, the only sequence ambiguities are within the anticodon and T 7 nt U-turn loops (Figure 1; red and purple segments). Sequence ambiguities arise because, within tRNA, the homologous anticodon loop and T loop sequences were separately selected in evolution for ~4 billion years. All other $tRNA^{Pri}$ sequences are known with very high confidence (essential certainty). When this model was first considered, it was a surprise to its originators, because tRNA was previously thought to arise by a chaotic process [6,8-14,21]. Because the entire $tRNA^{Pri}$ and entire tRNAomes from ancient organisms are ordered, however, no chaotic model can possibly be correct for tRNA evolution [1-3]. Only the 3-31 nt minihelix theorem can possibly account for $tRNA^{Pri}$ evolution and the evolution of tRNAomes of ancient organisms.

The 3-31 nt minihelix theorem for evolution of V loops requires that the alignment of type I and type II tRNAs be reconsidered. Sequence-based alignments of type I and type II tRNAs were biased by V loop residues that were strongly selected to make specific contacts. The V loop 5′ and 3′ ends, for instances, were selected by interacting bases. The 5′ variable loop base (V_1) was selected according to its interaction with the tRNA-26 base. The 3′ variable loop base (V_n for a V loop of n bases) forms the Levitt base pair with the base at the tRNA-15 position. So, in Archaea, V_nC forms a Levitt reverse Watson-Crick base pair (two hydrogen bonds) with tRNA-G15 modified to archaeosine [22]. In Archaea,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2022 doi:10.20944/preprints202209.0189.v1

the tRNA^{Leu} type II variable loop was selected to include an AG sequence to directly contact the LeuRS-IA C-terminal domain (Supplementary File #1). We propose to re-align type I and type II V loops as indicated in the evolutionary model shown in Figure 1. So, generally, type I V loops would align to the first 5 nt in type II V loops. Because V loops undergo sequence expansions and/or contractions in evolution, potential indels must be considered in aligning some sequences.

Figure 2 shows the 3-31 nt minihelix theorem applied to type I tRNAs. Figure 2A shows a type I tRNA^{Phe} from Saccharomyces cerevisiae [23] colored according to the model. Figures 2B and 2C show tRNA^{Gly}, the oldest tRNA and the one most similar in sequence to tRNA^{Pri} [24]. Figure 2D shows the alignment of tRNA^{Pri} (type I), tRNA^{Typical} and tRNA^{Gly} (typical; consensus from 9 tRNA^{Gly}). Very clearly, these sequences are nearly identical, indicating: 1) that the 3-31 nt minihelix model is accurate and correct; 2) that alternate tRNA evolution models are falsified; and 3) that the model can be used to align type I and type II tRNAs in the V loop region.

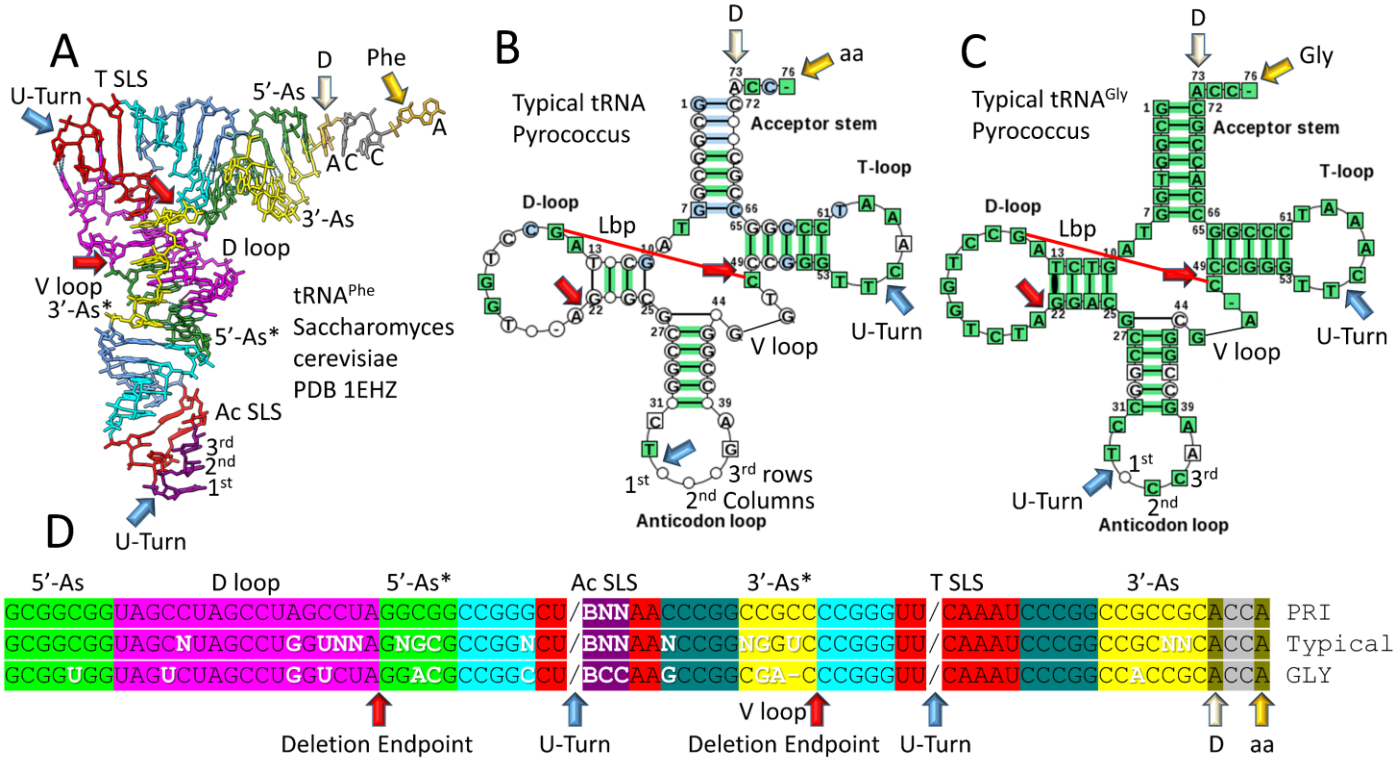


Figure 2. Type I tRNA. TRNA^{Pri} and tRNA^{Gly} radiated to entire tRNAomes. Consistent with a polyglycine world, tRNA^{Gly} was the first tRNA. A) TRNA^{Phe} from Saccharomyces cerevisiae colored according to the theorem. B) A typical tRNA from 3 Pyrococcus species. Note that the 5'-acceptor stem is a perfect GCG repeat. Note the obvious homology of the 17 nt Ac SLS and the 17 nt T SLS. TDNA sequence is shown. The red line connecting 15G and 48C (V_5C) is the Levitt reverse Watson-Crick base pair. C) A typical tRNA^{Gly} from 3 Pyrococcus species. Note that, after ~4 billion years of evolution, the tRNA^{Gly} 17 nt D loop sequence (magenta) matches the theorem in all but 3 positions. Lbp) Levitt base pair. D) An alignment of type I tRNA^{Pri}, tRNA^{typical} and tRNA^{Gly} colored according to the theorem. Colors and colored arrows are as described in Figure 1.

Figure 3 shows the 3-31 nt minihelix model applied to type II tRNAs [5]. Figure 3A shows a type II tRNA^{Leu} from Pyrococcus horikoshii [25] colored according to the model. The variable loop was derived from a 3'-acceptor stem (7 nt yellow segment) ligated to a 5'-acceptor stem (7 nt green segment). Figures 3B and 3C show typical type II tRNA diagrams for tRNA^{Leu} and tRNA^{Ser} from Pyrococcus furiosis. In Archaea, tRNA^{Leu} and tRNA^{Ser} are type II tRNAs. We argue that the primordial length of type II tRNA V loops

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2022 doi:10.20944/preprints202209.0189.v1

was 14 nt, as in P. furiosis tRNA^{Leu} and as shown in the model (Figure 1; Supplementary File #1). In P. furiosis, the tRNA^{Ser} V loop is 15 nt, but other achaeal species have tRNA^{Ser} V loops of 14 nt, the primordial length (Supplementary File #2) [17]. We conclude that the P. furiosis 15 nt tRNA^{Ser} V loop length represents a 1 nt insert expansion from the primordial length. Figure 3D shows alignment of type II tRNA^{Pri}, tRNA^{Leu} (typical, from P. furiosis) and tRNA^{Ser} (typical, from P. furiosis). These sequences are nearly identical. We posit that archaeal tRNA^{Leu} and tRNA^{Ser} coevolved and were counterselected for cognate tRNA charging from type II tRNA^{Pri}.

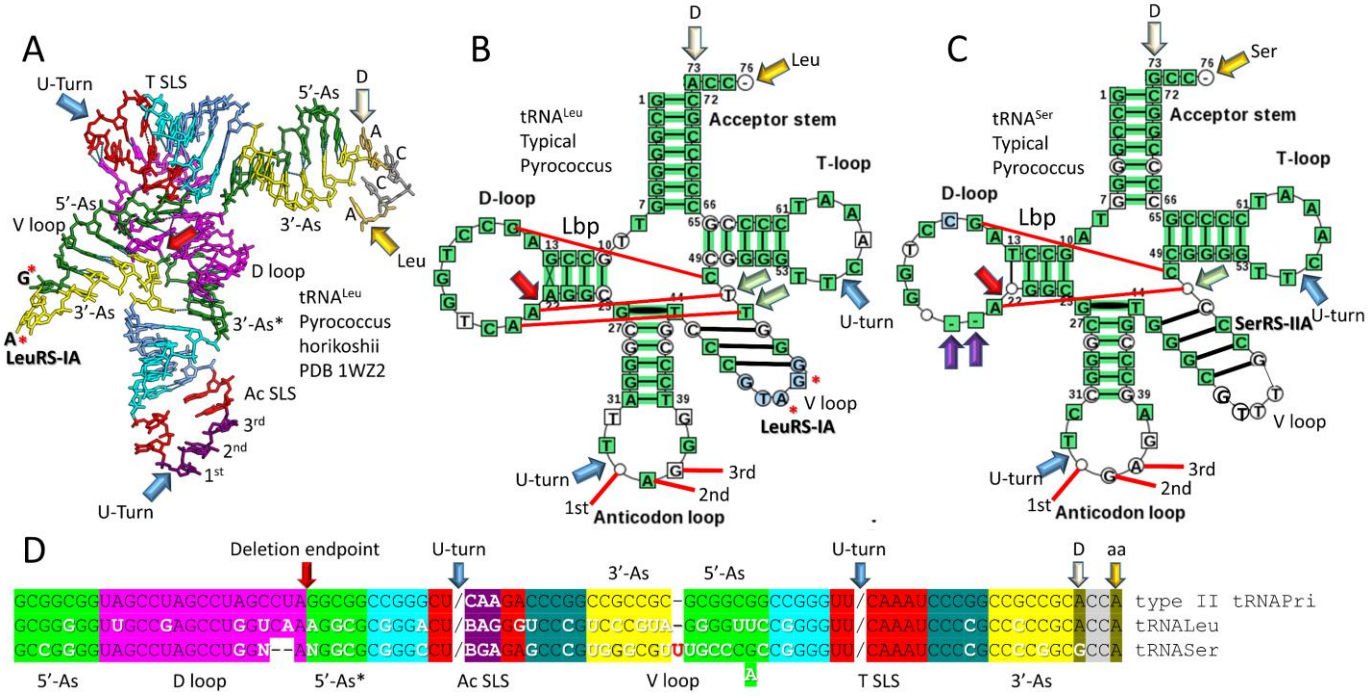


Figure 3. Type II tRNA. TRNA^{Leu} and tRNA^{Ser} have type II V loops with different trajectories from the tRNA body that were selected for accurate aminoacylation by LeuRS-IA and SerRS-IIA. A) X-ray structure of tRNA^{Leu} from Pyrococcus horikoshii. LeuRS-IA contacts the sequence AG on the V loop (red asterisks). B) A typical tRNA^{Leu} from Pyrococcus furiosis. C) A typical tRNA^{Ser} from P. furiosis. Light green arrows indicate unpaired bases between the 3′-V stem and the Levitt base (V_n) (Lbp). Purple arrows indicate deleted bases in the D loop of tRNA^{Ser}. D) An alignment of type II tRNA^{Pri} with tRNA^{Leu} and tRNA^{Ser} from P. furiosis. Shading colors and arrow colors are consistent in Figures 1-3. Note that tRNA^{Ser} includes two perfect UAGCC repeats in its D loop sequence. Continuing with the D loop sequence in tRNA^{Leu} gives the sequence UGGUCAA, which is a close match to the tRNA^{Pri} sequence UAGCCUA from the theorem.

Type II variable loops evolved from a sequence that can pair along its entire length (CCGCCGCGCGGCGG) to a more favorable sequence that forms a short stem of variable trajectory, a loop and may also have an unpaired base or unpaired bases just 5′ of the 3′-V_n (V loop of n nts; counting from the V_1 to the V_n base) base that forms the Levitt base pair. P. furiosis tRNA^{Leu} has 2 unpaired bases in this interval (Figure 2B; light green arrows). P. furiosis tRNA^{Ser} has 1 unpaired base. So far as we know, although the lengths of type II V loops may vary in evolution, these numbers of unpaired tRNA^{Leu} and tRNA^{Ser} V loop bases are maintained throughout Archaea (Supplementary Files #1 and #2). The number of unpaired bases just prior to the Levitt base pair determines the trajectory of the V arm from the tRNA (Figure 4). With two unpaired bases, the V arm angles to the left, in the view shown. With one unpaired base, the V arm projects straight out toward the viewer. With no unpaired bases, the V arm angles sharply to the right. In Archaea, tRNA^{Leu} has 2 unpaired bases and tRNA^{Ser} has 1 unpaired base just 5′ of the Levitt base. In Bacteria, tRNA^{Tyr} has 2 unpaired bases, tRNA^{Leu} has 1 unpaired base and tRNA^{Ser} has 0

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2022 doi:10.20944/preprints202209.0189.v1

unpaired bases. It appears that, in prokaryotes, cognate type II tRNAs were evolved to have distinguishable V loop trajectories with different numbers of unpaired bases just 5′ of the Levitt V_n base.

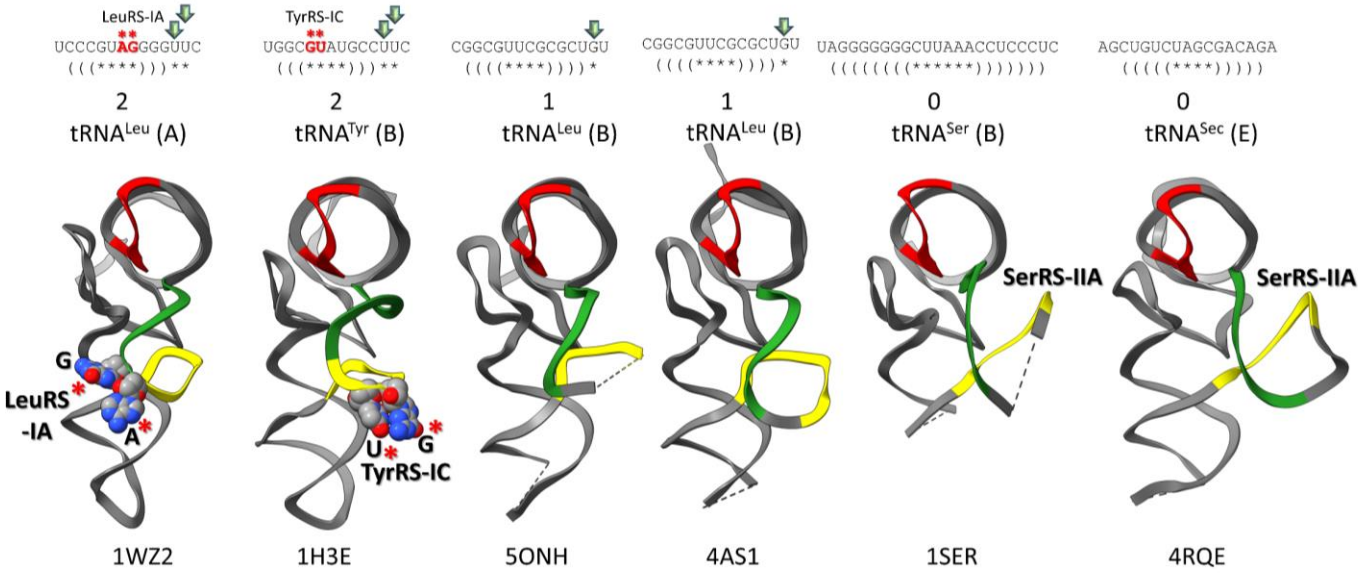


Figure 4. The number of unpaired V loop bases separating the 3′-V stem and the Levitt base (V_n; light green arrows) determines the trajectory of the V stem from the tRNA. A, B and E represent Archaea, Bacteria and Eukarya. V loop bases that are contacted by LeuRS-IA and TyrRS-IC are indicated by red asterisks and space-filling representations in the structures. TRNA is gray. The T loop is red. Yellow indicates the segment of the V loop derived from the 3′-acceptor stem. Green indicates the segment of the V loop derived from the 5′-acceptor stem. Some relevant structures are not available at the time of writing.

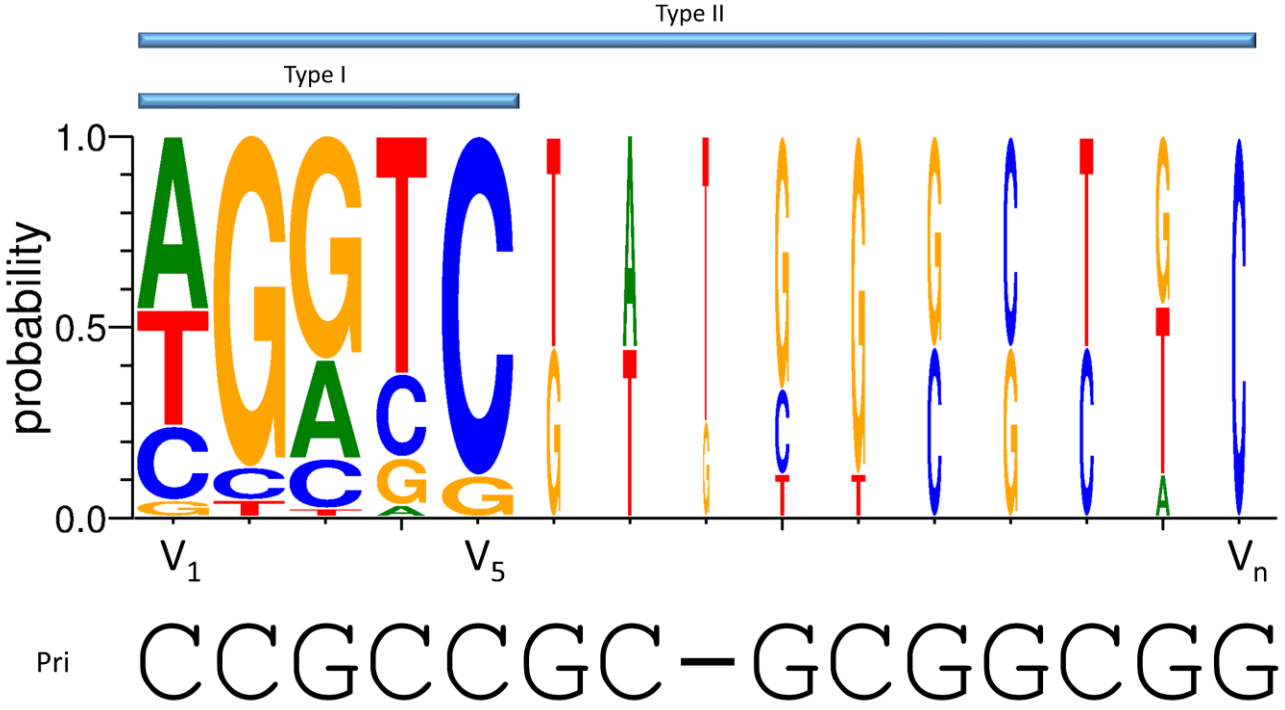


Figure 5 shows a corrected alignment for Pyrococcus furiosis type I and type II V loops adjusted according to the 3-31 nt minihelix theorem. As in Figure 1, type I tRNAs align within the first 5 bases of type II tRNAs. A sequence logo is shown for the 46 tRNAs in P. furiosis. The primordial V loop sequence is shown for reference. The primordial V loop

 doi:10.20944/preprints202209.0189.v1

sequence appears to match to P. furiosis V loop sequences in all but 3 positions, all of which are specific to type II V loops that diverged for LeuRS-IA and SerRS-IIA discrimination.

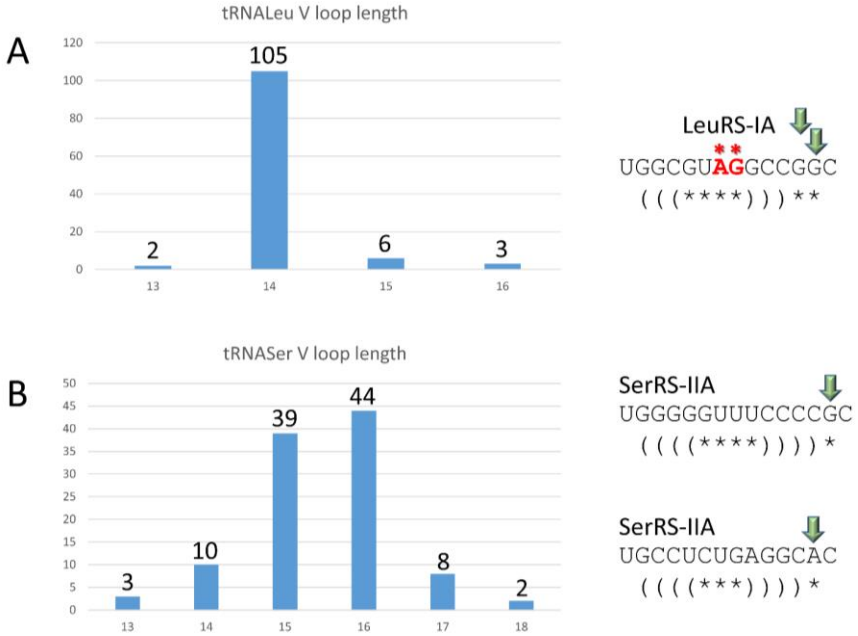


Figure 6. Lengths of type II V loops in Archaea. A) The tRNA^{Leu} V loop is primarily 14 nt, which is the primordial length (Figure 1). B) The tRNA^{Ser} V loop varies from 13-18 nt and is primarily 15 or 16 nt. Characteristic V loop sequences are shown to the right of the graphs (see also Supplementary Files #1 and #2). Light green arrows indicate unpaired bases just 5′ to the Levitt base V_n.

The trajectory of the V arm has consequences for LeuRS-IA contacts in Archaea and Bacteria (Figure 7). The type II V loop of tRNA^{Leu} from Pyrococcus horikoshii has 2 unpaired bases between the 3′-V stem and V_{14}C that forms the Levitt base pair to 15G (see Figures 3B and 4) [25]. V_7A and V_8G contact the C-terminal domain of P. horikoshii LeuRS-IA (Figure 7A). In Bacteria, by contrast, the V arm trajectory is changed. In Bacteria, tRNA^{Leu} has only a single unpaired base separating the 3′-V stem and the base that forms the Levitt base pair (see Figure 4). In Bacteria, the tRNA^{Leu} V loop makes no apparent contacts to LeuRS-IA (Figure 7B) [26]. Instead the C-terminal β-hairpin that contacts the V loop in Archaea contacts the T loop in Bacteria (compare Figures 7A and 7B).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2022 doi:10.20944/preprints202209.0189.v1

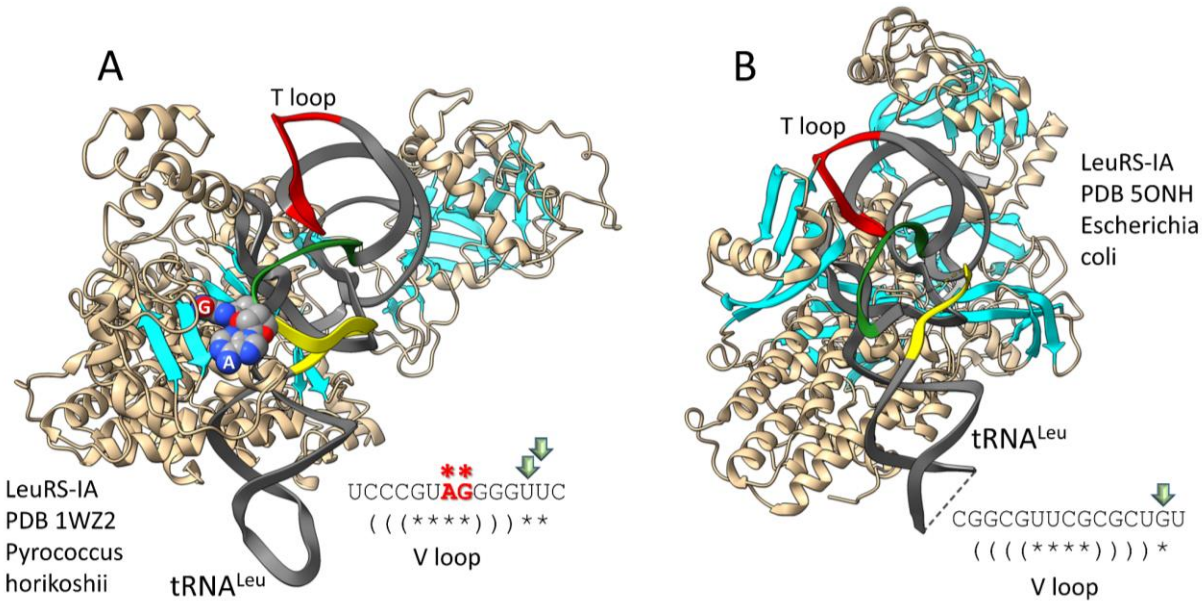
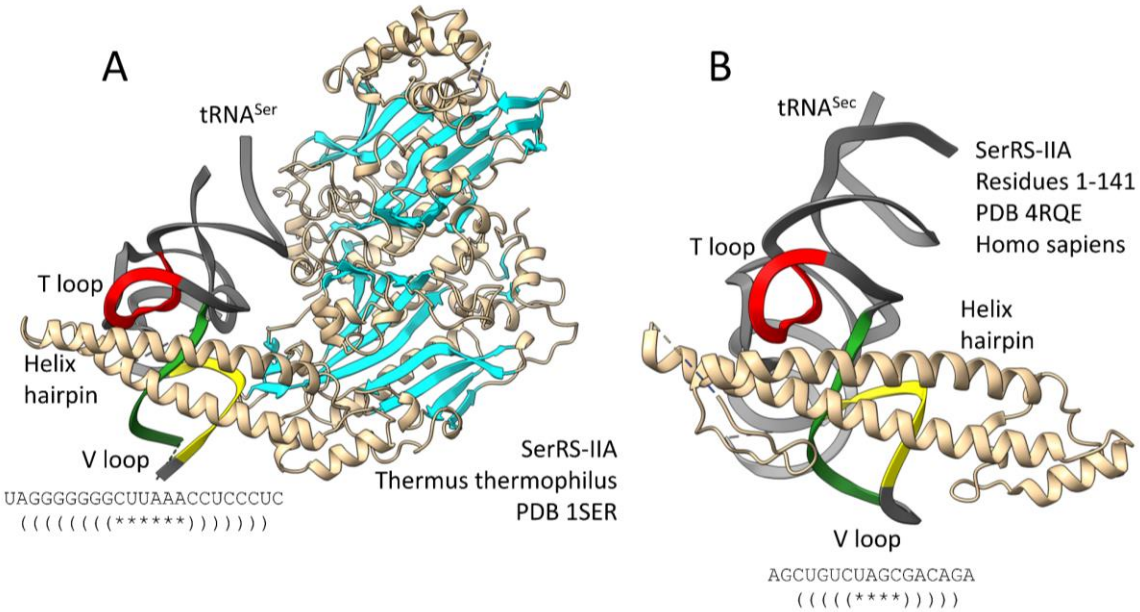


Figure 7. Different V loop determinants in LeuRS-IA recognition of $tRNA^{Leu}$ in Archaea and Bacteria. A) In Archaea, the C-terminal β-sheet contacts the AG motif in the $tRNA^{Leu}$ V loop. B) By contrast, in Bacteria, the C-terminal β-sheet contacts the T loop in $tRNA^{Leu}$. Protein is beige and cyan (β-sheets). TRNA is gray. The V loop is yellow (derived from the 3′-acceptor stem) and green (derived from the 5′-acceptor stem). The T loop is red. V loop sequences are annotated as in Figure 4.

SerRS-IIA contacts the V stem of $tRNA^{Ser}$ and $tRNA^{Sec}$ (Sec for selenocysteine), but not the V loop, through its N-terminal helix hairpin (Figure 8). The SerRS-IIA N-terminal helix hairpin forms a brace between the V stem and the T loop of the $tRNA^{Ser}$ (Figure 8A) [27] and $tRNA^{Sec}$ (Figure 8B) [28]. Unfortunately, no structure is currently available for Pyrococcus SerRS-IIA with a bound $tRNA^{Ser}$. Our interest in an archaeal structure is that it will have a different V arm trajectory than in Bacteria (see Figures 3C and 4). The V loop of $tRNA^{Ser}$ in Archaea has 1 unpaired base separating the 3′-V stem from the base that participates in the Levitt base pair (Supplementary File #2). In Bacteria, by contrast, the type II V loops of $tRNA^{Ser}$ have 0 unpaired bases. SerRS-IIA charges $tRNA^{Sec}$ with serine, which is then modified to selenocysteine.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2022 doi:10.20944/preprints202209.0189.v1

Figure 8. SerRS-IIA contacts the V stem and the T loop with an N-terminal helix hairpin. A) SerRS-IIA bound to $tRNA^{Ser}$ from Thermus thermophilus. B) SerRS-IIA (residues 1-141) bound to $tRNA^{Sec}$ from Homo sapiens. Colors and annotations as in Figure 7.

In Bacteria, $tRNA^{Leu}$ has 1 unpaired base separating the 3′-V stem from the Levitt base (V_n) versus 2 unpaired bases in $tRNA^{Leu}$ in Archaea (Figure 7). In Archaea, $tRNA^{Tyr}$ is a type I tRNA. In Bacteria, by contrast, $tRNA^{Tyr}$ is a type II tRNA [17]. Thermus thermophilus TyrRS-IC contacts V_5G and V_6U for accurate charging of $tRNA^{Tyr}$, which has a 14 nt V loop, the primordial length (Figure 9A). V loop binding is through the C-terminal TyrRS-IC domain. Figure 9B shows a sequence logo of bacterial $tRNA^{Tyr}$ 14 nt V loops. High conservation of V_5G or U and V_6U in the logo indicates that, at least in most species represented, TyrRS-IC must make similar V loop contacts.

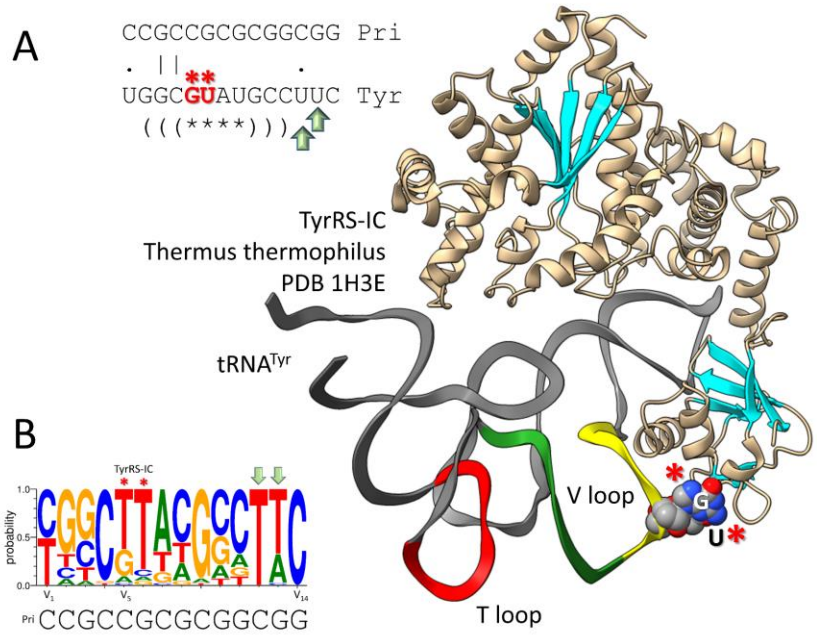


Figure 9. In Thermus thermophilus, TyrRS-IC contacts the $tRNA^{Tyr}$ V loop through a C-terminal domain. A) T. thermophilus TyrRS-IC bound to the $tRNA^{Tyr}$ V loop. B) A sequence logo of 14 nt $tRNA^{Tyr}$ V loops in Bacteria. Red asterisks indicate TyrRS-IC V loop contacts.

Shortening of the $tRNA^{Tyr}$ V loop, however, appears to have consequences for TyrRS-IC recognition of the V loop. In Escherichia coli and many other Bacteria, the $tRNA^{Tyr}$ V loop was shortened to 12 or 13 nt, and, from inspection of sequences, TyrRS-IC contacts to these shorter V loops appear not to be utilized [17]. The shortening of the V loop may partly correlate with the evolution of the wobble queousine modification of $tRNA^{Tyr}$, which may in part create an improved determinant to replace the V loop contact that appears to be lost in some bacterial species.

It appears that class I AARS enzymes have not evolved to access a V loop with 1 or 0 unpaired bases separating the 3′-V stem and the Levitt base (V_n) (see Figure 4). LeuRS-IA from Bacteria, for instance, has lost all contact with the $tRNA^{Leu}$ V loop (Figure 7B), which has 1 unpaired base in this interval. Because, in Bacteria, $tRNA^{Tyr}$ evolved with a V loop with 2 unpaired bases just 5′ of the Levitt base, $tRNA^{Leu}$ in Bacteria appears to have evolved toward a new mode of recognition (compare Figures 7A and 7B). In prokaryotes, type II tRNAs encoding different amino acids appear to not have the same trajectory from the tRNA without compromising tRNA discrimination (Figure 4).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2022 doi:10.20944/preprints202209.0189.v1

Thermus thermophilus is an ancient Bacterium, but its type II V loops are more innovated than archaeal type II V loops and more diverged from the primordial length of 14 nt (Figures 10-12). Escherichia coli is a much more specialized Bacterium than T. thermophilus, and E. coli has more innovated type II V loops than T. thermophilus (not shown) [17]. Figure 10 shows a typical tRNA^{Leu} from T. thermophilus and its V loop sequences. As noted, the bacterial tRNA^{Leu} has 1 unpaired base separating the 3′-V stem from the base that forms the Levitt base pair, which in this case is V$_n$U. V$_n$U forms a Levitt reverse Watson-Crick base pair with tRNA 15A. The length of the tRNA^{Leu} V loop in T. thermophilus varies (13-17 nt). Figure 11 shows a typical tRNA^{Ser} from T. thermophilus and its V loop sequences. In Bacteria, tRNA^{Ser} has 0 bases separating the 3′-V stem from the V$_n$ base that forms the Levitt base pair, which in this case is V$_n$C that pairs with tRNA 15G, as in archaeal tRNA^{Leu} and tRNA^{Ser} (Supplementary Files #1 and #2). The length of the T. thermophilus tRNA^{Ser} V loop varies (19-22 nt). Type II tRNA^{Tyr} in T. thermophilus has 2 bases separating the 3′-V stem from the base that forms the Levitt base pair (Figure 9), which is V$_n$C, as in archaeal tRNA^{Leu} and tRNA^{Ser}.

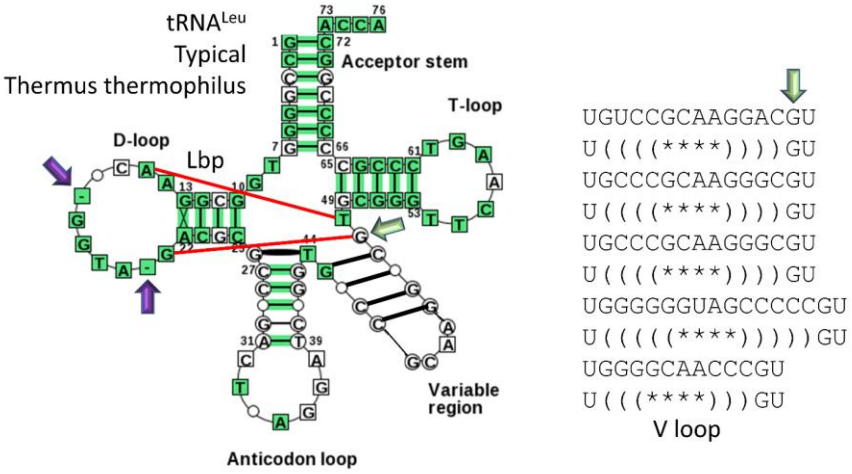


Figure 10. A typical tRNA^{Leu} of Thermus thermophilus. V loop lengths range from 13-17 nt. The Levitt base pair is 15A-V$_n$U. Purple arrows indicate D loop deletions that are not observed in Pyrococcus furiosis tRNA^{Leu} (Figure 3B) but may be seen in other Archaea (Supplementary File #1). Other annotations are as in previous figures. The website program does a poor job of aligning V loops of different lengths [17].

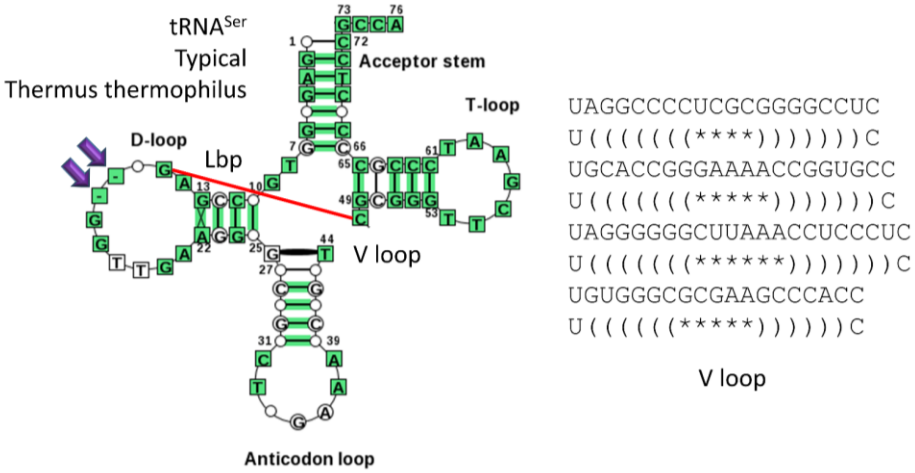


Figure 11. A typical tRNA^{Ser} from Thermus thermophilus. V loop lengths range from 19-22 nt. Purple arrows indicate deleted bases in T. thermophilus tRNA^{Ser}. These D loop bases are also deleted in some archaeal tRNA^{Ser} (Supplementary File #2).

To summarize type II tRNA data from Archaea and Bacteria, Figure 12 is shown. It appears that, to aid discrimination of cognate tRNAs in prokaryotes, type II tRNAs

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2022 doi:10.20944/preprints202209.0189.v1

evolved to have distinct trajectories from the tRNA body. In Archaea, $tRNA^{Leu}$ has 2 unpaired bases and $tRNA^{Ser}$ has 1 unpaired base just 5′ to the Levitt base (V_n) (Supplementary Files #1 and #2). In Bacteria, $tRNA^{Tyr}$ has 2 unpaired bases, $tRNA^{Leu}$ has 1 unpaired base and $tRNA^{Ser}$ has 0 unpaired bases just 5′ to the Levitt base. In Archaea, $tRNA^{Tyr}$ is a type I tRNA that is closely related, in ancient species (i.e., Pyrococcus), to $tRNA^{Asn}$, which has a similar anticodon, indicating that $tRNA^{Tyr}$ and $tRNA^{Asn}$ may have been coevolved in ancient Archaea by duplication and repurposing [24].

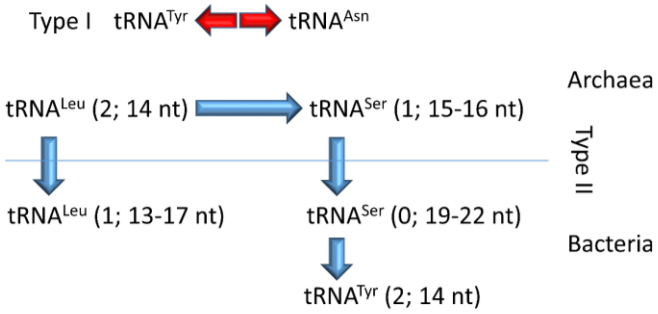


Figure 12. A model for evolution of type II tRNAs in Archaea and Bacteria. Archaea are considered to be the older Domain and more similar to LUCA compared to Bacteria. Divergence of Archaea and Bacteria is hypothesized to have occurred very early, perhaps when all type II tRNA V loops were still 14 nt in length. Dominant lengths of archaeal tRNA V loops are indicated compared to the lengths of T. thermophilus type II tRNAs. $TRNA^{Tyr}$ in Archaea is a type I tRNA that is closely related in sequence to $tRNA^{Asn}$ (red arrows).

4. Discussion

To our knowledge, type II V loops have not previously been analyzed and compared in detail in Archaea and Bacteria. We consider this comparison to be important for three reasons: 1) to analyze and understand the evolution of tRNA; 2) to describe the recognition of tRNAs by their cognate AARS enzymes; and 3) to describe the divergence of Bacteria from Archaea. We consider further analysis of type II tRNAs from Eukarya to be an interesting future project that we judged to be beyond the scope of the current work.

Our group has proposed the 3-31 nt minihelix theorem to describe evolution of type I and type II tRNAs [1-3]. We consider this to be a theorem and not merely a model because the evidence for the scheme from tRNA sequence analysis is so compelling that we cannot consider the possibility of the model being rationally questioned or falsified. Notably, the theorem provides four independent points of emphasis: 1) primordial tRNA sequences were composed of repeats and inverted repeats of known sequence (Figures 1-3); 2) evolution of tRNA required ligation of 3-31 nt minihelices of almost completely known sequence (Figure 1); 3) the theorem accounts for obvious homology of the 17 nt anticodon and 17 nt T stem-loop-stems (Figure 2B); and 4) the theorem completely describes the evolution of both type I and type II tRNAs (Figures 1-3). Notably, the longer type II V loop is not a V loop expansion, as some have described it, but, rather, a failure to process by an orderly 9 nt deletion (Figure 1).

It is notable to stress that the 3-31 nt minihelix theorem was initially proposed only for type I tRNAs [15,29]. In collaboration with Dr. Bruce Kowiatek and at his suggestion, we applied the theorem to type II tRNAs, as well [5]. The solution for type II tRNAs, therefore, was a major prediction of the 3-31 nt minihelix theorem and provided very strong support for the theorem as initially proposed. No alternate model for tRNA evolution can adequately describe either type I or type II tRNA evolution. As we have previously explained, no model that includes random indels (insertions and/or deletions) can possibly describe early tRNA evolution because of the clear patterning of $tRNA^{Pri}$ sequences (Figures 1-3). Furthermore, no 2 minihelix model can possibly describe tRNA early evolution. Only a 3-31 nt minihelix model is adequate. Thus, competing tRNA evolution models fail on multiple counts: 1) competing models fail to describe patterning (RNA repeats and inverted repeats) in $tRNA^{Pri}$ and tRNAomes (i.e., Pyrococcus; Figures 2-3); 2) those models

 doi:10.20944/preprints202209.0189.v1

fail to account for obvious homology of the 17 nt anticodon and 17 nt T stem-loop-stems (Figure 2B); 3) competing models require chaotic indels, which are in contradiction of the obvious tRNA sequence patterning (Figures 1-3); and 4) those models fail to account for both type I and type II tRNAs.

In Figure 12, a model is proposed for the radiation of type II V loops in Archaea and Bacteria. We describe archaeal type II V loops in detail (Supplementary Files #1 and #2). To describe bacterial type II V loops in similar detail would require additional analysis that we have not completed. For the current work, we show data for Thermus thermophilus, which we consider to be a reasonable reference organism as an ancient Bacterium that may be close to the divergence point for Archaea and Bacteria. Based on multiple analyses by us and by others, we consider Archaea to be older organisms than Bacteria and more similar to LUCA, particularly for translation systems. Our analysis of type II V loops appears to support this view (Figure 12).

Prokaryotes distribute the conformations of type II V loop arms for different cognate tRNAs. In Archaea, $tRNA^{Leu}$ and $tRNA^{Ser}$ have different trajectories of their arms from the tRNA body (Figures 3B-C, 4, 7A and 12). In Bacteria, $tRNA^{Tyr}$, $tRNA^{Leu}$ and $tRNA^{Ser}$ have different trajectories of their type II V arms from the tRNA body (Figures 4, 7B, 8-12). The trajectories of V arms can be related to two linked features: 1) the angles of hydrogen bonds in V stem pairing (Figures 3B-C); and 2) the number of unpaired bases (2, 1 or 0) separating the 3′ V stem and the Levitt base V_n (Figures 3B-C, 4, 7- 9 and 12).

Secondly, Bacteria appear to have replaced the type I $tRNA^{Tyr}$ from Archaea with a type II $tRNA^{Tyr}$ that may be derived from a primitive $tRNA^{Ser}$ [5]. If this idea is correct, duplication and repurposing of $tRNA^{Ser}$ into $tRNA^{Tyr}$ probably occurred when both tRNAs had V loops of 14 nt. In Thermus thermophilus, however, $tRNA^{Ser}$ has V loops of 19-22 nt and $tRNA^{Tyr}$ has a V loop of 14 nt, indicating that the bacterial $tRNA^{Ser}$ V loop expansions occurred after Bacteria diverged from Archaea and after type II $tRNA^{Tyr}$ formed but before Bacteria were fully established as a separate Domain. In support of this idea, $tRNA^{Ser}$ and $tRNA^{Tyr}$ in T. thermophilus have the same deletions in the D loop region of their tRNAs, indicating that the D loop deletions may have occurred prior to genesis of the type II $tRNA^{Tyr}$ and also prior to the $tRNA^{Ser}$ V loop expansions. We propose that Archaea and Bacteria diverged very early after LUCA, but that a full bacterial identity was only established much later in evolution. We have previously discussed divergence of archaeal and bacterial transcription systems [30], which fits a similar narrative. We posit that divergence of Archaea and Bacteria into separate Domains was largely driven by the divergence of their transcription systems. Archaea utilize RNA polymerase, TFD, TFB and TFE. Bacteria utilize a simplified RNA polymerase coevolved with bacterial sigma factors. Interestingly, archaeal TFB and bacterial sigma factors are helix-turn-helix factors and homologs.

Because type II $tRNA^{Tyr}$ in Bacteria has the same trajectory as $tRNA^{Leu}$ in Archaea, bacterial $tRNA^{Leu}$ evolved to alter its V loop trajectory, changing its contacts to LeuRS-IA from the V loop in Archaea to the T loop in Bacteria (Figures 4 and 7). Also, $tRNA^{Ser}$ in Bacteria altered its trajectory from the tRNA body relative to $tRNA^{Ser}$ from Archaea (compare Figures 3C, 4 and 8). At the time of writing, there is no Pyrococcus SerRS-IIA structure with bound $tRNA^{Ser}$ to make the fullest comparison. It appears that, in prokaryotes, type II V loops for different tRNAs must have different trajectories from the tRNA body in order to support the accurate activities of the interacting proteomes including cognate AARS enzymes.

We consider the analysis and comparisons of type II V loops to: 1) strongly support the 3-31 nt minihelix theorem for tRNA evolution; and 2) give insight into the divergence of Archaea and Bacteria.

5. Conclusions

We conclude that the comparison of type I and type II V loops and tRNAs in archaeal systems provides overwhelming support for the 3-31 nt minihelix theorem and falsifies:

 doi:10.20944/preprints202209.0189.v1

1) all accretion models (chaotic models with random indels) for tRNA evolution; and 2) all 2 minihelix models for tRNA evolution. The 3-31 nt minihelix theorem is the only description of both type I and type II tRNA evolution. No other published model provides this comparison. We understand that there are no theorems in biology. We propose that the 3-31 nt minihelix theorem be generally accepted as one and perhaps the only one. The 3-31 nt minihelix theorem describes pre-life processes and chemical evolution leading to life on Earth.

The current work demonstrates the importance of identifying and analyzing ancient root sequences to understand the pre-life to life transition and the divergence of Archaea and Bacteria. Identification of the patterning of the $tRNA^{Pri}$ root sequence provides insight into polymer world and minihelix world (Figure 1). Type I and type II $tRNA^{Pri}$ appear to be the root sequences for all tRNAomes on Earth. Our group has previously analyzed root trees of class I and class II AARS enzymes and correlated them with the evolution of the genetic code [1,2].

Bottom-up approaches attempt to reproduce pre-life chemistry. Our analysis, by contrast, is a top-down, sequence-based approach that nonetheless provides insight into a pre-life world. The patterning of tRNA sequences allows success of our top-down strategy. Note that chaotic sequences are highly unlikely to generate patterned sequences through subsequent random events. Because tRNA sequences include repeats and inverted repeats, tRNA must have been generated from patterned polymers and minihelices, and chaotic models are falsified (Figure 1).

Life may have evolved from an RNA-peptide world [2,31-35]. Also, modification of RNA at the 2′ position (i.e., 2′-O-methyl) may have stabilized RNA (i.e., to OH^- hydrolysis) and aided evolution of 5′→3′ complementary replication of RNA [31]. Complementary replication of RNA has, so far, been challenging to reproduce in vitro using primitive catalysts, such as ribozymes [36-38]. We show, however, that pre-life tRNA evolution required inverted repeats, snap-back primers, complementary sequences and complementary replication (Figure 1). We predict that, with modified procedures, efficient complementary 5′→3′ RNA replication can be reproduced in vitro using primitive catalysts. We posit that polyglycine was an important product of polymer world, minihelix world and the initial tRNA world before tRNAomes. We posit that pre-life evolved from highly ordered polymer and minihelix worlds to tRNA world to tRNAomes and the genetic code and, finally, to DNA genomes, cells and life. This history was recorded in the tRNA sequences of living organisms (Figures 1-3).

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Supplementary File #1: Archaeal $tRNA^{Leu}$; Supplementary File #2: Archaeal $tRNA^{Ser}$.

Author Contributions: We report the following author contributions: conceptualization, L.L. and Z. F. B.; methodology, L.L. and Z. F. B.; software, L.L. and Z. F. B..; validation, L.L. and Z. F. B.; formal analysis, L.L. and Z. F. B.; investigation, L.L. and Z. F. B.; resources, L.L. and Z. F. B.; data curation, L.L. and Z. F. B.; writing—original draft preparation, Z. F. B.; writing—review and editing, L.L. and Z. F. B.; visualization, L.L. and Z. F. B.; supervision, L.L. and Z. F. B.; project administration, Z. F. B.. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: See Supplementary Materials.

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

References

1. Lei, L.; Burton, Z.F. Evolution of the genetic code. *Transcription* **2021**, *12*, 28-53, doi:10.1080/21541264.2021.1927652.

2. Lei, L.; Burton, Z.F. Evolution of Life on Earth: tRNA, Aminoacyl-tRNA Synthetases and the Genetic Code. *Life (Basel)* **2020**, *10*, doi:10.3390/life10030021.
3. Burton, Z.F. The 3-Minihelix tRNA Evolution Theorem. *J Mol Evol* **2020**, *88*, 234-242, doi:10.1007/s00239-020-09928-2.
4. 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 (Basel)* **2019**, *9*, doi:10.3390/life9020037.
5. 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*, doi:10.3390/ijms19103275.
6. 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, doi:10.1007/s00239-020-09929-1.
7. Di Giulio, M. An RNA Ring was Not the Progenitor of the tRNA Molecule. *J Mol Evol* **2020**, *88*, 228-233, doi:10.1007/s00239-020-09927-3.
8. Di Giulio, M. A comparison between two models for understanding the origin of the tRNA molecule. *J Theor Biol* **2019**, *480*, 99-103, doi:10.1016/j.jtbi.2019.07.020.
9. Demongeot, J.; Seligmann, H. Theoretical minimal RNA rings mimick molecular evolution before tRNA-mediated translation: codon-amino acid affinities increase from early to late RNA rings. *C R Biol* **2020**, *343*, 111-122, doi:10.5802/crbiol.1.
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, doi:10.1007/s10441-019-09356-w.
11. Nagaswamy, U.; Fox, G.E. RNA ligation and the origin of tRNA. *Orig Life Evol Biosph* **2003**, *33*, 199-209, doi:10.1023/a:1024658727570.
12. 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, doi:10.1007/s00239-021-10002-8.
13. 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 (Basel)* **2016**, *6*, doi:10.3390/life6020015.
14. 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, doi:10.1016/j.fob.2014.01.010.
15. 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, doi:10.1080/21541264.2017.1318811.
16. Sun, F.J.; Caetano-Anolles, G. The Evolutionary Significance of the Long Variable Arm in Transfer RNA. *Complexity* **2009**, *14*, 26-39, doi:10.1002/cplx.20255.
17. 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*, D159-162, doi:10.1093/nar/gkn772.
18. 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, doi:10.1002/pro.3943.
19. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res* **2004**, *14*, 1188-1190, doi:10.1101/gr.849004.
20. Schneider, T.D.; Stephens, R.M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* **1990**, *18*, 6097-6100, doi:10.1093/nar/18.20.6097.
21. Demongeot, J.; Seligmann, H. The primordial tRNA acceptor stem code from theoretical minimal RNA ring clusters. *BMC Genet* **2020**, *21*, 7, doi:10.1186/s12863-020-0812-2.
22. Sajek, M.P.; Wozniak, T.; Sprinzl, M.; Jaruzelska, J.; Barciszewski, J. T-psi-C: user friendly database of tRNA sequences and structures. *Nucleic Acids Res* **2020**, *48*, D256-D260, doi:10.1093/nar/gkz922.

23. 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, doi:10.1017/s1355838200000364.
24. Pak, D.; Du, N.; Kim, Y.; Sun, Y.; Burton, Z.F. Rooted tRNAomes and evolution of the genetic code. *Transcription* **2018**, *9*, 137-151, doi:10.1080/21541264.2018.1429837.
25. 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, doi:10.1016/j.jmb.2004.11.060.
26. Dulic, M.; Cvetesic, N.; Zivkovic, I.; Palencia, A.; Cusack, S.; Bertosa, B.; Gruic-Sovulj, I. Kinetic Origin of Substrate Specificity in Post-Transfer Editing by Leucyl-tRNA Synthetase. *J Mol Biol* **2018**, *430*, 1-16, doi:10.1016/j.jmb.2017.10.024.
27. 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, doi:10.1126/science.8128220.
28. Wang, C.; Guo, Y.; Tian, Q.; Jia, Q.; Gao, Y.; Zhang, Q.; Zhou, C.; Xie, W. SerRS-tRNASec complex structures reveal mechanism of the first step in selenocysteine biosynthesis. *Nucleic Acids Res* **2015**, *43*, 10534-10545, doi:10.1093/nar/gkv996.
29. Root-Bernstein, R.; Kim, Y.; Sanjay, A.; Burton, Z.F. tRNA evolution from the proto-tRNA minihelix world. *Transcription* **2016**, *7*, 153-163, doi:10.1080/21541264.2016.1235527.
30. Lei, L.; Burton, Z.F. Early Evolution of Transcription Systems and Divergence of Archaea and Bacteria *Frontiers in Molecular Biosciences* **2021**, *8*, doi:10.3389/fmolb.2021.651134.
31. 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, doi:10.1038/s41586-022-04676-3.
32. Iglesias-Artola, J.M.; Drobot, B.; Kar, M.; Fritsch, A.W.; Mutschler, H.; Dora Tang, T.Y.; Kreysing, M. Charge-density reduction promotes ribozyme activity in RNA-peptide coacervates via RNA fluidization and magnesium partitioning. *Nat Chem* **2022**, *14*, 407-416, doi:10.1038/s41557-022-00890-8.
33. Gospodinov, A.; Kunnev, D. Universal Codons with Enrichment from GC to AU Nucleotide Composition Reveal a Chronological Assignment from Early to Late Along with LUCA Formation. *Life (Basel)* **2020**, *10*, doi:10.3390/life10060081.
34. Nainyte, M.; Muller, F.; Ganazzoli, G.; Chan, C.Y.; Crisp, A.; Globisch, D.; Carell, T. Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World. *Chemistry* **2020**, *26*, 14856-14860, doi:10.1002/chem.202002929.
35. Kunnev, D.; Gospodinov, A. Possible Emergence of Sequence Specific RNA Aminoacylation via Peptide Intermediary to Initiate Darwinian Evolution and Code Through Origin of Life. *Life (Basel)* **2018**, *8*, doi:10.3390/life8040044.
36. Setterholm, N.A.; Haratipour, P.; Kashemirov, B.A.; McKenna, C.E.; Joyce, G.F. Kinetic Effects of beta,gamma-Modified Deoxynucleoside 5'-Triphosphate Analogues on RNA-Catalyzed Polymerization of DNA. *Biochemistry* **2021**, *60*, 1-5, doi:10.1021/acs.biochem.0c00779.
37. 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 U S A* **2020**, *117*, 2906-2913, doi:10.1073/pnas.1914282117.
38. 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, doi:10.1021/acssynbio.9b00044.