Review

Not peer-reviewed version

Early Evolution of Type II tRNA Variable Loops, Aminoacyl-tRNA Synthetase Allostery from Distal Determinants and Origin and Diversification of Life

Lei Lei and Zachary Frome Burton *

Posted Date: 8 July 2024

doi: 10.20944/preprints202407.0578.v1

Keywords: tRNA evolution; type II tRNA; allostery; aminoacyl-tRNA synthetase; LeuRS; SerRS; TyrRS; origin of life; divergence of Archaea and Bacteria

Review

Early Evolution of Type II tRNA Variable Loops, Aminoacyl-tRNA Synthetase Allostery from Distal Determinants and Origin and Diversification of Life

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

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

Abstract: The 3 31 nucleotide minihelix tRNA theorem describes evolution of type I and type II tRNAs to the last nucleotide. In databases, type I and type II tRNA V loops (V for variable) were improperly aligned, but alignment based on the theorem is accurate. Type II tRNA V arms were a 3′-acceptor stem (initially CCGCCGC) ligated to a 5′-acceptor stem (initially GCGGCGG). The type II V arm evolved to form a stem-loop-stem. In Archaea, $tRNA^{Leu}$ and $tRNA^{Ser}$ are type II. In Bacteria, $tRNA^{Leu}$, $tRNA^{Ser}$ and $tRNA^{Tyr}$ are type II. The positioning of the type II V arm is determined by the number of unpaired bases just 5′ of the Levitt base (V_{max}). For Archaea, $tRNA^{Leu}$ has 2 unpaired bases, and $tRNA^{Ser}$ has 1 unpaired base. For Bacteria, $tRNA^{Tyr}$ has 2 unpaired bases, $tRNA^{Leu}$ has 1 unpaired base and $tRNA^{Ser}$ has 0 unpaired bases. Thus, the number of type II tRNAs is limited by the possible set points of the arm. From analysis of aminoacyl-tRNA synthetase structures, contacts to type II V arms appear to adjust allosteric tension communicated primarily via tRNA to aminoacylating and editing active sites. To enhance allostery, it appears that type II V arm end loop contacts may tend to evolve to V arm stem contacts.

Keywords: tRNA evolution; type II tRNA; allostery; aminoacyl-tRNA synthetase; LeuRS; SerRS; TyrRS; origin of life; divergence of Archaea and Bacteria

1. Introduction

Type II variable (V) loops of tRNAs have been misunderstood, and type I and type II tRNA V loops have been improperly aligned in tRNA databases [1–4]. Here, we describe: 1) the early evolution of type II V loops; 2) their proper alignment to type I V loops; and 3) their interactions with aminoacyl-tRNA synthetase (AARS) enzymes. We posit that type II V arms are important for allosteric communication with aminoacylating and editing active sites of cognate AARS.

The 3 31 nt minihelix tRNA evolution theorem fully describes the origin of type I and type II tRNAs [5–10]. Because type I and type II tRNAs first evolved ~4 billion years ago, this is a remarkable observation. The pre-life sequences of tRNAs and tRNA precursor molecules, however, are known with essential certainty because these sequences are ordered and conserved in living organisms. Type I and type II tRNAs evolved from RNA repeats and inverted repeats, allowing the initial pre-life sequence to be determined. Evolution of tRNA was from a 93 nt precursor molecule formed by ligation of 3 31 nt minihelices. The D loop 31 nt minihelix had the sequence GCGGCGG_UAGCCUAGCCUAGCCUA_CCGCCGC. The anticodon and T loop 31 nt minihelices had the sequence GCGGCGG_CCGGG_CU/???AA_CCCGG_CCGCCGC (_ delineates acceptor stems and stem-loop-stems; / indicates a U-turn; ? indicates sequence ambiguity).

The type I V loop was processed from the type II V loop by an internal 9 nt deletion. The initial type I V loop sequence was 5 nt in length (CCGCC; a fragment of a 3′-acceptor stem). The initial type II V loop sequence was 14 nt in length (CCGCCGCGCGGCGG). The type II V loop arose from a 3′-acceptor stem (CCGCCGC) ligated to a 5′-acceptor stem (GCGGCGG). In type II tRNAs, the type II

 doi:10.20944/preprints202407.0578.v1

V loop evolved to a stem-loop-stem. In a domain (i.e., Archaea or Bacteria), type II V arms have different trajectories from the tRNA, allowing for distinct recognition (i.e., by cognate AARS).

For translation functions, Archaea appear older than Bacteria and closer to LUCA (the last universal common (cellular) ancestor). We posit that Bacteria diverged from Archaea very early after LUCA [11]. After a significant time of separation, Bacteria assumed their new and distinct identity. Archaeal tRNAomes are more highly ordered and simpler than bacterial tRNAomes [5–8,12]. The archaeal genetic code is simpler and more ordered than the bacterial code [13]. Analysis of type II V loops, type II tRNAs and cognate AARS provides further insight into the divergence of Archaea and Bacteria.

The type II V arm interacts with AARS enzymes as a determinant for cognate AARS and as an anti-determinant for non-cognate AARS [13–16]. Leucine and serine may have entered the evolving genetic code at about the same time and both leucine and serine utilize type II tRNAs. Leucine, serine and arginine occupy 6 codon sectors within the code. LeuRS-IA, SerRS-IIA and AlaRS-IID are the AARS that lack anticodon loop recognition. During code evolution, serine jumped between column 2 and column 4, and serine is the only amino acid that occupies separate columns of the code. It is likely that serine jumping between columns required $tRNA^{Ser}$ being a type II tRNA and, also, jumping required SerRS-IIA lacking anticodon loop recognition. We posit that serine jumping during code evolution may have related to initial cysteine incorporation into the code.

Knowledge remains incomplete about allostery in cognate AARS charging of tRNAs [17,18]. It appears to us that, at least in some cases, recognition of tRNA determinants by AARS may generate allosteric communication largely via the tRNA (i.e., acting similarly to a coiled spring) to the tRNA 3′-end for cognate tRNA aminoacylation. For many type I tRNAs, allostery is largely initiated through accurate anticodon recognition to position the tRNA 3′-end for aminoacylation. Because LeuRS-IA and SerRS-IIA lack anticodon recognition and because $tRNA^{Leu}$ and $tRNA^{Ser}$ have type II V arms, the V arm assumes a more prominent role to tune allosteric communication. Because LeuRS-IA has separate aminoacylating and proofreading/editing active sites, allostery, mediated through different contacts of the V arm, directs the $tRNA^{Leu}$ 3′-end to switch between aminoacylating and editing active sites. In Bacteria but not Archaea, $tRNA^{Tyr}$ is a type II tRNA, and $tRNA^{Tyr}$ type II V arm contacts help direct accurate $tRNA^{Tyr}$ charging in Bacteria.

2. Type I and Type II tRNA V Loops

Figure 1 shows type I and type II tRNAs and proper alignments of V loops. The initial type I V loop sequence was CCGCC. The initial type II V loop sequence was CCGCCGC_GCGGCGG, derived from ligation of a 3′-acceptor stem (CCGCCGC) to a 5′-acceptor stem (GCGGCGG) [5]. Coloring of the tRNAs is according to the 3 31 nt tRNA evolution theorem that completely describes evolution of type I and type II tRNAs [5]. Colors in common indicate homologous sequences that were identical sequences in pre-life. In existing tRNA databases [3,4,19], type I and type II V loops were improperly aligned to one another because the 3 31 nt minihelix tRNA evolution theorem was not applied.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0578.v1

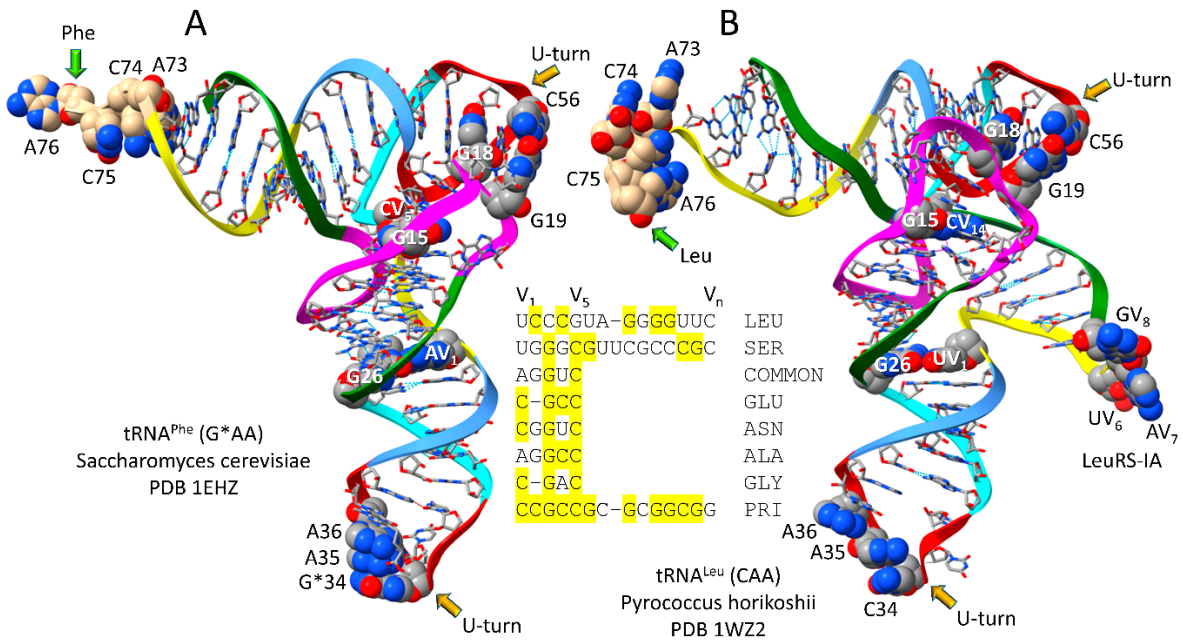


Figure 1. Comparison of type I and type II tRNAs and their proper alignment. A) a type I $tRNA^{Phe}$ from Saccharomyces cerevisiae (a Eukaryote) [20]. B) a type II $tRNA^{Leu}$ from Pyrococcus horikoshii (an ancient Archaeon) [21]. Colors reflect the 3 31 nt tRNA evolution theorem: green) 5′-acceptor stems and 5′-acceptor stem remnants; magenta) D loop; cyan-red-cornflower blue) anticodon and T stem-loop-stems; yellow) 3′-acceptor stems and 3′-acceptor stem remnants. In the images, some bases are emphasized using space-filling representation. ChimeraX was used for molecular graphics [22–24]. Alignment of some type I and type II V loops from Pyrococcus furiosus is based on the theorem. PRI indicates the $tRNA^{Pri}$ type II V loop (PRI for primordial; the initial pre-life sequence). COMMON indicates the most common P. furiosus type I V loop sequence.

3. Different Trajectories of the Type II tRNA V Arms

A gallery of type II tRNAs is shown in Figure 2 emphasizing the contacts and trajectories of the V arms. The type II $tRNA^{Pri}$ V arm is self-complementary along its entire length and can form many different or tangled C=G pairings. V arms, therefore, evolved to form stem-loop-stems that could be discriminated by cognate AARS. We find that the trajectory of the V arm depends on the number of unpaired bases just 5′ to the Levitt base pair (2, 1 or 0; sometimes -1). The Levitt base pair is a reverse Watson-Crick pair between tRNA base 15 and V_n (for a V loop of n bases (numbered V_1-V_n)). A reverse Watson-Crick base pair is a standard Watson-Crick pair, as in DNA, with one of the bases flipped over.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0578.v1

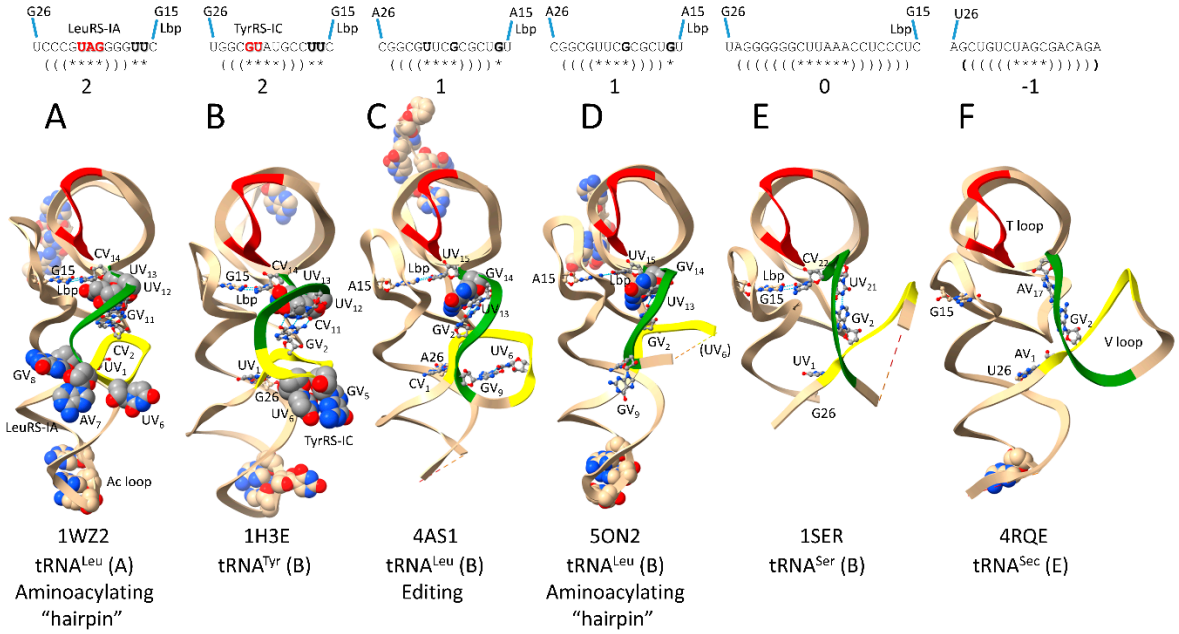


Figure 2. Distinct trajectories of type II V arms depend on the number of unpaired bases just 5′ of the Levitt base V_n (2, 1, 0 or -1). A) Archaeal (A for archaeal) $tRNA^{Leu}$ in the aminoacylating "hairpin" conformation [21,25]; B) Bacterial (B for bacterial) $tRNA^{Tyr}$ [26]; C) Bacterial $tRNA^{Leu}$ in the editing/proofreading conformation [27]; D) Bacterial $tRNA^{Leu}$ in the aminoacylating "hairpin" conformation [28]; E) Bacterial $tRNA^{Ser}$ [29]; and F) Human (E for Eukarya) $tRNA^{Sec}$ (Sec for selenocysteine) [30]. Colors: red) T loop; yellow) 1st 7 nt of the V loop; and green) last 7 nt of the V loop. Space-filling and ball and stick bases are shown for emphasis. Parentheses indicate stems; * indicates loop. Lbp indicates the Levitt base pair. See the text for details.

Figure 2A shows archaeal Pyrococcus horikoshii $tRNA^{Leu}$ (CAA) in the aminoacylating "hairpin" conformation [21,25]. Structures of tRNAs are from co-crystal structures with cognate AARS. "Hairpin" relates to the bent down 3′-end of the $tRNA^{Leu}$ into the LeuRS-IA aminoacylating active site. V loops are numbered V_1-V_n for a V loop of n bases. For historical reasons, standard numbering of tRNAs breaks down in the D loop and V loop, explaining why we use the V_1 to V_n numbering here. P. horikoshii (an ancient Archaeon) $tRNA^{Leu}$ (CAA) has a V loop of 14 nt, which is the primordial (pre-life) length [5–7]. UV_1 forms a wobble pair with tRNA-G26 (UV_1~G26). V_2-CCC-V_4 is the V arm 5′-stem. V_5-GUAG-V_8 is the V arm end loop. V arm end loop bases V_6-UAG-V_8 are highly conserved in Archaea and bind to archaeal LeuRS-IA, as indicated (see below). V_9-GGG-V_{11} forms the V arm 3′-stem. V_{12}-UU-V_{13} are 2 unpaired bases, just 5′ of the Levitt base (CV_{14}). The Levitt base (CV_{14}) forms a reverse Watson-Crick base pair with tRNA-G15 (CV_{14}=G15) (the Levitt base pair). The number of unpaired bases just 5′ of the Levitt base V_n determines the trajectory of the V arm from the tRNA body.

Figure 2B shows bacterial Thermus thermophilus $tRNA^{Tyr}$ (GU*A) (U* for pseudouridine) from a co-crystal with TyrRS-IC [26]. The V loop is 14 nt, which is the primordial length. UV_1 forms a wobble pair with tRNA-G26 (UV_1~G26). The V arm 5′-stem is V_2-GGC-V_4. The V arm end loop is V_5-GUAU-V_8. V arm end loop bases V_5-GU-V_6, which along with V_5-UU-V_6 are highly conserved in Bacteria with $tRNA^{Tyr}$ V loops of 14 nt, bind to TyrRS-IC, as indicated (see below). Type II $tRNA^{Tyr}$ V loops of less than 14 nt lose the conserved V_5 and V_6 bases and, presumably, also V arm end loop contacts by TyrRS-IC (see below). V_9-GCC-V_{11} form the V arm 3′-stem. V_{12}-UU-V_{13} are unpaired bases just 5′ of the Levitt base (CV_{14}), which forms the Levitt base pair with tRNA-G15 (CV_{14}=G15).

Figure 2C shows bacterial Escherichia coli $tRNA^{Leu}$ (UAA) from a co-crystal with LeuRS-IA in the editing/proofreading conformation (the 3′-end of $tRNA^{Leu}$ is in the LeuRS-IA editing/proofreading active site rather than the aminoacylating active site) [27]. The $tRNA^{Leu}$ (UAA) V loop is 15 nt in E. coli. CV_1 interacts with tRNA-A26 (CV_1~A26). V_2-GGCG-V_5 is the V arm 5′-stem.

 doi:10.20944/preprints202407.0578.v1

V_6-UUCG-V_9 is the V arm end loop. In the editing conformation, UV_6 and GV_9 interact, and the V arm end loop is ordered. V_{10}-CGCU-V_{13} is the V arm 3′-stem. GV_{14} is the one unpaired base that determines the V arm trajectory, just 5′ of the Levitt base UV_{15}. UV_{15} forms the Levitt reverse Watson-Crick base pair with tRNA-A15 (UV_{15}=A15).

Figure 2D shows the same Escherichia coli $tRNA^{Leu}$ (UAA) from a LeuRS-IA co-crystal but in the aminoacylating hairpin conformation [28]. The major differences from the editing/proofreading conformation image in Figure 2C are as follows: 1) the conformation of the V arm 3′-stem (green) is altered; 2) the V arm end loop is disordered (Figure 2D); 3) the orientation of GV_9 is changed; and 4) UV_6 has lost contact to GV_9 and is disordered (Figure 2D). We posit that allosteric interactions initiated by appropriate (aminoacylating) or inappropriate (editing/proofreading) contacts at the $tRNA^{Leu}$ (UAA) 3′-end are communicated to and amplified by contacts at the V arm (see below).

Figure 2E shows bacterial Thermus thermophilus $tRNA^{Ser}$ (GGA) from a co-crystal with SerRS-IIA [29]. The $tRNA^{Ser}$ (GGA) is significantly disordered, and the structure is not in an aminoacylating conformation (the $tRNA^{Ser}$ 72-CGCCA 3′-end is disordered). The V loop is 22 nt. UV_1 can probably interact with tRNA-G26 (UV_1~G26) (tRNA-G26 is mostly disordered in the structure). V_2-AGGGGGG-V_8 is the V arm 5′-stem. The V arm end loop is V_9-CUUAAA-V_{14} (mostly disordered). The V arm 3′-stem is V_{15}-CCUCCCU-V_{21}. CV_{22} is the Levitt base that pairs with tRNA-G15 (CV_{22}=G15). No unpaired bases are present just 5′ of the Levitt reverse Watson-Crick base pair (trajectory set point of 0). In Bacteria, $tRNA^{Ser}$ V arm stems are generally longer than archaeal $tRNA^{Ser}$ V arm stems. We posit that the lengthening of the stem may reflect a need to stabilize the V_2=$V_{(n-1)}$ pair. In Archaea, it is the V_2=$V_{(n-2)}$ pair that forms (in Archaea, the set point trajectory of the $tRNA^{Ser}$ V arm is 1, in contrast to 0 in Bacteria).

Figure 2F shows a somewhat unusual case that is included here mostly for completeness. Human $tRNA^{Sec}$ (UCA) is shown from a co-crystal with SerRS-IIA (Sec for selenocysteine) [30]. SerRS-IIA attaches serine to $tRNA^{Sec}$ (Ser-$tRNA^{Sec}$), which is then converted to Sec-$tRNA^{Sec}$ by other enzymes utilizing tRNA-linked chemistry. The V loop is 17 nt. In this case, AV_1 probably interacts with tRNA-U26 (AV_1=U26). V_2-GCUGUC-V_7 is the V arm 5′-stem. V_8-UAGC-V_{11} is the V arm end loop. V_{12}-GACAGA-V_{17} is the V arm 3′-stem. The GV_2~AV_{17} interaction is notable. Because AV_{17} interacts with GV_2, the Levitt base pair to tRNA-G15 cannot form. Because human $tRNA^{Sec}$ is somewhat of an unusual case, it is not described in more detail here.

We conclude that V arm trajectories are different with 2, 1 or 0 unpaired bases just 5′ of the Levitt base. In the view shown in Figure 2, Figure 2A,B (trajectory score of 2; 2 unpaired bases just 5′ of the Levitt base) have the type II V arm extending almost straight back. Figure 2C,D (score of 1) have the V arm angling more to the right, and Figure 2E,F (scores of 0 and -1) have the V arm pointing to the right. At the time of writing, some desired images were not available. For instance, an archaeal SerRS-IIA-$tRNA^{Ser}$ structure would be useful (V loop trajectory score 1 in Archaea versus 0 in Bacteria).

4. LeuRS-IA-$tRNA^{Leu}$ Co-Crystal Structures

Figure 3 shows a co-crystal of archaeal Pyrococcus horikoshii LeuRS-IA complexed with $tRNA^{Leu}$ (CAA) [21,25]. P. horikoshii is an ancient Archaeon with a translation system very similar to the one that must have functioned at LUCA (the last universal common (cellular) ancestor). The $tRNA^{Leu}$ (CAA) is in the “hairpin” conformation with the tRNA 3′-end bent down into the aminoacylating active site. We could not identify an archaeal co-crystal structure with LeuRS-IA-$tRNA^{Leu}$ in an editing/proofreading conformation for comparison.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0578.v1

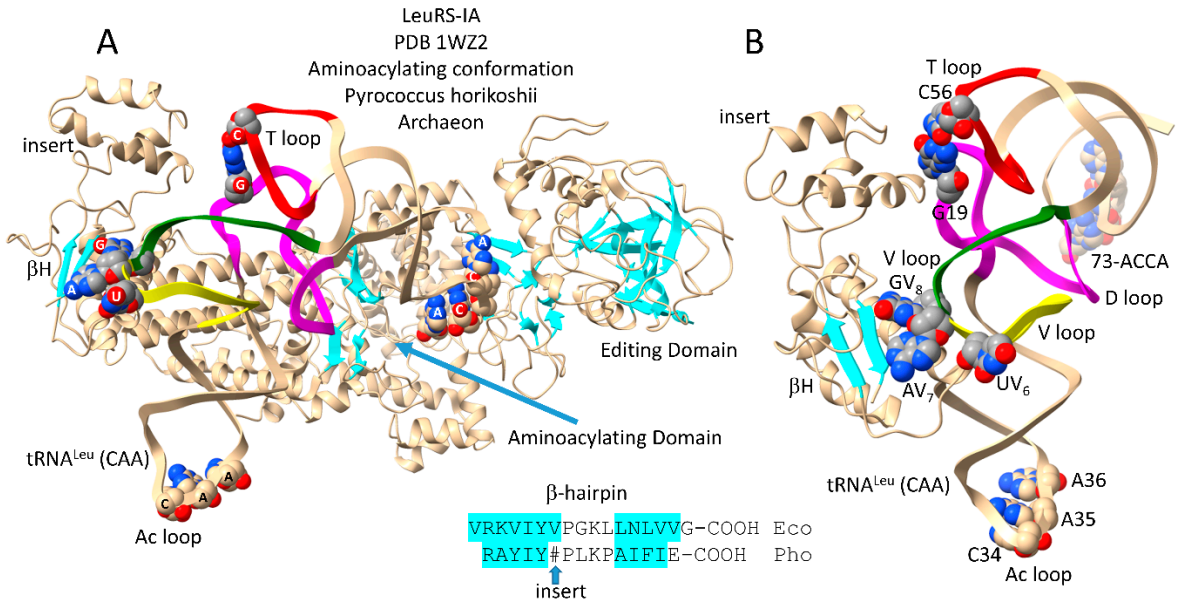


Figure 3. Archaeal LeuRS-IA bound with tRNA^{Leu} (CAA) in the aminoacylating conformation [21,25]. A) The entire structure; B) A C-terminal LeuRS-IA detail emphasizing V arm end loop contacts. β-sheets are cyan. Colors: magenta) D loop; yellow) 1st 7 nt of the V loop; green) last 7 nt of the V loop; and red) T loop. The sequence of a β-hairpin (βH) at the C terminus of Escherichia coli (Eco) and P. horikoshii (Pho) LeuRS-IA is shown. # indicates an ~93 amino acid insert in the Pho C-terminal β-hairpin.

Figure 3A shows the archaeal LeuRS-IA-tRNA^{Leu} (CAA) structure in the aminoacylating conformation. LeuRS-IA has separate aminoacylating and editing/proofreading active sites. The tRNA^{Leu} (CAA) 3′-end is in the hairpin conformation for aminoacylation. Because LeuRS-IA is a class I AARS, the aminoacylating active site is also identified by parallel β-sheets. Figure 3B shows a C terminal fragment of LeuRS-IA interacting with the tRNA^{Leu} V arm. In Archaea, the highly conserved V arm end loop sequence V_6-UAG-V_8 interacts with the C terminal LeuRS-IA β-hairpin. In Pyrococcus horikoshii but not in Bacteria, the β-hairpin has an ~93 amino acid insert that interacts with the tRNA^{Leu} elbow, where the D loop and T loop interact (i.e., where tRNA-G19 pairs with tRNA-C56 (a bent Watson-Crick pair)). The length of the insert in the archaeal C-terminal LeuRS-IA β-hairpin depends on how sequences are aligned. The V loop is 14 nt, which is the primordial (pre-life) length [5–7]. More detail about the tRNA^{Leu} (CAA) V loop is shown in Figure 2A. Comparison of an archaeal editing/proofreading LeuRS-IA-tRNA^{Leu} structure would enrich this discussion.

Figure 4 shows a bacterial Escherichia coli LeuRS-IA-tRNA^{Leu} (UAA) co-crystal structure in the aminoacylating hairpin conformation [27]. Figure 4A shows the intact structure. Figure 4B shows a LeuRS-IA C-terminal domain detail, highlighting interactions with the type II V loop. In contrast to archaeal LeuRS-IA (Figure 3), bacterial LeuRS-IA does not make contact to the V arm end loop (Figure 4). Furthermore, the C-terminal domain of bacterial LeuRS-IA, which lacks the ~93 amino acid insert of archaeal LeuRS-IA, is significantly rearranged and redirected compared to the archaeal C-terminal domain. Notably, the bacterial C-terminal β-hairpin interacts with the tRNA^{Leu} elbow (i.e., the tRNA-G19 bent Watson-Crick pair with tRNA-C56). K809, R811 and R837 interact with the V arm 3′-stem (green). The V arm end loop is disordered, and V arm end loop base GV_9 is flipped out of the end loop. See also Figure 2D. We posit that K809, R811 and K837 interaction with the V arm 3′-stem (green) generates allosteric communication with the tRNA^{Leu} 3′-end, particularly in the aminoacylating conformation, and that tighter binding to the V arm 3′-stem in the aminoacylating conformation disrupts the structure of the V arm end loop.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0578.v1

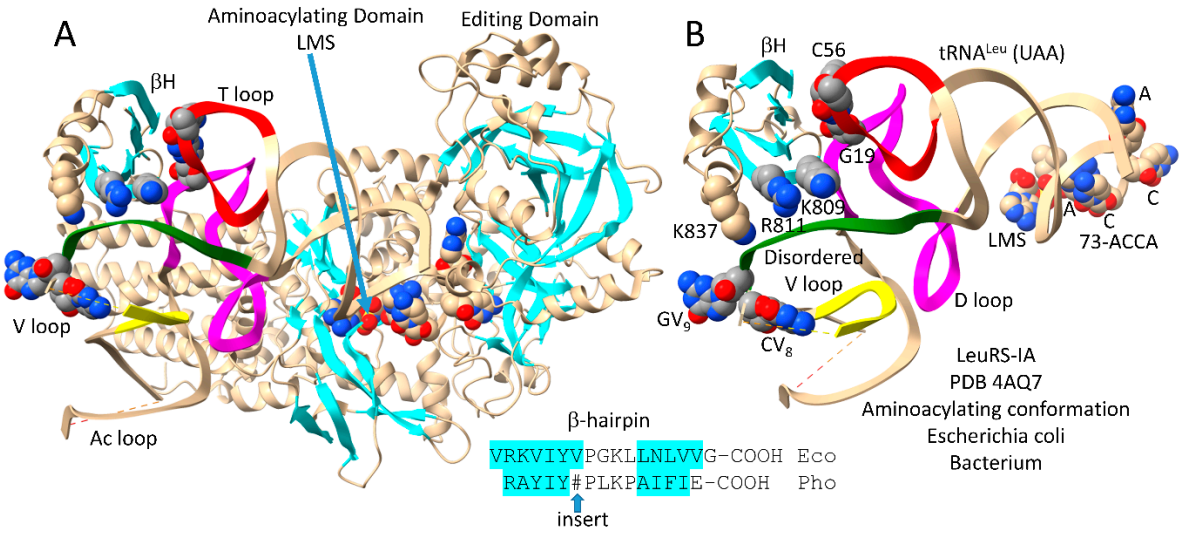


Figure 4. Bacterial Escherichia coli LeuRS-IA-tRNA^{Leu} (UAA) co-crystal structure in the aminoacylating hairpin conformation [27]. A) the full structure; B) a C-terminal LeuRS-IA detail emphasizing V arm 3′-stem contacts and the conformation of the V loop. LMS is a non-reactive leucine-AMP reaction intermediate analogue. βH indicates the C-terminal LeuRS-IA β-hairpin. See the text for details. Colors as in other figures.

Figure 5 shows bacterial Escherichia coli LeuRS-IA-tRNA^{Leu} (UAA) in the editing/proofreading conformation [27]. The tRNA^{Leu} 3′-end, which was chemically modified, locates to the editing active site. Figure 5A shows the entire structure. Figure 5B highlights C-terminal LeuRS-IA-tRNA^{Leu} V arm contacts, which are significantly altered from the aminoacylating conformation (Figure 4). The C-terminal β-hairpin maintains its contact to the tRNA-G19=tRNA-C56 pair at the tRNA^{Leu} elbow. K809 and R811 move away from the tRNA^{Leu} V arm 3′-stem. K837 that is visualized in the aminoacylating conformation (Figure 4) is unstructured in the editing/proofreading conformation (Figure 5). Because of weakened contacts to the V arm 3′-stem in the editing conformation, the tRNA^{Leu} V arm end loop is structured. GV_9 stacks on CV_8 and interacts with UV_6 (see Figure 2C; compare to Figure 4). We posit that distal tRNA^{Leu} V arm and elbow determinant contacts act as allosteric effectors to influence accurate tRNA charging and editing. Allosteric communication is transmitted back and forth from the tRNA^{Leu} 3′-end and the type II V arm, such that contacts to the V arm amplify allostery initiated by tRNA^{Leu} 3′-end contacts. It is much more difficult to imagine strong allosteric communication being transmitted through the LeuRS-IA protein structure. For instance, the protein connection of the C-terminal LeuRS-IA domain to the aminoacylating active site is very flexible (Figure 5A).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0578.v1

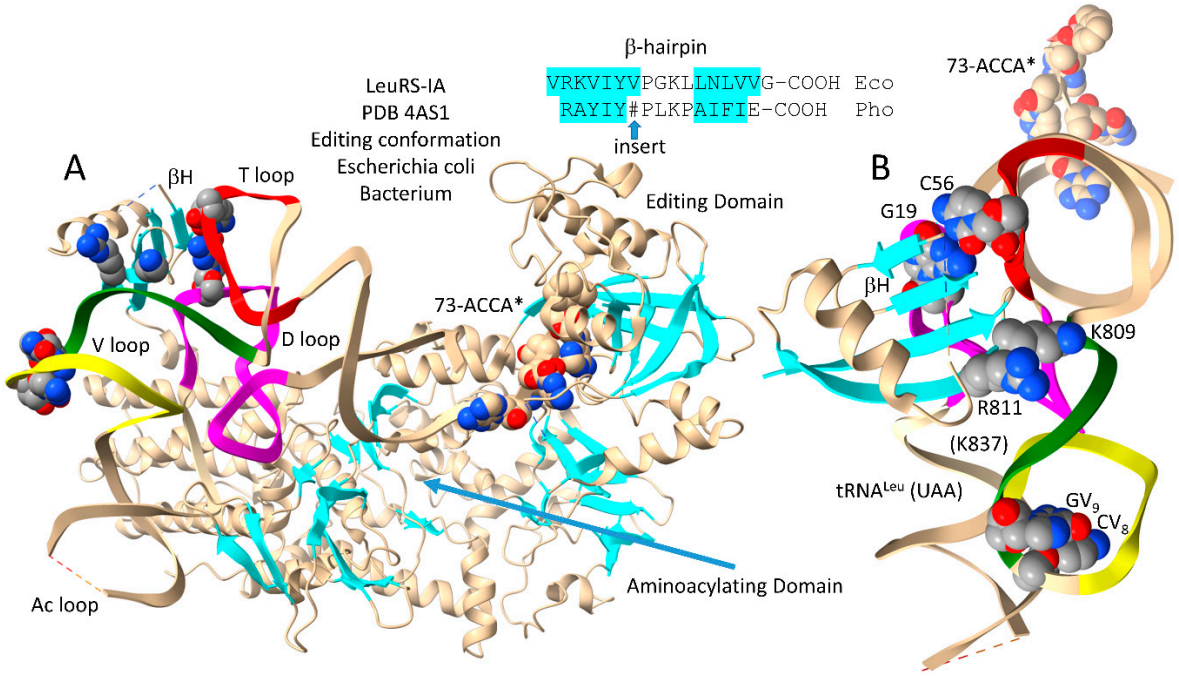
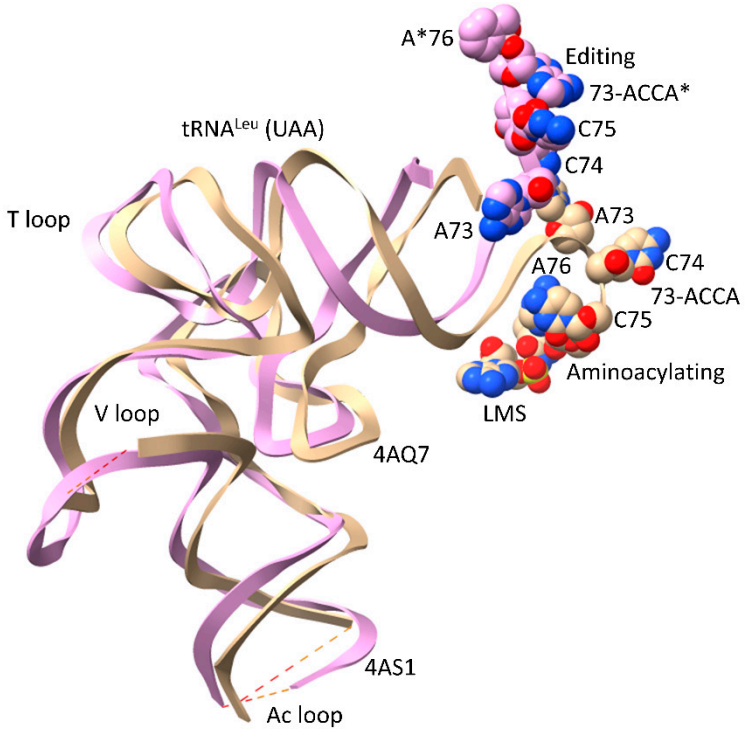


Figure 5. Bacterial Escherichia coli LeuRS-IA bound to $tRNA^{Leu}$ (UAA) in the editing/proofreading conformation [27]. A) The full structure; B) C-terminal LeuRS-IA-$tRNA^{Leu}$ contacts emphasizing elbow and weakened V arm 3′-stem contacts (compare to Figure 4). A* indicates a modified 3′-A76 base that directs the $tRNA^{Leu}$ 3′-end into the editing/proofreading active site. Colors as in previous figures.

Figure 6 shows an overlay of Escherichia coli $tRNA^{Leu}$ (UAA) in the aminoacylating and editing/proofreading conformations to indicate possible allosteric communication [18,27]. Structures were aligned based on LeuRS-IA protein. Because leucine occupies a 6 codon box in the genetic code, LeuRS-IA makes no contacts with the $tRNA^{Leu}$ anticodon loop. Instead, allosteric communication is established between the $tRNA^{Leu}$ V arm and the 3′-end, reinforcing the appropriate 3′-end placement into the aminoacylating or the editing/proofreading active site.



 doi:10.20944/preprints202407.0578.v1

Figure 6. Overlay of LeuRS-IA-tRNA^{Leu} (UAA) co-crystal structures indicates allosteric communication linking the tRNA^{Leu} 3′-end and the type II V arm 3′-stem [27]. The beige tRNA^{Leu} is in the aminoacylating hairpin conformation, with an unstructured V arm end loop (see Figure 4). LMS is a non-reactive leucine-AMP reaction intermediate analogue bound in the aminoacylating active site. The pink tRNA^{Leu} is in the editing/proofreading conformation, with a structured V arm end loop (see Figure 5). A*76 indicates a modified 3′-A* that directed the tRNA^{Leu} 3′-end into the editing/proofreading active site.

4. SerRS-IIA-tRNA^{Ser} and -tRNA^{Sec} Co-Crystal Structures

Figure 7 shows a bacterial Thermus thermophilus SerRS-IIA-tRNA^{Ser} co-crystal structure [29,31]. As do most or all class II AARS, SerRS-IIA functions as an α_2-dimer. Serine is in a 6 codon box in the genetic code. Consistent with a 6 codon box, SerRS-IIA lacks tRNA^{Ser} anticodon recognition. Figure 7A is the entire structure (1SER) supplemented with a light blue N-terminal helix hairpin and SSA non-reactive reaction intermediate analogue from 1SET. Figure 7B highlights SerRS-IIA-tRNA^{Ser} type II V arm contacts. An N-terminal helix hairpin forms a brace that contacts the tRNA elbow (tRNA-G19=tRNA-C56) (magenta and red), the V arm 3′-stem (yellow) and the V arm 5′-stem (green). Perhaps because only a single tRNA^{Ser} is bound, the N-terminal helix hairpin from the other α subunit was not visualized in the 1SER structure. The tRNA^{Ser} 3′-end 72-CGCCA is disordered, so the structure does not fully reflect an aminoacylating conformation. Perhaps the reaction must proceed to the serine-AMP stage to more stably attract the tRNA^{Ser} 3′-end. The tRNA^{Ser} structure shows a significant amount of disorder, possibly indicating allosteric communication linking the V arm and 3′-end contacts. The mode of SerRS-IIA-tRNA^{Ser} binding may have facilitated serine jumping from column 2 to column 4 of the genetic code. Serine is the only amino acid to have split between two genetic code columns. The binding of only a single tRNA^{Ser} by SerRS-IIA may indicate negative cooperativity in tRNA^{Ser} binding.

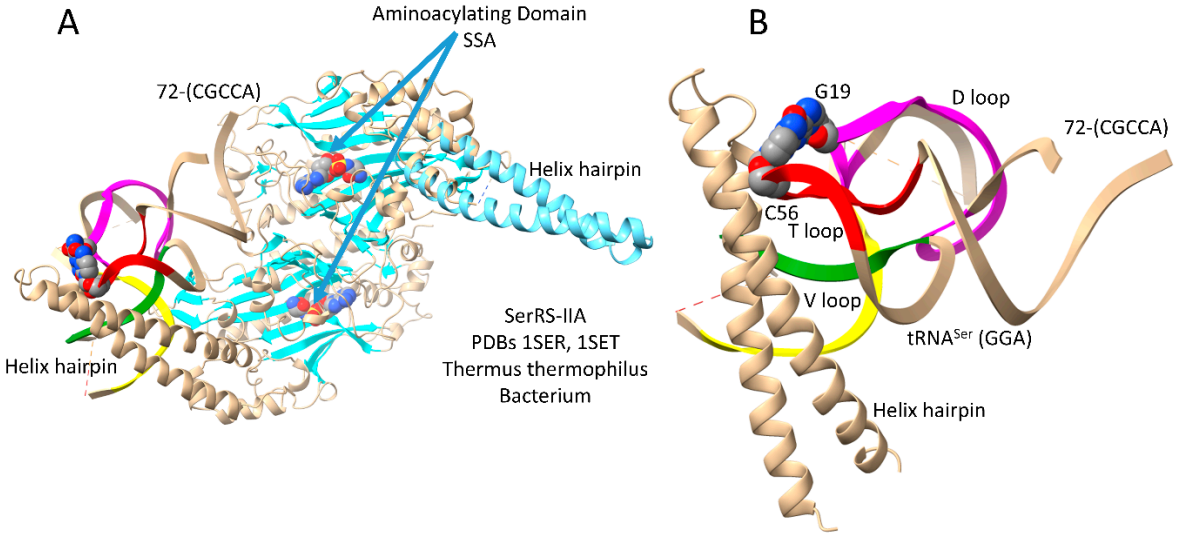


Figure 7. A Thermus thermophilus SerRS-IIA-tRNA^{Ser} (GGA) co-crystal structure [29,31]. The SerRS-IIA active site is formed on a surface of antiparallel β-sheets. A) The entire structure. The structure shown is a composite of the 1SER (beige) and 1SET (light blue; SSA) structures. B) A detail that highlights V arm stem contacts. Colors as in other figures.

Figure 8 shows a human SerRS-IIA-tRNA^{Sec} (UCA) co-crystal structure [30]. Because not all organisms encode selenocysteine (Sec), this is somewhat of an unusual case. Anticodon UCA represents stop codon UGA. The structure is important for the current discussion partly because it helps clarify some aspects of the structure shown in Figure 7. In Figure 8A, two tRNA^{Sec} (UCA) are bound, so both N-terminal helix hairpins are observed, rendering the structure more intuitive to observe (compare to Figure 7). The aminoacylating active site is identified by: 1) a surface of

 doi:10.20944/preprints202407.0578.v1

antiparallel β-sheets; 2) ANP binding (ANP is a non-reactive ATP analogue); and 3) serine binding. Because the tRNA^{Sec} 3′-end (72-CGCCA) is disordered, the structure does not fully reflect an aminoacylating conformation. The tRNA^{Sec} (UCA) has a somewhat unusual V loop structure with a broken Levitt pair and a GV_2=AV_{17} interaction (Figure 2F). The altered tRNA^{Sec} (UCA) type II V arm trajectory (score of -1 (tRNA^{Sec}) versus score of 0 (tRNA^{Ser})) may aid in specifying subsequent tRNA-linked chemistry to convert Ser-tRNA^{Sec} to Sec-tRNA^{Sec} and to not improperly convert Ser-tRNA^{Ser} to Sec-tRNA^{Ser}. It may be that apparent negative cooperativity observed in SerRS-IIA-tRNA^{Ser} binding (Figure 7) is relieved in SerRS-IIA-tRNA^{Sec} binding (Figure 8).

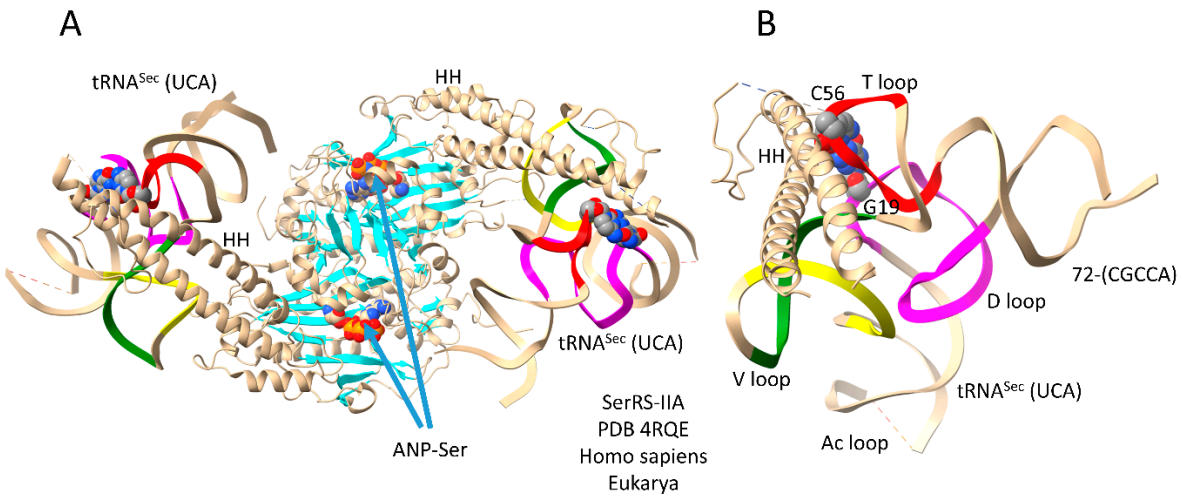


Figure 8. A human SerRS-IIA-tRNA^{Sec} (UCA) co-crystal structure [30]. A) The entire structure; B) an image highlighting tRNA^{Sec} (UCA) V arm stem and elbow contacts. ANP is a non-reactive ATP analogue. HH indicates helix hairpin. Colors as in other figures.

5. ArgRS-IA-tRNA^{Arg}

This section is included for completeness and to tie analysis of type II tRNAs into the evolution of the genetic code. Leucine, serine and arginine are within 6 codon sectors of the genetic code and probably entered the genetic code at about the same time [14–16]. In contrast to LeuRS-IA and SerRS-IIA, ArgRS-IA utilizes a type I tRNA and, also, anticodon recognition for tRNA^{Arg} (Figure 9). Because ArgRS-IA utilizes 5 tRNA^{Arg} anticodons from two genetic code rows (2 and 3), there is ambiguity in reading the anticodon sequence. The structure shown is a yeast ArgRS-IA-tRNA^{Arg} (ICG) (I for inosine). Inosine is formed by deamination of adenine. In Archaea, wobble tRNA-34A is not utilized. When A is modified to I in Bacteria and Eukarya, the corresponding G anticodon (i.e., GCG) is not utilized. Because tRNA-34I reads mRNA wobble A, C and U, wobble inosine can only be utilized in a 3 or 4 codon sector [32,33]. The structure in Figure 9 has tRNA^{Arg} (ICG) in the hairpin conformation with the tRNA^{Arg} 73-GCCA bending down into the aminoacylating active site, which is also identified by parallel β-sheets and arginine binding. The anticodon loop has the sequence 34-ICGAA-38 [34]. Because of ambiguity in tRNA^{Arg} anticodon reading, the anticodon loop is substantially unwound, exposing tRNA-38A to interact with ArgRS-IA. Because of anticodon ambiguity, the strongest sequence-specific contacts are expected to tRNA-35C and tRNA-38A. Substantial unwinding of the anticodon loop is expected to generate torque on the tRNA^{Arg} 3′-end to support the aminoacylating conformation.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0578.v1

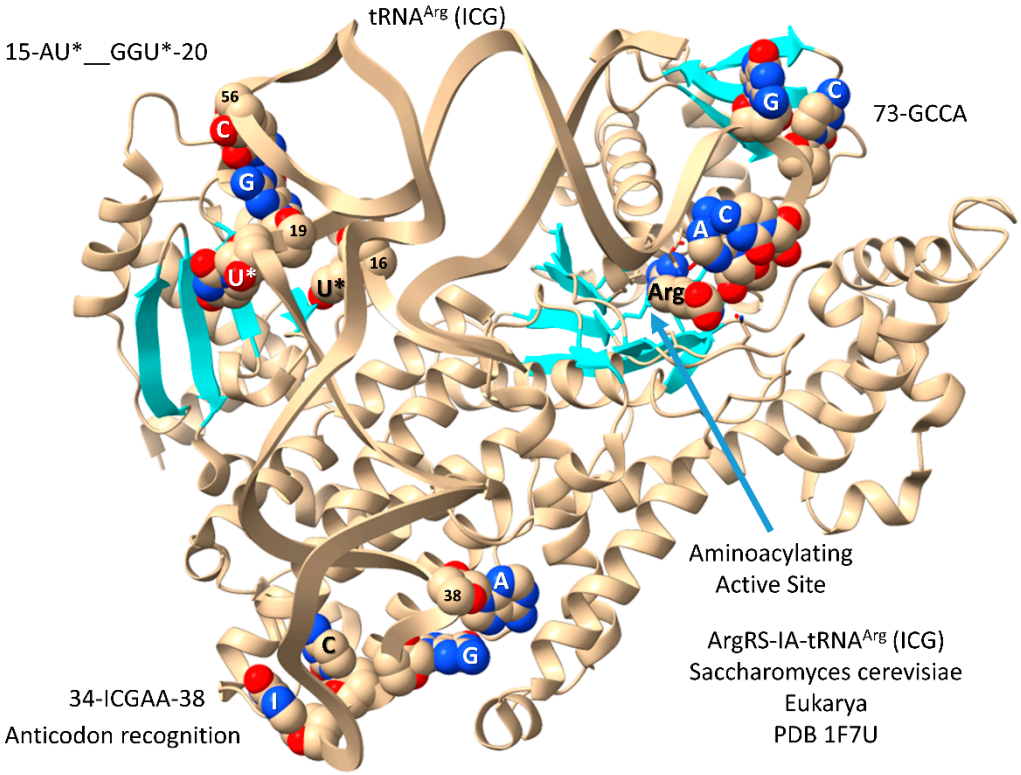


Figure 9. Saccharomyces cerevisiae ArgRS-IA-tRNA^{Arg} (ICG) [34]. U* for 5,6-dihydrouridine. This is a very good structure for molecular dynamics studies.

Numbering in the tRNA D loop is confusing, because, for historical reasons, numbering was based on eukaryotic tRNAs with 3 deleted D loop nts. The D loop evolved from a 17 nucleotide UAGCC repeat (i.e., D_1-UAGCCUAGCCUAGCCUA-D_{17}) [5,6]. According to improved numbering, 15-AU*--GGU*-20 would be numbered D_8-AU*--GGU*-D_{14} (with 2 deleted nts (D_{10} and D_{11})). The tRNA^{Arg} (ICG) elbow (G19=C56) and D loop make contact to ArgRS-IA, as shown.

6. TyrRS-IC-tRNA^{Tyr} (GUA) in Archaea and Bacteria

Interestingly, tRNA^{Tyr} is a type I tRNA in Archaea and a type II tRNA with a longer V loop in Bacteria [35]. An archaeal TyrRS-IC-tRNA^{Tyr} co-crystal is shown in Figure 10. In contrast to most other class I AARS, which are monomers, class IC AARS are obligate α_2-dimers with anticodon-binding domains and aminoacylating domains in opposite subunits. Via TyrRS-IC protein, the anticodon-interaction domain is only loosely connected to the aminoacylating domain, indicating that allosteric contacts at the anticodon-binding region may be communicated to the aminoacylating active site mostly via the tRNA. Because the tRNA^{Tyr} (GUA) 3′-end is disordered (only A73 of 73-ACCA is ordered), the structure is not in a fully aminoacylating conformation. The aminoacylating active site is indicated by binding of tyrosine and parallel β-sheets.

 doi:10.20944/preprints202407.0578.v1

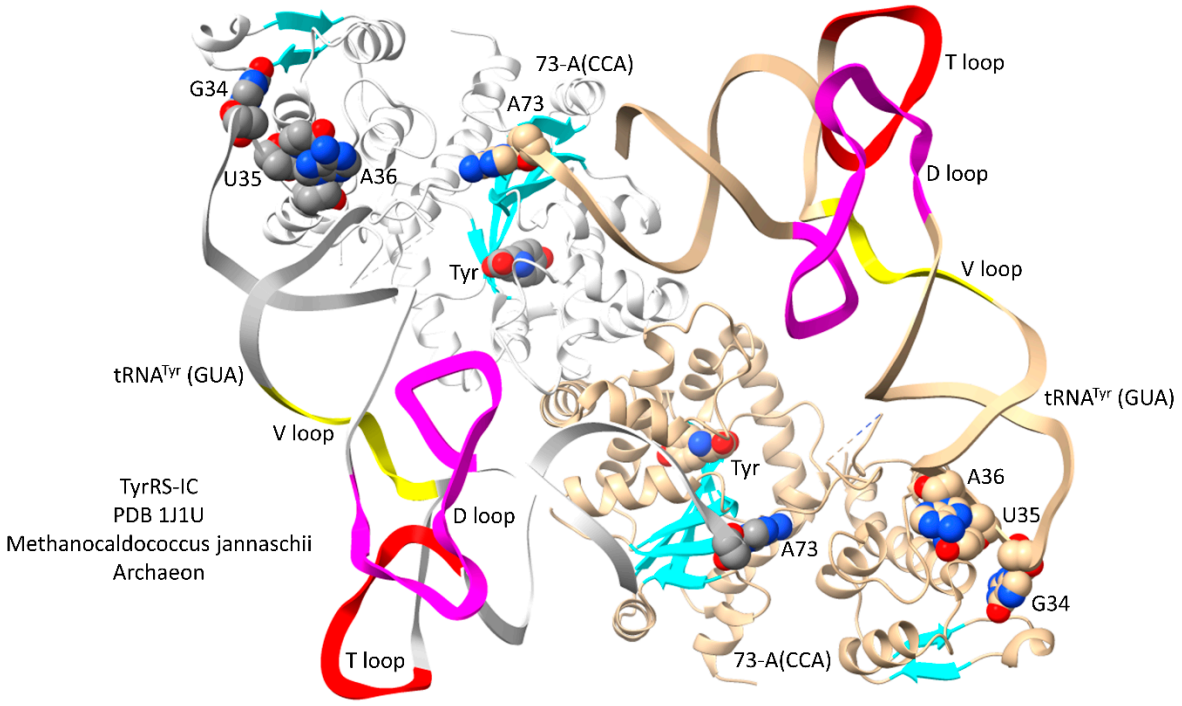


Figure 10. Archaeal TyrRS-IA bound to type I $tRNA^{Tyr}$ [35]. Colors as in other figures. One α subunit and $tRNA^{Tyr}$ are white.

In Bacteria, $tRNA^{Tyr}$ is a type II tRNA with a longer V loop [26]. In the ancient Bacterium Thermus thermophilus, $tRNA^{Tyr}$ has a type II V loop of 14 nt, the primordial length. A co-crystal structure of T. thermophilus TyrRS-IC bound to $tRNA^{Tyr}$ is shown in Figure 11. Figure 11A shows the entire structure. Figure 11B shows more detailed contacts by a C-terminal TyrRS-IC fragment to the $tRNA^{Tyr}$ V arm 5′-stem and end loop. The aminoacylating active site is identified by: 1) parallel β-sheets; 2) ATP binding; and 3) TYE (a non-reactive tyrosine analogue) binding. Because $tRNA^{Tyr}$ 74-CCA is unstructured, the image does not fully represent an aminoacylating conformation. The type II V arm is contacted by TyrRS-IC R388 and R389 on its 5′ stem (yellow). V arm end loop bases V_5-GU-V_6 bind TyrRS-IC. The TyrRS-IC C-terminal domain is loosely tethered to the anticodon-binding domain, which is loosely tethered to the aminoacylating domain. From the image, it appears that allosteric effects from $tRNA^{Tyr}$ anticodon and V arm 5′-stem contacts may be mostly communicated to the $tRNA^{Tyr}$ 3′-end via $tRNA^{Tyr}$ more than through the TyrRS-IC protein. The protein linkage of the C-terminal TyrRS-IC domain with the aminoacylating active site is very flexible, and the most relevant communication would be with the aminoacylating active site in the opposite α subunit.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0578.v1

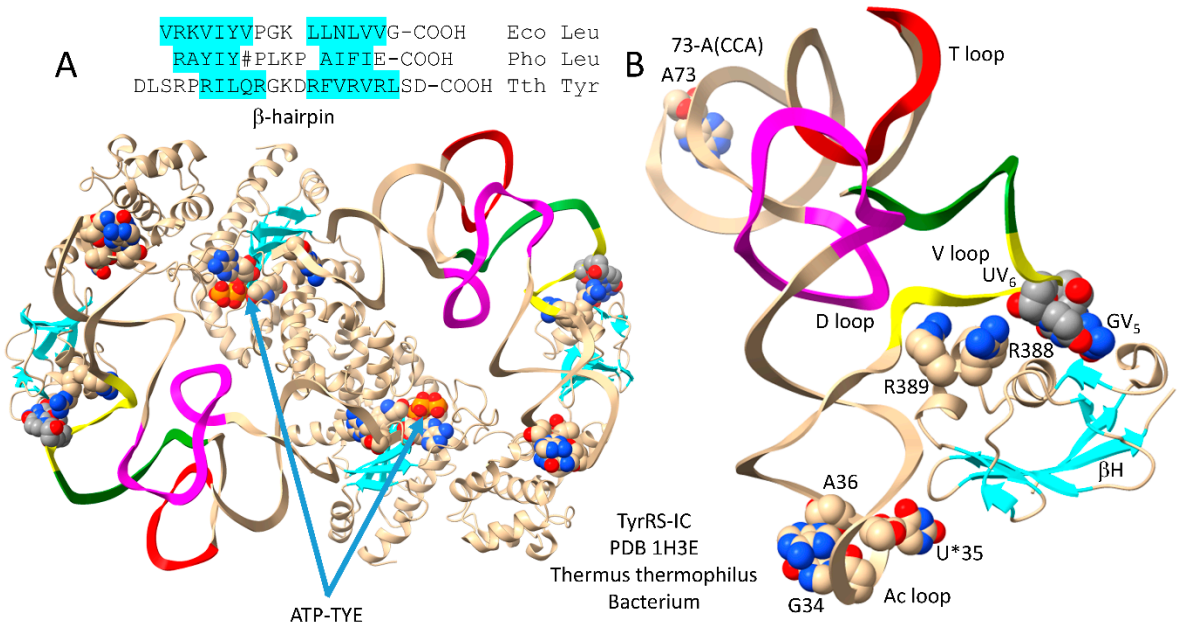


Figure 11. Bacterial TyrRS-IC bound to type II $tRNA^{Tyr}$ (GU*A) (U* for pseudouridine) [26]. A) The entire structure; B) the C-terminal TyrRS-IC domain bound to $tRNA^{Tyr}$ (GU*A). The aminoacylating active site binds ATP and TYE (a non-reactive tyrosine analogue). A β-hairpin at the Thermus thermophilus (Tth) TyrRS-IC C-terminus may relate to C-terminal β-hairpins in Escherichia coli (Eco) and Pyrococcus horikoshii (Pho) LeuRS-IA. # indicates an ~93 amino acid insert in the Pho C-terminal β-hairpin sequence. Colors are consistent with other figures.

Comparing the archaeal and bacterial TyrRS-IC-$tRNA^{Tyr}$ structures, archaeal TyrRS-IC lacks the C-terminal domain that binds the bacterial $tRNA^{Tyr}$ type II V arm. Because the archaeal TyrRS-IC enzyme recognizes a type I $tRNA^{Tyr}$, absence of the type II V arm-binding domain in archaeal TyrRS-IC is as expected. Either the bacterial TyrRS-IC C-terminal domain was a bacterial addition or an archaeal deletion. We posit, however, that the TyrRS-IC C-terminal domain in Bacteria may be distantly homologous to the LeuRS-IA C-terminus because of the possible similarities of the C-terminal β-hairpins. If this idea is correct, archaeal and bacterial TyrRS-IC may have diverged very early in evolution (i.e., at LUCA), and archaeal $tRNA^{Tyr}$ may have subsequently deleted the C-terminal domain. It appears to us that divergence of archaeal and bacterial TyrRS-IC may have occurred very early in evolution at the time that type I and type II tRNAs were first sorted. Because Thermus thermophilus $tRNA^{Tyr}$ has a 14 nt V loop, which is the primordial length, this observation is also consistent with early sorting and divergence of type I and type II tRNAs.

7. Type II V Loops in an Ancient Bacterium

Figure 12 compares type II V loops for Thermus thermophilus [4]. In Archaea, $tRNA^{Leu}$ and $tRNA^{Ser}$ are type II tRNAs with V arm set points of 2 and 1, respectively. In Bacteria, $tRNA^{Tyr}$, $tRNA^{Leu}$ and $tRNA^{Ser}$ are type II tRNAs with V arm trajectory set points of 2, 1 and 0, respectively. We consider T. thermophilus to be a reasonable reference organism for the earliest recognizable divergence of Archaea and Bacteria.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0578.v1

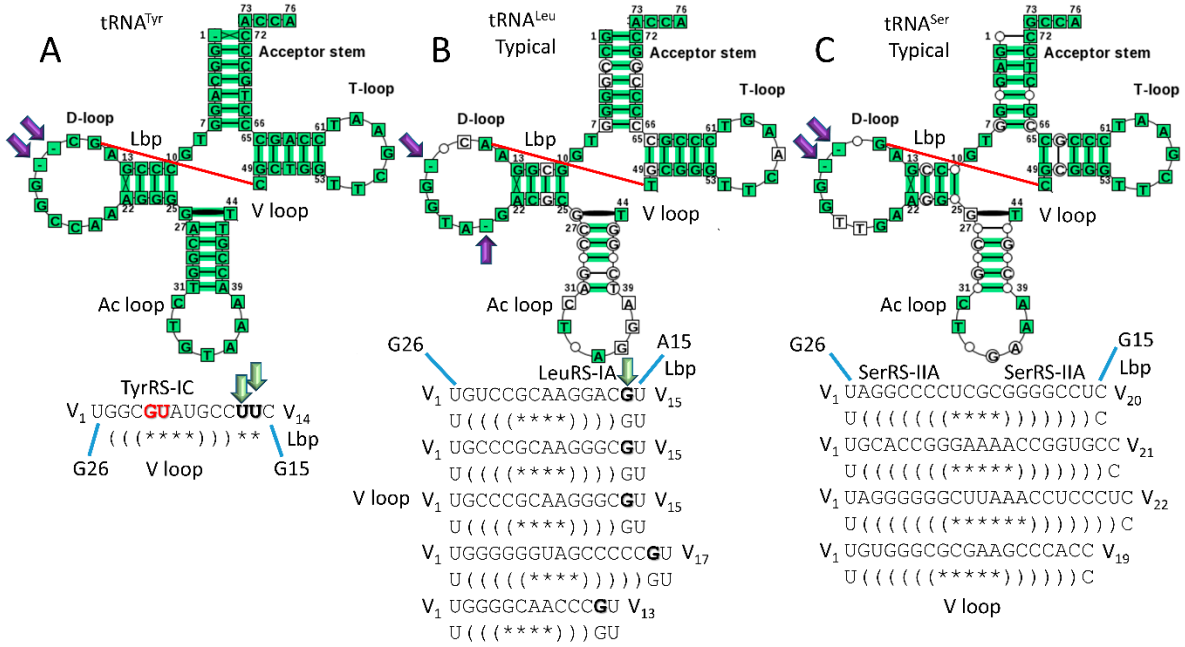


Figure 12. Type II V loops in Thermus thermophilus (typical tDNA sequences are shown) [4]. A) tRNA^{Tyr}; B) tRNA^{Leu}; and C) tRNA^{Ser}. Lbp indicates the Levitt base pair. Deleted bases in the D loop are indicated with purple arrows. Green arrows indicate unpaired bases just 5′ of the Levitt base that determine the trajectory set point of the V arm from the tRNA body (Figure 2). Notations are consistent with other figures. TyrRS-IC binds the V arm 5′-stem and V arm end loop bases V_5-GU-V_6 (red; Figure 11). LeuRS-IA binds the V arm 3′-stem but not the end loop (Figures 4 and 5). SerRS-IIA binds the elbow and the V arm 5′- and 3′-stems (Figures 7 and 8).

Figure 12A represents tRNA^{Tyr} (trajectory set point 2). Figure 12B represents a typical tRNA for tRNA^{Leu} (set point 1). Figure 12C represents a typical tRNA for tRNA^{Ser} (set point 0). Typical tRNAs represent a consensus tRNA with relaxed scoring [4]. The program used to generate the cloverleaf diagrams does not handle V loop sequences appropriately, so the relevant V loop sequences are shown individually. For all type II tRNAs in T. thermophilus, UV_1 forms a wobble pair with tRNA-G26, as in Archaea. Only TyrRS-IC interacts with V arm end loop bases (V_5-GU-V_6 in Tth). TyrRS-IC contacts are also made to the V arm 5′-stem (Figure 11). LeuRS-IA interacts with the V arm 3′-stem, making stronger contacts to the stem in the aminoacylating conformation compared to the editing/proofreading conformation (compare Figures 4 and 5). SerRS-IIA interacts with both the V arm 5′- and 3′-stems. In Bacteria, the tRNA^{Ser} V arm set point of 0 may correlate with the longer lengths of the tRNA^{Ser} V arm stems compared to Archaea. Longer tRNA^{Ser} stems in Bacteria may be necessary to maintain V_2-$V_{(n-1)}$ pairing. Also, the longer tRNA^{Ser} V arm stems may otherwise help accurately discriminate three type II tRNAs in Bacteria compared to two type II tRNAs in Archaea.

Because the tRNA^{Tyr} V loop is 14 nt, which is the primordial length, type II tRNA^{Tyr} in Bacteria may be as ancient as a time when all or most type II V loops were 14 nt in length. It appears to us that bacterial type II tRNA^{Tyr} may have evolved from an ancient type II tRNA^{Ser} before expansion of the tRNA^{Ser} V arm [8]. In T. thermophilus, tRNA^{Tyr} and tRNA^{Ser} are similar tRNAs. The deleted bases in the D loop are the same, the Levitt base pair is the same, and the T loops are the same. By contrast, tRNA^{Leu} has a different deleted base in the D loop, a different Levitt base pair and a different T loop base (tRNA-57A versus tRNA-57G). We posit that, in Bacteria, type II tRNA^{Tyr} may have evolved from an ancient tRNA^{Ser} with a 14 nt V loop.

8. Missing Data

Here, we have attempted to assemble a coherent story for evolution and divergence of type II V loops in Archaea and Bacteria based on evolution of type I and type II tRNAs. We have previously

 doi:10.20944/preprints202407.0578.v1

shown that tRNAomes for Archaea are simpler than for Bacteria and closer to LUCA [8]. This comparison holds for type II V loops. To improve the current analyses, additional data will be useful. Many AARS-tRNA structures should be done or modeled using native, modified tRNAs. Modeling can be done with modified tRNAs and reactive intermediates. Structures should be done using cryo-electron microscopy and natural tRNAs. Perhaps, structures could be modeled using AlphaFold 3 [36]. Also, molecular dynamics simulations and analyses of AARS-tRNA allostery should be done. We were surprised not to find more such papers. For instance, LeuRS-IA switching between aminoacylating and editing/proofreading modes could be analyzed by simulation. In Archaea, observing a LeuRS-IA-tRNA^{Leu} structure in the editing/proofreading conformation would be useful. In Archaea, observing a SerRS-IIA-tRNA^{Ser} structure (V arm trajectory set point of 1 in Archaea versus 0 in Bacteria) would add to the current discussion. In Bacteria, an Escherichia coli TyrRS-IC-tRNA^{Tyr} structure would be of interest. Molecular dynamics simulation of the ArgRS-IA-tRNA^{Arg} (ICG) structure (Figure 9) should provide insight into allostery involving distal AARS-tRNA determinants (anticodon loop and elbow).

9. Allostery

We posit that allosteric communication between distal tRNA contacts (determinants) and the tRNA 3′-end may be important for accurate aminoacylation of cognate tRNAs [17]. Several simulations have been done on AARS enzymes, but few appear to address the issue of allostery linking distal tRNA determinants (i.e., anticodon loop, elbow and V arm) and the tRNA 3′-end. A study of MetRS-IA-tRNA^{Met} indicates that most allosteric communication in this AARS is through the MetRS-IA protein not the tRNA^{Met} [37]. In MetRS-IA, however, the protein is well-connected, extending from the anticodon-binding to the tRNA^{Met} 3′-end. For some other AARS enzymes, distal cognate tRNA determinants appear to be loosely tethered through the protein to the aminoacylating active site and the tRNA 3′-end. For instances, see LeuRS-IA-tRNA^{Leu} (Figures 3–5) and TyrRS-IC-tRNA^{Tyr} (Figures 10 and 11). Most simulations to date appear to best describe events at the aminoacylating or editing/proofreading active sites [38–43]. We posit that, for some AARS, events at the tRNA 3′-end may be coupled to distal tRNA determinants primarily through the tRNA, acting in similar fashion to a coiled spring. We suggest this type of allosteric communication for LeuRS-IA-tRNA^{Leu} and TyrRS-IC-tRNA^{Tyr}. In such a case, events at the aminoacylating or editing active sites would be amplified by distal tRNA determinant contacts. For instance, the C-terminal domain of bacterial LeuRS-IA makes distinct tRNA^{Leu} V arm 3′-stem contacts in the aminoacylating and editing/proofreading conformations (compare Figures 4–6).

10. Divergence of Archaea and Bacteria

Evolution of type II tRNA V arms appears to relate a simple story about divergence of the archaeal and bacterial domains. We support the following model. From LUCA, Archaea and Bacteria diverged. For translation functions, Archaea are most similar to LUCA, and Bacteria are more distinct. Bacteria appear to have assumed their separate identity after significant isolation from Archaea. For instance, no intermediate organisms separating Archaea and Bacteria have been identified. Bacteria partly diverged because of their different transcription system. Bacteria rely on coevolution of sigma factors, bacterial promoters and a streamlined RNA polymerase [11,44,45]. For translation functions, Bacteria appear more diverged from LUCA than Archaea. Bacterial divergence is evident by inspection of tRNAomes and the genetic code. In many ways, Bacteria appear to be a more successful and innovated prokaryote compared to Archaea.

Table 1 summarizes scores and lengths of type II V loops in Pyrococcus furiosus (an ancient Archaeon) and Thermus thermophilus (an ancient Bacterium) [4]. In P. furiosus, tRNA^{Tyr} (1 tRNA^{Tyr}) is a type I tRNA (5 nt V loop). In P. furiosus, tRNA^{Leu} is a type II tRNA with a trajectory score of 2 and a length of 14 nt (5 tRNA^{Leu}), the primordial length. Also, tRNA^{Ser} has a trajectory score of 1 and a length of 15 nt (4 tRNA^{Ser}). In T. thermophilus, tRNA^{Tyr} has a score of 2 and a length of 14 nt, the primordial length (1 tRNA^{Tyr}). Also, tRNA^{Leu} has a score of 1 and lengths of 13-17 nt (5 tRNA^{Leu}), and tRNA^{Ser} has a score of 0 and lengths of 19-22 nt (4 tRNA^{Ser}). It appears to us that, within a domain,

 doi:10.20944/preprints202407.0578.v1

synonymous tRNAs with type II V loops must have different trajectory set point scores and, thus, distinct trajectories of V arms from the tRNA body (Figure 2). Archaeal type II V arms for $tRNA^{Leu}$ are generally shorter compared to bacterial V arms. In Archaea, $tRNA^{Tyr}$ sorted to become a type I tRNA (closely related to $tRNA^{Asn}$) [8]. Because Archaea only utilize type II tRNAs encoding leucine and serine, there was little pressure to lengthen V arms, and $tRNA^{Leu}$ and $tRNA^{Ser}$ generally maintained shorter V arms in Archaea than in Bacteria. Because type II $tRNA^{Tyr}$ (score of 2; V_n of 14 nt) was adopted in Bacteria, $tRNA^{Leu}$ was downgraded to a trajectory score of 1 (i.e., V_n of 13-17 nt in T. thermophilus), and $tRNA^{Ser}$ was downgraded to a score of 0 (V_n of 19-22 nt in T. thermophilus), relative to Archaea. In Bacteria, we posit that longer $tRNA^{Ser}$ V arm stems may have evolved to stabilize the V_2=$V_{(n-1)}$ pairing. Because of the different V arm set point in Archaea, V_2=$V_{(n-2)}$ pairing is utilized. In order to sort three type II tRNA amino acids in Bacteria, $tRNA^{Tyr}$ is generally 14 nt or shorter. 14 nt is the primordial length. Interestingly, the $tRNA^{Tyr}$ V loop is 13 nt in Escherichia coli and the V arm end loop sequence of V_5-GU-V_6 that binds TyrRS-IC in T. thermophilus (or V_5-UU-V_6 in some Bacteria) is not present. We posit that, in E. coli, contact to the V arm end loop is not utilized, and only contacts to the V arm 5′-stem are maintained for allosteric contacts. No structure is available currently to test this notion. AlphaFold 3 modeling could perhaps be used to address this issue [36].

Table 1. Type II tRNA V loops in Archaea (Pfu for Pyrococcus furiosus) and Bacteria (Tth for Thermus thermophilus). NA indicates "not applicable". The trajectory score is the number of unpaired bases in a type II V loop just 5′ of the Levitt base (V_n).

	Archaeon (Pfu)		Bacterium (Tth)	
	Score	**Length**	**Score**	**Length**
$tRNA^{Tyr}$	NA	(5)	2	14
$tRNA^{Leu}$	2	14	1	13-17
$tRNA^{Ser}$	1	15	0	19-22

Above, we have argued that the β-hairpin at the C-terminus of bacterial TyrRS-IC may relate distantly to the β-hairpins at the C-termini of LeuRS-IA (Figure 11). We agree that the sequence match is insufficient to fully demonstrate this idea. If this idea is correct, however, we would argue that evolution of type II $tRNA^{Tyr}$ and TyrRS-IC in Bacteria may have been as ancient an occurrence as the initial divergence of Archaea and Bacteria, perhaps as ancient as when all or most type II V arms were 14 nt in length. In such a scenario, Archaea could have adopted a type I $tRNA^{Tyr}$ and deleted the TyrRS-IC C-terminal domain that was necessary only to recognize a type II $tRNA^{Tyr}$. Bacteria would have adopted or maintained a type II $tRNA^{Tyr}$ and maintained a TyrRS-IC with a C-terminal domain capable of recognizing the type II V arm.

Table 2 summarizes type II V arm and elbow tRNA distal determinants for LeuRS-IA, SerRS-IIA and TyrRS-IC. Missing data for a full comparison are identified. For LeuRS-IA, Archaea and Bacteria have evolved homologous C-terminal domains that are massively modified and rearranged to make very different tRNA V arm and elbow contacts. In Bacteria, TyrRS-IC has a C-terminal domain that interacts with the V arm end loop and V arm 5′-stem in Thermus thermophilus and probably only the V arm 5′-stem in Escherichia coli (no structure is available). SerRS-IIA utilizes an N-terminal helix hairpin to bind the $tRNA^{Ser}$ V arm 5′- and 3′-stems and the elbow. LeuRS-IA and SerRS-IIA do not utilize cognate tRNA anticodon recognition, consistent with leucine and serine occupying 6 codon sectors of the genetic code. Serine is the only amino acid that is split between two genetic code columns (columns 2 and 4). As described below, we attempt to explain how SerRS-IIA-$tRNA^{Ser}$ recognition may have facilitated the jump. Among missing data are: 1) archaeal LeuRS-IA-$tRNA^{Leu}$ in an editing/proofreading conformation; 2) archaeal SerRS-IIA-$tRNA^{Ser}$; and 3) Escherichia coli TyrRS-IC-$tRNA^{Tyr}$. We do not know how tRNA elbow recognition affects allosteric communication to a cognate AARS active site(s).

 doi:10.20944/preprints202407.0578.v1

Table 2. Comparison summary of type II V loop and elbow tRNA allosteric contacts in Archaea and Bacteria. NA for not applicable. Pho for Pyrococcus horikoshii. Tth for Thermus thermophilus.

AARS-tRNA	Domain--conformation	Score	elbow	V loop
LeuRS-IA-tRNA^{Leu}	Archaea--Aminoacylating	2	93 aa insert	β-hairpin-V arm end loop V_6-UAG-V_8 (Pho)
LeuRS-IA-tRNA^{Leu}	Archaea--Editing	2	**no structure**	**no structure**
LeuRS-IA-tRNA^{Leu}	Bacteria--Aminoacylating	1	β-hairpin	V arm 3'-stem
LeuRS-IA-tRNA^{Leu}	Bacteria--Editing	1	β-hairpin	(V arm 3'-stem)—weakened contact
SerRS-IIA-tRNA^{Ser}	Archaea	1	**no structure**	**no structure**
SerRS-IIA-tRNA^{Ser}	Bacteria	0	helix hairpin	helix hairpin--5'- and 3'-V arm stems
TyrRS-IC-tRNA^{Tyr}	Archaea	NA	no contact	no contact
TyrRS-IC-tRNA^{Tyr}	Bacteria	2	no contact	β-hairpin domain, 5'-V arm stem, V arm end loop V_5-GU-V_6 (Tth)

11. Serine Jumping in Genetic Code Evolution

Serine is the only amino acid that is split between two genetic code columns (columns 2 and 4) (see below). We posit that serine jumping was possible because tRNA^{Ser} has a type II V loop and SerRS-IIA lacks tRNA^{Ser} anticodon loop recognition [14–16]. We further suggest that serine jumping during genetic code establishment may relate to incorporation of cysteine into the code. Serine can be converted to cysteine through tRNA-linked chemistry [46–48]. Because cysteine is important for Zn binding, there is reason to believe that cysteine was an early addition to the genetic code. First proteins that coevolved with the code may have utilized Zn-binding for their initial folding [49]. Cysteine, however, now occupies disfavored row 1 (tRNA-36A) of the genetic code, indicating that cysteine may have been a late addition. Row 1 (tRNA-36A) appears to be the last row of the code to fill. Complex aromatic amino acids (phenylalanine, tyrosine and tryptophan) and stop codons locate to row 1 of the code, indicating that row 1 filled late [50]. It has been posited that row 1 filled late because tRNA-36 was initially a wobble position [14–16]. Unmodified A is not observed in a wobble position in Archaea (no unmodified tRNA-34A). Suppression of wobbling at tRNA-36A involved a tightening conformation (closing) of the 30S ribosomal subunit [51–55] and modifications of tRNA-37 [13]. At the base of code evolution, it appears that tRNA-37m^1G was necessary to read tRNA-36A and tRNA-37t^6A was necessary to read tRNA-36U. Such observations are consistent with tRNA-36 having been a wobble position before wobbling could be suppressed.

We suggest that serine jumped from column 2 to column 4 from an enlarged serine block within the code (i.e., tRNA^{Ser} (GGU→GCU)). Cysteine, however, may have entered the genetic code by modification of serine (i.e., Ser-tRNA^{Cys}→Cys-tRNA^{Cys} through tRNA-linked chemistry) [46–48]. In this way, cysteine could have invaded the genetic code early but settled in its final position in the code late (tRNA^{Cys} (GCA)). Anticodon GGU now encodes threonine, which is chemically related to serine. We are suggesting that both serine conversion to cysteine by tRNA-linked chemistry and type II tRNA^{Ser} V arm recognition by SerRS-IIA may have enabled serine jumping from column 2 to column 4 of the genetic code. Serine jumping during code establishment is of interest because this is some of the only observed chaos in evolution of the code [14–16].

12. Type II tRNA Evolution and the Origin of the Genetic Code

The effort to understand type II tRNA diversification at the origin of life and during the great divergence at LUCA of Archaea and Bacteria is part of a larger effort to understand evolution of the genetic code [5,14–16,56,57]. Type I and type II tRNAs were sorted very early in evolution. It appears

 doi:10.20944/preprints202407.0578.v1

that type II $tRNA^{Leu}$ and $tRNA^{Ser}$ were sorted before divergence of Archaea and Bacteria. Type II $tRNA^{Tyr}$ was subsequently adopted in Bacteria but rejected in Archaea. The number of type II tRNAs in a prokaryote was limited by the number of potential trajectory set points of the V arm. Archaea adopted two set points. Bacteria adopted three. In Bacteria, having three type II V arm trajectory set points is correlated with expansion of $tRNA^{Ser}$ V arm stems. We posit that adoption of three trajectory set points for type II V arms in Bacteria: 1) resulted in lengthening of $tRNA^{Ser}$ V arm stems; 2) altered the set points of $tRNA^{Leu}$ and $tRNA^{Ser}$ V arms; 3) caused alterations in how the $tRNA^{Leu}$ V arm and elbow are utilized as determinants for LeuRS-IA recognition; and 4) contributed to divergence of Archaea and Bacteria.

Figure 12 shows an archaeal codon-anticodon table [14–16,49]. The complexity of the code is a maximum of 32 assignments. The table lists the encoded amino acid and its cognate AARS. Colors emphasize related amino acids and AARS enzymes that mostly align in columns. To suppress superwobbling and allow 2 codon sectors, U must be modified by methylation at the 5-carbon [13]. Most evolution is in code columns (tRNA-35). Columns 1, 2 and 4 contain 4 and 6 codon sectors. Column 3 is entirely 2 codon sectors. Rows 1, 2, 3 and 4 relate to tRNA-36. Only purine-pyrimidine discrimination is achieved at a wobble position (tRNA-34; A and B rows).

The genetic code is highly ordered. Significant evolution is observed in genetic code columns. Related amino acids Leu, Ile, Met and Val locate to column 1 of the code. LeuRS-IA, IleRS-IA, MetRS-IA and ValRS-IA are all closely homologous class IA AARS enzymes. Ser and Thr are chemically-related amino acids, and Ser, Pro, Thr and Ala are neutral amino acids that locate to column 2. SerRS-IIA, ProRS-IIA and ThrRS-IIA are closely homologous class IIA AARS enzymes. AlaRS-IID is a significantly different AARS, which may have replaced a now extinct AlaRS-IIA before LUCA to suppress translation errors. In Archaea, in column 3, an ordered striped pattern is observed. His, Asn and Asp occupy column 3, rows 2A, 3A and 4A (tRNA-34G). HisRS-IIA, AsnRS-IIB and AspRS-IIB are closely homologous AARS. Gln, Lys and Glu occupy column 3, rows 2B, 3B and 4B (tRNA-34U*/C; U* is modified U to suppress superwobbling; i.e., cnm^5U) [13]. GlnRS-IB, LysRS-IB (in Archaea) and GluRS-IB are closely homologous AARS. In column 4, CysRS-IA and ArgRS-IA are closely homologous class IA AARS. Glycine occupies the most favored sector in the genetic code: column 4 (tRNA-35C) and row 4 (tRNA-36C).

It appears that glycine was the first encoded amino acid (tRNA-35C, tRNA36C) [14–16,58,59]. Glycine, alanine, aspartic acid and valine (GADV) are the simplest amino acids that occupy the most favored row 4 (tRNA-36C). It appears that GADV were the first 4 encoded amino acids [60–65]. An adequate model for evolution of the genetic code must specify an order of addition of amino acids into the code. An adequate model for evolution of the code must account for evolution of 6 codon sectors (Leu, Ser and Arg), 4 codon sectors (Val, Pro, Thr, Ala and Gly), the 3 codon sector (Ile), 2 codon sectors (Phe, Tyr, His, Gln, Asn, Lys, Asp, Glu and Cys) and one codon sectors (Met and Trp). Leu and Ser occupy 6 codon sectors, and $tRNA^{Leu}$ and $tRNA^{Ser}$ are type II tRNAs that utilize their longer V arms for LeuRS-IA and SerRS-IIA recognition and discrimination. ArgRS-IA, which utilizes a type I $tRNA^{Arg}$, unwinds the anticodon loop to better recognize this feature (Figure 9) [34]. An adequate model for code evolution must rationalize why leucine and serine utilize type II tRNAs and occupy 6 codon sectors. An adequate model must rationalize serine jumping between column 2 and column 4 of the genetic code.

The genetic code evolved around tRNA and the tRNA anticodon [14–16,49]. Degeneracy explains why the genetic code encodes 21 assignments: 20 amino acids plus stops. The genetic code has the capacity to encode up to 32 assignments (Figure 13). At a wobble position (tRNA-34), only purine versus pyrimidine resolution has been achieved. At Watson-Crick positions (tRNA-35 and tRNA-36), codon A, G, C and U can be read. Thus, the code was limited by tRNA reading to 2x4x4=32 assignments. But, tRNA-34 (wobble) and tRNA-36 positions show similarities for utilization of weakly pairing bases U and A. At tRNA-34, tRNA-34U must be modified to suppress "superwobbling" [13]. Superwobbling, in which tRNA-34U reads mRNA wobble A, G, C and U (as in mitochondria), can only be utilized in a 4 codon box [66,67]. To support 2 codon sectors, tRNA-34U must be modified (i.e., tRNA-$34cnm^5U$; 5-cyanomethyluridine). In Archaea, tRNA-34A is not

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 July 2024 doi:10.20944/preprints202407.0578.v1

observed. At tRNA-36, at the base of the code, tRNA-36A is supported by adjacent tRNA-37m^{1}G modification. Also, tRNA-36U is supported by adjacent tRNA-37t^{6}A modification. We posit that tRNA-34 and tRNA-36 were originally both wobble positions, and only a single wobble position could be read at a time. With both tRNA-34 and tRNA-36 as wobble positions, the complexity of the genetic code was 8 assignments (2x4 or 4x2) (only one wobble position could be read at a time). When tRNA-36 was a wobble position, we posit that columns 1, 2 and 4 of the genetic code evolved primarily around tRNA-35 (Watson-Crick) and tRNA-36 (wobble). Column 3 of the genetic code evolved around tRNA-34 (wobble) and tRNA-35 (Watson-Crick). This explains why 6, 4 and 3 codon boxes locate to columns 1, 2 and 4. This also explains why only 2 codon boxes are found in column 3, and why column 3 fractionates on A and B rows (tRNA-34; wobble).

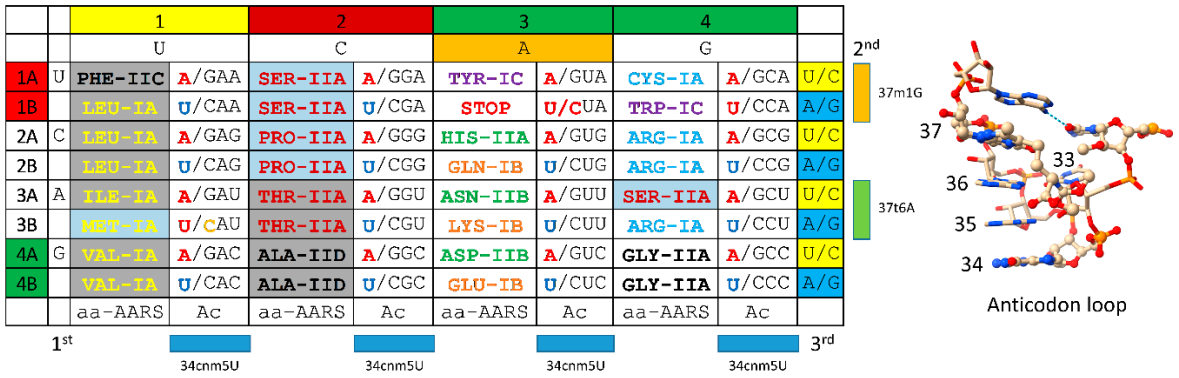


Figure 13. The archaeal genetic code as a codon-anticodon table (maximum complexity 32 assignments). aa-AARS) amino acid-aminoacyl-tRNA synthetase (class I or II and subclass A, B, C or D); Ac) anticodon. 1st, 2nd, 3rd) codon. Colors were selected to emphasize similar amino acids and AARS enzymes. Red bases in the anticodon are not utilized. Blue U indicates a modified tRNA-34cnm^{5}U or similar modification. Orange C indicates differential anticodon C*AU modification. In Archaea, C* is agmatidine to encode isoleucine. C* is methylated to encode elongator methionine, and C is unmodified for initiator methionine. Gray shading indicates AARS enzymes with separate editing/proofreading active sites. Light blue shading indicates AARS enzymes with editing functions in the aminoacylating active site. 37m^{1}G evolved to read tRNA-36A. 37t^{6}A evolved to read tRNA-36U. To the right of the figure, the anticodon loop is shown. Bases indicated in ball and stick representation are modified bases [20].

Because tRNA-36 was originally a wobble position, we posit that disfavored row 1 of the genetic code (tRNA-36A) was the last row to fill. Stop codons locate to disfavored row 1. Stop codons are read by protein release factors that read the mRNA codon directly, so there is no tRNA that corresponds to a stop codon (except in suppressor strains) [68]. Aromatic amino acids Phe, Tyr and Trp locate to row 1. Phe, Tyr and Trp are the most complex amino acids, so it is reasonable that they were added late after wobbling was suppressed at tRNA-36 [50]. Suppression of wobbling at tRNA-36 was partly via tRNA-37m^{1}G modification to read tRNA-36A and tRNA-37t^{6}A modification to read tRNA-36U. Also, to suppress tRNA-36 wobbling, a conformational change of the 30S ribosomal subunit tightens the anticodon-codon interaction, dehydrates the base pairs and locks the translation frame that helps to maintain translational fidelity and, also, helps to set the reading frame in place [51–55,69–71]. Wobbling cannot be suppressed in the same manner at tRNA-34 because modification of adjacent bases does not assist tRNA-34 reading. Modification of tRNA-33 will not help to suppress tRNA-34 wobbling because tRNA-33 is on the other side of the anticodon loop U-turn. Also, tRNA-35 cannot be easily modified because this is a Watson-Crick base that must pair with mRNA. Too many different and constrained tRNA-35 modifications might be necessary (i.e., 2-4) to suppress wobbling at tRNA-34 for such a mechanism to evolve.

In the pre-life world, RNA-linked and tRNA-linked chemistry were common. In evolution of the genetic code, tRNA-linked reactions may have promoted incorporation of leucine (Val→Leu; 5 steps), tyrosine (Phe→Tyr; 1 step), glutamine (Glu→Gln; 1 step), asparagine (Asp→Asn; 1 step) [72,73],

 doi:10.20944/preprints202407.0578.v1

arginine (Orn→Arg; 2 steps; Orn for ornithine) [74] and cysteine (Ser→Cys; 2 steps) [46–48]. Because of tRNA-linked reactions, an 8 amino acid genetic code can be significantly enriched to encode the first RNA sequence-dependent proteins. For instance, in addition to encoding GADVLSER, an 8 aa code with tRNA-34 and tRNA-36 wobbling could also utilize CQN through tRNA-linked reactions.

Adopting a tRNA-centric view of pre-life chemical evolution indicates that the genetic code was initially utilized to generate polyglycine, a component of protocells [14–16]. Subsequently, the code progressed to generate GADV polymers. Then, probably, GADVLSER was encoded with CQN added through tRNA-linked chemistry. Leucine and serine, therefore, may have entered the code at about the same time to utilize type II tRNAs and to eventually settle into 6 codon boxes. First proteins emerged at about the 11 amino acid stage. Suppression of tRNA-36 wobbling allowed the code to expand. The code froze at 20 amino acids plus stops because of fidelity mechanisms.

We consider the utilization of type II $tRNA^{Leu}$, $tRNA^{Ser}$ and $tRNA^{Tyr}$ (in Bacteria) to support this narrative.

13. Conclusions

We conclude that type II V loops in Archaea and Bacteria relate a simple story about the original sorting of type I and type II tRNAs. In a domain (i.e., Archaea and Bacteria), type II tRNA V loops must have distinct trajectory set points determined by the number of unpaired bases just 5′ of the Levitt base (V_n). For Archaea, $tRNA^{Leu}$ has a set point of 2, and $tRNA^{Ser}$ has a set point of 1. For Bacteria, $tRNA^{Tyr}$ has a set point of 2. To accommodate a type II $tRNA^{Tyr}$ with a set point of 2 in Bacteria, $tRNA^{Leu}$ has a set point of 1, and $tRNA^{Ser}$ has a set point of 0. The longer lengths of the $tRNA^{Ser}$ V arm stems in Bacteria may relate to the need to stabilize the V_2=$V_{(n-1)}$ base pair. Sharing tRNAs between Archaea and Bacteria is awkward, in part, because of the incompatibility of type II tRNAs. Also, tRNA modification systems in Archaea and Bacteria are largely incompatible.

As organisms became more derived in evolution, V arm end loop contacts by a cognate AARS appear to have given way to V arm stem contacts. This trend appears to be supported by LeuRS-IA-$tRNA^{Leu}$ V arm end loop contacts in Archaea being replaced by LeuRS-IA-$tRNA^{Leu}$ V arm 3′ stem contacts in Bacteria. Also, V arm end loop and 5′-stem contacts in TyrRS-IC-$tRNA^{Tyr}$ of Thermus thermophilus appear to give way to V arm 5′-stem contacts in TyrRS-IC-$tRNA^{Tyr}$ of Escherichia coli (no structure is currently available). We posit that V arm stem contacts may exert greater allosteric communication to the tRNA 3′ end compared to V arm end loop contacts, which would be expected to be more flexible and more weakly allosteric than contacts to a V arm stem.

The 3 31 nt minihelix tRNA evolution theorem completely describes the evolution of type I and type II tRNAs, to the last nucleotide [5]. The reason that tRNA evolution could be solved with such high confidence is that tRNA sequences chemically evolved from RNA repeats and inverted repeats conserved from pre-life. Solution of type II tRNA evolution was predicted based on the model for type I tRNA evolution, making the 3 31 nt minihelix tRNA evolution theorem powerfully predictive [7]. Evolution of type II tRNAs in Archaea and Bacteria is fully supportive of the theorem. Evolution of type I and type II tRNAs forms the core, successful and conserved strategy and pathway in evolution of life on Earth. After ~4 billion years, it is remarkable that such a clear record of pre-life worlds survived in the tRNA sequences of living organisms.

Author Contributions: L. L. and Z.F.B. wrote the paper and prepared the figures. The authors have read and approved the final manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. Chan, P.P.; Lowe, T.M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods Mol Biol* **2019**, *1962*, 1-14. https://doi.org/10.1007/978-1-4939-9173-0_1.
2. Lowe, T.M.; Chan, P.P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* **2016**, *44*, W54-57. https://doi.org/10.1093/nar/gkw413.
3. Chan, P.P.; Lowe, T.M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **2016**, *44*, D184-189. https://doi.org/10.1093/nar/gkv1309.
4. 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. https://doi.org/10.1093/nar/gkn772.
5. Lei, L.; Burton, Z.F. The 3 31 Nucleotide Minihelix tRNA Evolution Theorem and the Origin of Life. *Life (Basel)* **2023**, *13*. https://doi.org/10.3390/life13112224.
6. Burton, Z.F. The 3-Minihelix tRNA Evolution Theorem. *J Mol Evol* **2020**, *88*, 234-242. https://doi.org/10.1007/s00239-020-09928-2.
7. 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*. https://doi.org/10.3390/ijms19103275.
8. Pak, D.; Du, N.; Kim, Y.; Sun, Y.; Burton, Z.F. Rooted tRNAomes and evolution of the genetic code. *Transcription* **2018**, *9*, 137-151. https://doi.org/10.1080/21541264.2018.1429837.
9. 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. https://doi.org/10.1080/21541264.2017.1318811.
10. Root-Bernstein, R.; Kim, Y.; Sanjay, A.; Burton, Z.F. tRNA evolution from the proto-tRNA minihelix world. *Transcription* **2016**, *7*, 153-163. https://doi.org/10.1080/21541264.2016.1235527.
11. Lei, L.; Burton, Z.F. Early Evolution of Transcription Systems and Divergence of Archaea and Bacteria. *Front Mol Biosci* **2021**, *8*, 651134. https://doi.org/10.3389/fmolb.2021.651134.
12. Spellman, C.D.; Burton, Z.T.; Ikuma, K.; Strosnider, W.H.J.; Tasker, T.L.; Roman, B.; Goodwill, J.E. Continuous co-treatment of mine drainage with municipal wastewater. *J Environ Manage* **2024**, *354*, 120282. https://doi.org/10.1016/j.jenvman.2024.120282.
13. Lei, L.; Burton, Z.F. "Superwobbling" and tRNA-34 Wobble and tRNA-37 Anticodon Loop Modifications in Evolution and Devolution of the Genetic Code. *Life (Basel)* **2022**, *12*. https://doi.org/10.3390/life12020252.
14. Lei, L.; Burton, Z.F. Evolution of the genetic code. *Transcription* **2021**, *12*, 28-53. https://doi.org/10.1080/21541264.2021.1927652.
15. Lei, L.; Burton, Z.F. Evolution of Life on Earth: tRNA, Aminoacyl-tRNA Synthetases and the Genetic Code. *Life (Basel)* **2020**, *10*. https://doi.org/10.3390/life10030021.
16. 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*. https://doi.org/10.3390/life9020037.
17. Li, R.; Macnamara, L.M.; Leuchter, J.D.; Alexander, R.W.; Cho, S.S. MD Simulations of tRNA and Aminoacyl-tRNA Synthetases: Dynamics, Folding, Binding, and Allostery. *Int J Mol Sci* **2015**, *16*, 15872-15902. https://doi.org/10.3390/ijms160715872.
18. Giege, R.; Springer, M. Aminoacyl-tRNA Synthetases in the Bacterial World. *EcoSal Plus* **2016**, *7*. https://doi.org/10.1128/ecosalplus.ESP-0002-2016.
19. Chan, P.P.; Lowe, T.M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* **2009**, *37*, D93-97. https://doi.org/10.1093/nar/gkn787.
20. 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. https://doi.org/10.1017/s1355838200000364.
21. Fukunaga, R.; Yokoyama, S. Aminoacylation complex structures of leucyl-tRNA synthetase and tRNALeu reveal two modes of discriminator-base recognition. *Nat Struct Mol Biol* **2005**, *12*, 915-922. https://doi.org/10.1038/nsmb985.
22. Meng, E.C.; Goddard, T.D.; Pettersen, E.F.; Couch, G.S.; Pearson, Z.J.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* **2023**, *32*, e4792. https://doi.org/10.1002/pro.4792.
23. 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. https://doi.org/10.1002/pro.3943.
24. Goddard, T.D.; Huang, C.C.; Meng, E.C.; Pettersen, E.F.; Couch, G.S.; Morris, J.H.; Ferrin, T.E. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **2018**, *27*, 14-25. https://doi.org/10.1002/pro.3235.
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. https://doi.org/10.1016/j.jmb.2004.11.060.
26. Yaremchuk, A.; Kriklivyi, I.; Tukalo, M.; Cusack, S. Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J* **2002**, *21*, 3829-3840. https://doi.org/10.1093/emboj/cdf373.

27. Palencia, A.; Crepin, T.; Vu, M.T.; Lincecum, T.L., Jr.; Martinis, S.A.; Cusack, S. Structural dynamics of the aminoacylation and proofreading functional cycle of bacterial leucyl-tRNA synthetase. *Nat Struct Mol Biol* **2012**, *19*, 677-684. https://doi.org/10.1038/nsmb.2317.
28. 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. https://doi.org/10.1016/j.jmb.2017.10.024.
29. 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. https://doi.org/10.1126/science.8128220.
30. 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. https://doi.org/10.1093/nar/gkv996.
31. Belrhali, H.; Yaremchuk, A.; Tukalo, M.; Berthet-Colominas, C.; Rasmussen, B.; Bosecke, P.; Diat, O.; Cusack, S. The structural basis for seryl-adenylate and Ap4A synthesis by seryl-tRNA synthetase. *Structure* **1995**, *3*, 341-352. https://doi.org/10.1016/s0969-2126(01)00166-6.
32. Rafels-Ybern, A.; Torres, A.G.; Camacho, N.; Herencia-Ropero, A.; Roura Frigole, H.; Wulff, T.F.; Raboteg, M.; Bordons, A.; Grau-Bove, X.; Ruiz-Trillo, I., et al. The Expansion of Inosine at the Wobble Position of tRNAs, and Its Role in the Evolution of Proteomes. *Mol Biol Evol* **2019**, *36*, 650-662. https://doi.org/10.1093/molbev/msy245.
33. Rafels-Ybern, A.; Torres, A.G.; Grau-Bove, X.; Ruiz-Trillo, I.; Ribas de Pouplana, L. Codon adaptation to tRNAs with Inosine modification at position 34 is widespread among Eukaryotes and present in two Bacterial phyla. *RNA Biol* **2018**, *15*, 500-507. https://doi.org/10.1080/15476286.2017.1358348.
34. Delagoutte, B.; Moras, D.; Cavarelli, J. tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *EMBO J* **2000**, *19*, 5599-5610. https://doi.org/10.1093/emboj/19.21.5599.
35. Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat Struct Biol* **2003**, *10*, 425-432. https://doi.org/10.1038/nsb934.
36. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J., et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, 10.1038/s41586-024-07487-w. https://doi.org/10.1038/s41586-024-07487-w.
37. Ghosh, A.; Vishveshwara, S. A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc Natl Acad Sci U S A* **2007**, *104*, 15711-15716. https://doi.org/10.1073/pnas.0704459104.
38. Chowdhury, S.; Nandi, N. Dynamics of the Catalytic Active Site of Isoleucyl tRNA Synthetase from Staphylococcus aureus bound with Adenylate and Mupirocin. *J Phys Chem B* **2022**, *126*, 620-633. https://doi.org/10.1021/acs.jpcb.1c08555.
39. Saha, A.; Nandi, N. Role of the transfer ribonucleic acid (tRNA) bound magnesium ions in the charging step of aminoacylation reaction in the glutamyl tRNA synthetase and the seryl tRNA synthetase bound with cognate tRNA. *J Biomol Struct Dyn* **2022**, *40*, 8538-8559. https://doi.org/10.1080/07391102.2021.1914732.
40. Dutta, S.; Nandi, N. Classical molecular dynamics simulation of seryl tRNA synthetase and threonyl tRNA synthetase bound with tRNA and aminoacyl adenylate. *J Biomol Struct Dyn* **2019**, *37*, 336-358. https://doi.org/10.1080/07391102.2018.1426498.
41. Dutta, S.; Kundu, S.; Saha, A.; Nandi, N. Dynamics of the active site loops in catalyzing aminoacylation reaction in seryl and histidyl tRNA synthetases. *J Biomol Struct Dyn* **2018**, *36*, 878-892. https://doi.org/10.1080/07391102.2017.1301272.
42. Torres-Larios, A.; Sankaranarayanan, R.; Rees, B.; Dock-Bregeon, A.C.; Moras, D. Conformational movements and cooperativity upon amino acid, ATP and tRNA binding in threonyl-tRNA synthetase. *J Mol Biol* **2003**, *331*, 201-211. https://doi.org/10.1016/s0022-2836(03)00719-8.
43. Rayevsky, A.V.; Sharifi, M.; Tukalo, M.A. Molecular modeling and molecular dynamics simulation study of archaeal leucyl-tRNA synthetase in complex with different mischarged tRNA in editing conformation. *J Mol Graph Model* **2017**, *76*, 289-295. https://doi.org/10.1016/j.jmgm.2017.06.022.
44. Burton, Z.F.; Opron, K.; Wei, G.; Geiger, J.H. A model for genesis of transcription systems. *Transcription* **2016**, *7*, 1-13. https://doi.org/10.1080/21541264.2015.1128518.
45. Burton, S.P.; Burton, Z.F. The sigma enigma: bacterial sigma factors, archaeal TFB and eukaryotic TFIIB are homologs. *Transcription* **2014**, *5*, e967599. https://doi.org/10.4161/21541264.2014.967599.
46. Mukai, T.; Crnkovic, A.; Umehara, T.; Ivanova, N.N.; Kyrpides, N.C.; Soll, D. RNA-Dependent Cysteine Biosynthesis in Bacteria and Archaea. *mBio* **2017**, *8*. https://doi.org/10.1128/mBio.00561-17.
47. Turanov, A.A.; Xu, X.M.; Carlson, B.A.; Yoo, M.H.; Gladyshev, V.N.; Hatfield, D.L. Biosynthesis of selenocysteine, the 21st amino acid in the genetic code, and a novel pathway for cysteine biosynthesis. *Adv Nutr* **2011**, *2*, 122-128. https://doi.org/10.3945/an.110.000265.

48. Hauenstein, S.I.; Perona, J.J. Redundant synthesis of cysteinyl-tRNACys in Methanosarcina mazei. *J Biol Chem* **2008**, *283*, 22007-22017. https://doi.org/10.1074/jbc.M801839200.
49. Pak, D.; Kim, Y.; Burton, Z.F. Aminoacyl-tRNA synthetase evolution and sectoring of the genetic code. *Transcription* **2018**, *9*, 205-224. https://doi.org/10.1080/21541264.2018.1467718.
50. Fournier, G.P.; Alm, E.J. Ancestral Reconstruction of a Pre-LUCA Aminoacyl-tRNA Synthetase Ancestor Supports the Late Addition of Trp to the Genetic Code. *J Mol Evol* **2015**, *80*, 171-185. https://doi.org/10.1007/s00239-015-9672-1.
51. Rozov, A.; Wolff, P.; Grosjean, H.; Yusupov, M.; Yusupova, G.; Westhof, E. Tautomeric G*U pairs within the molecular ribosomal grip and fidelity of decoding in bacteria. *Nucleic Acids Res* **2018**, *46*, 7425-7435. https://doi.org/10.1093/nar/gky547.
52. Rozov, A.; Demeshkina, N.; Westhof, E.; Yusupov, M.; Yusupova, G. New Structural Insights into Translational Miscoding. *Trends Biochem Sci* **2016**, *41*, 798-814. https://doi.org/10.1016/j.tibs.2016.06.001.
53. Rozov, A.; Westhof, E.; Yusupov, M.; Yusupova, G. The ribosome prohibits the G*U wobble geometry at the first position of the codon-anticodon helix. *Nucleic Acids Res* **2016**, *44*, 6434-6441. https://doi.org/10.1093/nar/gkw431.
54. Rozov, A.; Demeshkina, N.; Westhof, E.; Yusupov, M.; Yusupova, G. Structural insights into the translational infidelity mechanism. *Nat Commun* **2015**, *6*, 7251. https://doi.org/10.1038/ncomms8251.
55. Loveland, A.B.; Demo, G.; Grigorieff, N.; Korostelev, A.A. Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* **2017**, *546*, 113-117. https://doi.org/10.1038/nature22397.
56. Koonin, E.V.; Novozhilov, A.S. Origin and Evolution of the Universal Genetic Code. *Annu Rev Genet* **2017**, *51*, 45-62. https://doi.org/10.1146/annurev-genet-120116-024713.
57. Koonin, E.V.; Novozhilov, A.S. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life* **2009**, *61*, 99-111. https://doi.org/10.1002/iub.146.
58. Bernhardt, H.S.; Patrick, W.M. Genetic code evolution started with the incorporation of glycine, followed by other small hydrophilic amino acids. *J Mol Evol* **2014**, *78*, 307-309. https://doi.org/10.1007/s00239-014-9627-y.
59. Bernhardt, H.S.; Tate, W.P. Evidence from glycine transfer RNA of a frozen accident at the dawn of the genetic code. *Biol Direct* **2008**, *3*, 53. https://doi.org/10.1186/1745-6150-3-53.
60. Ikehara, K. Why Were [GADV]-amino Acids and GNC Codons Selected and How Was GNC Primeval Genetic Code Established? *Genes (Basel)* **2023**, *14*. https://doi.org/10.3390/genes14020375.
61. Ikehara, K. Evolutionary Steps in the Emergence of Life Deduced from the Bottom-Up Approach and GADV Hypothesis (Top-Down Approach). *Life (Basel)* **2016**, *6*. https://doi.org/10.3390/life6010006.
62. Ikehara, K. [GADV]-protein world hypothesis on the origin of life. *Orig Life Evol Biosph* **2014**, *44*, 299-302. https://doi.org/10.1007/s11084-014-9383-4.
63. Ikehara, K. Pseudo-replication of [GADV]-proteins and origin of life. *Int J Mol Sci* **2009**, *10*, 1525-1537. https://doi.org/10.3390/ijms10041525.
64. Oba, T.; Fukushima, J.; Maruyama, M.; Iwamoto, R.; Ikehara, K. Catalytic activities of [GADV]-peptides. Formation and establishment of [GADV]-protein world for the emergence of life. *Orig Life Evol Biosph* **2005**, *35*, 447-460. https://doi.org/10.1007/s11084-005-3519-5.
65. Ikehara, K. Possible steps to the emergence of life: the [GADV]-protein world hypothesis. *Chem Rec* **2005**, *5*, 107-118. https://doi.org/10.1002/tcr.20037.
66. Alkatib, S.; Scharff, L.B.; Rogalski, M.; Fleischmann, T.T.; Matthes, A.; Seeger, S.; Schottler, M.A.; Ruf, S.; Bock, R. The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet* **2012**, *8*, e1003076. https://doi.org/10.1371/journal.pgen.1003076.
67. Rogalski, M.; Karcher, D.; Bock, R. Superwobbling facilitates translation with reduced tRNA sets. *Nat Struct Mol Biol* **2008**, *15*, 192-198. https://doi.org/10.1038/nsmb.1370.
68. Burroughs, A.M.; Aravind, L. The Origin and Evolution of Release Factors: Implications for Translation Termination, Ribosome Rescue, and Quality Control Pathways. *Int J Mol Sci* **2019**, *20*. https://doi.org/10.3390/ijms20081981.
69. Demeshkina, N.; Jenner, L.; Westhof, E.; Yusupov, M.; Yusupova, G. New structural insights into the decoding mechanism: translation infidelity via a G.U pair with Watson-Crick geometry. *FEBS Lett* **2013**, *587*, 1848-1857. https://doi.org/10.1016/j.febslet.2013.05.009.
70. Demeshkina, N.; Jenner, L.; Westhof, E.; Yusupov, M.; Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **2012**, *484*, 256-259. https://doi.org/10.1038/nature10913.
71. Loveland, A.B.; Demo, G.; Korostelev, A.A. Cryo-EM of elongating ribosome with EF-Tu*GTP elucidates tRNA proofreading. *Nature* **2020**, *584*, 640-645. https://doi.org/10.1038/s41586-020-2447-x.
72. Min, B.; Pelaschier, J.T.; Graham, D.E.; Tumbula-Hansen, D.; Soll, D. Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation. *Proc Natl Acad Sci U S A* **2002**, *99*, 2678-2683. https://doi.org/10.1073/pnas.012027399.

 doi:10.20944/preprints202407.0578.v1

73. Salazar, J.C.; Zuniga, R.; Raczniak, G.; Becker, H.; Soll, D.; Orellana, O. A dual-specific Glu-tRNA(Gln) and Asp-tRNA(Asn) amidotransferase is involved in decoding glutamine and asparagine codons in Acidithiobacillus ferrooxidans. *FEBS Lett* **2001**, *500*, 129-131. https://doi.org/10.1016/s0014-5793(01)02600-x.
74. Longo, L.M.; Despotovic, D.; Weil-Ktorza, O.; Walker, M.J.; Jablonska, J.; Fridmann-Sirkis, Y.; Varani, G.; Metanis, N.; Tawfik, D.S. Primordial emergence of a nucleic acid-binding protein via phase separation and statistical ornithine-to-arginine conversion. *Proc Natl Acad Sci U S A* **2020**, *117*, 15731-15739. https://doi.org/10.1073/pnas.2001989117.