

Article

Not peer-reviewed version

Nature and Nurture in Archaeal Synthetases and their Effects on tRNA Aminoacylation

[Alan M Laibelman](#) *

Posted Date: 2 June 2025

doi: 10.20944/preprints202506.0050.v1

Keywords: archaea; tRNA synthetases; comparative genomics; aminoacylation mechanism



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

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

Article

Nature and Nurture in Archaeal Synthetases and Their Effects on tRNA Aminoacylation

Alan M. Laibelman

Independent Researcher, USA; laibelman@yahoo.com

Abstract: Genes affect lifestyle preferences affect genes affect lifestyle preferences; the cycle never stops. For no lifeform is this more dramatically demonstrated than Archaea, the majority of whom are extreme in their preference for heat or cold, acidity or alkalinity, degree of salinity, presence or absence of oxygen. This comparative genomic study investigated encoded aminoacyl-tRNA synthetase primary sequences for 150 species. In conjunction with a previous inquiry about tRNA sequences, implications arise for mechanisms of aminoacylation for each tRNA type by its cognate amino acid with synthetase as catalyst. One strong conclusion emerges: there is a severe shortage of kinetic data available for these organisms. Regardless of theoretical developments for bacterial systems, on which much effort has been expended over time, they cannot be reliably extrapolated into Archaea to produce a comprehensive understanding of the process of translation in unicellular creatures.

Keywords: archaea; tRNA synthetases; comparative genomics; aminoacylation mechanism

Introduction

Scientists attempt to establish order out of apparent randomness, so they obsess over patterns; as a conglomerate of exploratory disciplines, science is grounded on the inherency of the preconception that, if one only seeks, systematization will reveal itself with respect to the natural universe. So abhorrent is chaos to a scientist that he/she sometimes tends to overcompensate by over-generalizing observational data. One field of inquiry in which this unfortunate trend exists is genomics. It is sometimes difficult to accept that genetics, the parent to genomics, is indisputably, unavoidably individualistic, with uniqueness at its core. A dynamic interplay of nature and nurture as inherent factors in the unceasing evolutionary development of lifeforms is not limited to sociological theory and practice; it is foundational for genetics. Students of genomics concentrate on the nature portion described by gene content and activity. They leave ecologists to explore nurture: facts and principles stipulating conditions permitting species to thrive in particular environments. It is only in combination that a coherent biology of living things is possible.

The bridge between nature and nurture can be perceived in attempts to establish a taxonomic tree of life. Taxonomic classification cannot achieve self-consistent separation into distinct groupings because traits upon which a sorting procedure is based will be found to cross organismal communities. Grounding on DNA (RNA) composition or protein functionality or physical cellular characteristics or environmental living conditions will inevitably result in finding exceptions to every proposed guiding rule. Phylogenetic trees are both source and result of taxonomic order, meaning their utility is inextricably tied to inalienable individuality found in lifeforms, which is why trees utilizing property A yield different nodal connections to those based on property B. The only rigorously self-consistent taxonomical hierarchy would place each species in a unique genus/family/order/class/phylum, and subdivide that species further into the number of subspecies equal to the population of that lifeform. The sole alternative to intrinsic uniqueness is cloning, and that procedure still fails to address issues of nurture and consequent behavior causing, for example, a set of identical human twins or triplets to be readily differentiable among those familiar with them.

To illustrate: human beings are superficially bilaterally symmetric in eyes, ears, arms, hands, legs, feet. In reality, dominance characteristics exist in strength, acuity, preferred usage. When it was confirmed that some people are preferentially righthanded while others are lefthanded, or that a distinction exists in visual and aural acuity between left and right sets of eyes and ears, *Homo sapiens* were not subsequently reclassified into subspecies *H. sapiens left* and *H. sapiens right*, although it would have been scientifically accurate. This nondecision reflected a choice to ignore complexity in relationship between genes in theory and genes in practice. This lack of taxonomic reclassification will undoubtedly be maintained throughout the animal and plant kingdoms despite mounting evidence that many species exhibit *limb chirality* or *bilateral asymmetry* even in those without neural networks (Rogers, 2023; Guerra et al., 2024).

Eukaryl taxonomy constitutes a stable system: other than addition of newly discovered organisms occurring rarely in multicellular animals and more frequently in plants, once genera and species names are assigned, reclassification does not transpire upon accumulations of new information regarding their genetics or behavior. In contrast, archaeal taxonomy depicts an unstable system. Not only do organisms undergo name change (e.g., *Methanotherx harundinacea* is *Methanocrinis harundinaceus*), but high-level categories such as phyla are renamed and sorted differently than they were previously.

Division of Archaea into Crenarchaeota and Euryarchaeota was originally based on 16S rRNA sequencing (Woese et al., 1990), as was the original separation of newly named Archaea from Bacteria as unique lifeforms (Woese & Fox, 1977). New studies demanded adjustments incorporating Crenarchaeota, Euryarchaeota, Korarchaeota, Nanoarchaeota, Thaumarchaeota (Brochier-Armanet et al., 2008). The most current edition of the Genome Taxonomy database (GTdb) stipulates the prior arrangement has now been superseded: (1) Crenarchaeota renamed to Thermoproteota; (2) Euryarchaeota divided into Halobacteriota, Methanobacteriota, Methanobacteriota A, Methanobacteriota B, Thermoplasmatota; (3) eliminating the remaining three former phyla through absorption into these new designations. Ongoing use of taxonomic names at all levels continue to suggest Archaea are offshoots of Bacteria (e.g., Halobacteriota) despite their distinction in ways physical/chemical/biological having been known for almost fifty years.

Ongoing linkage of Bacteria with Archaea is of more consequence than nomenclature. Theories regarding the process of translation whereby the genetic code is transformed into functional bioproteins are constructed mostly from bricks provided by studies of Bacteria (*E. coli*, *T. thermophilus*, *B. subtilis*). This feature is comprehensible in practical terms: *E. coli* is perceived as the superstar of genetics, the superior model species because ubiquitous in nature, including its incorporation into human bodies. It is easily cultured (37°C, pH 7.0, sugars, inorganic salts, H₂O, O₂) and amenable to manipulation for investigations. These characteristics specifically render it a terrible choice by which to elaborate ideas pertaining to translation in Archaea.

Many Archaea are extreme in their environmental preferences for survival and growth. *Haloferax volcanii* needs NaCl concentrations of 1.7–2.5M and lives in Dead Sea sediment (Hartman et al., 2010). *Archaeoglobus fulgidus* metabolizes sulfur-containing compounds and grows in culture at 60–95°C, with 83°C optimal (Klenk et al., 1997). Anaerobic *Thermoproteus uzoniensis* was extracted from a Kamchatka volcano (Mardanov et al., 2011); it grows at pH 5.5. Deep-sea Archaea in hydrothermal vents experience hydrostatic pressures hundreds of times greater than atmospheric (Jebbar et al., 2015). *Adaptive radiation* conveys the thought that these kinds of lifeforms have developed biological traits that, of necessity, must be other than those promulgated by *E. coli* (Yoder et al., 2010; Gong et al., 2020).

This paper has three parts: it provides details concerning archaeal tRNA synthetase sequences and their cross-species comparison; reverses prior emphasis (Laibelman, 2022) on tRNA differences in cross-species characteristics by concentrating on nucleotide invariance in position and identity within the chain; offers a discussion on implications generated from the first two areas for the aminoacylation component of the broader process of translation.

If there is a single, overarching take-home lesson from this research, it would be that geneticists should expunge *absolutely conserved*, *universally conserved*, *strictly conserved* from their descriptive vocabularies when extrapolating beyond data in hand. A collection of invariant amino acids or nucleotides within a select group of structures does not imply strict conservation of those same residues in sequences of organisms known but unchecked. Even more obfuscatory for complete understanding of mechanistic details in processes of transcription and translation is the (often implicit) suggestion that one instance of event occurrence in a single specimen is tantamount to establishing what is true for every organism. One does not constitute a pattern for anything; neither does two, three, or ten *if* those results are intended to cover lifeforms beyond that collection.

To borrow a legal phrase heard in courtrooms, pronouncements of universal XYZ “assumes facts not in evidence.” There are two ways to interpret this jargon: the narrow technical version indicates facts are known but not introduced into the courtroom according to prescribed rules of evidence; the more broad-based meaning is that facts are not known, but merely presupposed for the sake of argument. It is the second interpretation that guides its usage in the present context. Despite enduring quests for general theoretical simplicity based upon limited information, model species for *any* biological activity found within Archaea, Bacteria, Eukarya do not exist.

Materials and Methods

Textual information is written in variable-width Times New Roman font; sequence comparisons are written in constant-width Courier font.

Archaea used to compare genome sequences for tRNA synthetases included all those studied in the investigation of tRNA sequences (Laibelman, 2022) except for multiple strains from the same species, which were removed to leave a single strain. After removal of extras, additional species were incorporated to bring the total to 150 organisms of interest. No attempt was made to include representatives from every genus. After species selection, relevant synthetase protein sequences were downloaded from UniProtKB (<https://www.uniprot.org>). This archive provides two related databases: (1) trEMBL for protein sequences obtained from three agencies comprising the *International Nucleotide Sequence Database Collaboration* (NCBI GenBank, EMBL European Nucleotide Archive, DNA Data Bank of Japan; Karsch-Mizrachi et al., 2018); (2) swiss-prot, a refined version of trEMBL in which items from INSDC are reviewed and annotated. Both sources were utilized and marked ‘tr’ or ‘sp’ as appropriate. Sequences were periodically downloaded between January and October 2024.

The ‘tr’ archive was cluttered: executing searches for any particular amino acid ligase (the official designation instead of synthetase) also retrieved indirectly related enzymes (e.g., editing proteins), as well as unrelated enzymes (e.g., lipooates); all extraneous proteins were ignored. On the other hand, sequences pertaining to amidotransferase subunits A–E were retained because essential for charging tRNA^{Asn} and tRNA^{Gln}, as will be discussed. Otherwise valid tRNA ligase strings were manually removed if containing: (1) indeterminate amino acids signified by ‘X’ in place of the conventional single-letter designation; (2) duplicates in both databases (‘tr’ versions are usually deleted by UniProtKB); (3) sequences not beginning with methionine; (4) short sequence fragments less than 300 residues *except* for amidotransferase subunit C since the longest of these is 111 amino acids.

Only strings related to the standard twenty amino acids in the genetic code were ultimately kept. Sequences for phosphoseryl-tRNA synthetase and pyrrollysyl-tRNA synthetase were available and also initially downloaded, but later removed from the collection because there were no tRNA corresponding to them on hand from the earlier compilation. Some species possess two enzyme versions (rarely three) associated with a single amino acid. Designation as a multigene synthetase always excludes PheRS α and β subunits. There were occasions, especially common for *Picrophilus oshimae*, in which these variants differed by eight or less mutated residues, and all were retained. The final dataset totaled 3626 synthetase and amidotransferase sequences, of which 2819 were from trEMBL and 807 from swiss-prot.

The authority for nomenclature is the Genome Taxonomy database (<http://gtdb.ecogenomic.org>). Adjustments made on this website are not automatically transferred either to the INSDC archives or into UniProtKB. For newly discovered and sequenced organisms, a lack of linkage is not problematic because their taxonomic information consists of simple addition to existing groupings. However, categories used for established Archaea are changed throughout the taxonomic hierarchy, as mentioned in the Introduction. This instability creates issues for unlinked resources and for researchers dependent upon them for accurate information. To illustrate: when the tRNA genome comparison study was undertaken, GTdb version R207 from June 2021 was authoritative. For the present tRNA synthetase examination, current is R220 from April 2024 (i.e., one revision every three months).

The NCBI database contains unique identifier GCAxxxxxxxx.x or GCFxxxxxxxx.x exhibiting minor variations due to research group-derived sequence revisions indicated by a post-decimal point digit. GTdb uses this identifier, so the current taxonomic version can be correlated even if taxonomy changes at the species level. UniProtKB species names agree with NCBI's since the latter, through INSDC, is input source for the former. In principle, all three (NCBI, GTdb, UniProtKB) can be integrated. A complication arises for Archaea and Bacteria: strains within species. UniProtKB employs strain names in addition to organism names. The question becomes how to relate strain designations with the correct NCBI identifier. This is accomplished by perusing the GTdb *change history* table showing which GTdb name connects to which NCBI variant for each species.

Consider *Archaeoglobus profundus* altered to *Archaeoglobus B profundus* according to the GTdb *change history* table. The genus has been subdivided into A, B, C variants in addition to retention of the original: *A. fulgidus*, *A. A sulfaticallidus*, *A. B profundus*, *A. C veneficus*. The archives for *profundus* are linked through NCBI ID GCF000025285.1 provided in taxonomic database R220. The UniProtKB strain name for *A. B profundus* is represented either by research group-appointed designator Av18 or sequence identifier DSM 5631, among other institutional identifiers signifying where different types of data for this organism have been stored. Although UniProtKB contains other *profundus* strings for tRNA ligases, none possess a strain designator. For this precise reason Av18/DSM 5631 was selected for study in preference to these others, which could not be otherwise conveniently differentiated.

Multiple transnational organizations assign identifiers for archaeal species and ignore research group-designated in-house labels. This causes confusion: beyond GTdb's official species name, Archaea are referenced differently depending upon host country. Multiple cross-referenced identifiers create the same situation as would occur in the United States if a person's social security number changed each time he/she moved to a new state (e.g., Georgia to North Dakota). It would be desirable for INSDC to make standard a single universal appellation beyond idiosyncratic researcher-derived common versions. As an admittedly extreme example, UniProtKB provides tRNA ligase sequences for:

Natrialba magadii ATCC 43099 = DSM 3394 = CCM 3739 = CIP 104546 = IAM 13178 = JCM 8861 = NBRC 102185 = NCIMB 2190 = MS3. ATCC→ American Type Culture Collection; DSM→ German Collection of Microorganisms (Deutsche Sammlung von Mikroorganismen); CCM→ Canadian College of Microbiologists; CIP→ Collection de l'Institut Pasteur (France); IAM→ Interactions Arbres-Microorganismes (France); JCM→ Japanese Collection of Microorganisms; NBRC→ NITE Biological Resource Center (Japan) where NITE→ National Institute of Technology and Evaluation; NCIMB→ National Collection of Industrial and Marine Bacteria (United Kingdom); MS3→ research group common name

In this paper, DSM number is used as strain name 106 of 150 times (70.7 %) because it applied to the largest number of species investigated; common names are employed next most frequently (37 of 150, 24.7%). If UniProtKB recorded neither, JCM number was used for *Haloarcula amylolytica*, *Halorubrum distributum*, *Natrinema limicola* because it was the only descriptor given beyond species name; ATCC number is used as strain name for *Haloarcula vallismortis* for an identical reason;

Acidianus B brierleyi, *Acidianus infernus*, *Geoglobus acetivorans* A lack supplemental identifiers altogether.

After complete enzyme sets were processed for 150 species, intergenome comparisons for each amino acid-affiliated set of aaRS were undertaken. The organisms were divided into subgroups by phyla: Halobacteriota (69), Methanobacteriota (38), Thermoproteota (43). Clustal Omega multiple sequence alignment tool (v.1.2.4, accessed December 2024–February 2025) performed the requisite comparisons (<https://www.ebi.ac.uk/jdispatcher>; Goujon et al., 2010). Residue counts for each sequence obtained from UniProtKB were verified by Clustal Omega totals. If discrepancies existed, those from the latter source were preferentially accepted.

Results were retrieved along with an automatically created phylogenetic tree generated from their neighbor-joining distance matrix algorithm (Madeira et al., 2024). On each branch of every tree is a five-place decimal value between zero and one intended to convey how different each sequence's amino acid composition is from immediately adjacent neighbors; values close to zero indicate high levels of sequence similarity, meaning that pair is more closely related from an evolutionary development perspective. After alignments were produced, it became clear each phylogenetic tree gave a different evolutionary picture for every aaRS. For example, Archaea closely related (nearest neighbors) in primary sequence for GlyRS strings may also be relatively disconnected in AlaRS according to their respective branch locations on that tree; decimal values lack cumulative probative utility, and for this investigation were regarded as pertinent to a rounded-off two-digit value. Usage was confined to determining whether alignments omitting strings > 0.30 would generate significantly more acceptable results in terms of the number of invariant amino acids produced across all species in that comparison.

Justification for accepting only exact amino acid matches in a given location is straightforward. Although the majority of residue substitutions probably preserve overall shape and/or functionality—Val for Ile; Lys for Arg; Ala for Thr, etc.—only exact matches can be deemed certain of such conservation. Since aminoacylation kinetics studies do not cover all species of interest, it seems prudent to accept identical sequences exclusively: kinetic rate variability may be undetermined, yet of potential significance for in vivo function. In particular, catalytic rate differences and/or adjustments in equilibrium binding constant data may reflect adaptation to an organism's environment.

The prior study of tRNA sequences (Laibelman, 2022) covered 186 Archaea, including multiple strains for single species, downloaded from Genomic tRNA database (GtRNAdb; <http://gtgnadb.ucsc.edu>) released on 19 June 2021. The collection was narrowed by omitting all but one strain for a given species. Others did not match organisms chosen for the current exploration of tRNA synthetase sequences. After their removal, the remaining 131 tRNA strings were divided into phyla, leading to sets of Halobacteriota (56), Methanobacteriota (37), Thermoproteota (38). Clustal Omega was employed to align sequences for each phylum.

Only tRNA representing the canonical twenty amino acids were used; encoded selenocysteine or pyrrolysine strings were ignored. In 129 of 131 instances, tRNA^{Ile(UAU)} is modified in the first anticodon base to tRNA^{Ile₂(C⁺AU)}, where C⁺ signifies altered cytidine base *agmatidine*. Exceptions are *Korarchaeum cryptofilum*, which retains UAU as anticodon, and *Caldivirga maquilensis*, utilizing neither standard nor changed nucleobase, encoding conventional tRNA^{Ile(GAU)} isoacceptor version only. A decision was made to exclude tRNA^{Ile₂(C⁺AU)} from alignments because it would represent the sole modified nucleobase, which seems arbitrary despite its abundance. Chemical alterations are not considered for other positions despite there being “greater than one hundred naturally occurring chemical modifications” (Paris et al., 2012) because there exists a lack of knowledge about such changes for the majority of Archaea (Wolff et al., 2020).

Results and Discussion

Part I. Synthetases and Amidotransferases

Most of the compiled data is too complex to be conveniently presented in tabular form within the main body of this report; it will be relegated to the Supplementary Information section accessible at the reader's request. Supplementary Table 1 lists the Archaea studied; it presents the currently accepted name, including strain designation, found within GTdb arranged in alphabetical order by genus. It also provides names according to NCBI, which is not authoritative, but many nonspecialists are more familiar with its nomenclature. GTdb and NCBI names are linked by a nonvarying NCBI ID descriptor even as taxonomic changes are recorded by newer versions of GTdb. Segregation into Halobacteriota, Methanobacteriota, Thermoproteota phyla are indicated; an asterisk (*) signifies species used in the sequence comparisons for which both aminoacyl tRNA synthetases and tRNA data are on-hand.

Supplementary Table 2 presents cell culture growth conditions (optimal plus range, if available) organized alphabetically by genus and species within each phylum. Information has been compiled for (1) temperature (°C); (2) pH (standard log units); (3) NaCl concentration (molar units); (4) O₂ necessity. In a few cases where the literature provides specific quantities, pressure (atm) is included with the temperature information. Oxygen necessity turns out to be more complicated than a simple aerobic *versus* anaerobic dichotomy. Some organisms grow under either condition relatively equally depending upon the nutrients supplied in a culture medium, while others are essentially anaerobic but able to tolerate 1–5% O₂ by volume and denominated *facultative* anaerobes.

In the early stages of information collection, individual articles were consulted; midway through the process, Google introduced an experimental AI search engine summarizing values from the literature based on a generic query: "Name (genus plus species) cell culture growth conditions." When presented, this summary was used without searching any linked articles. If more than one reference was uncovered, occurring when multiple species are associated with a single genus, or in rare cases if sources did *not* agree quantitatively, both sets of data were included in the tabulation. At the other extreme, upon perusal of Table S2, it will be observed some parametric values were not recorded. This was the case frequently for effects of salt on growth for Methanobacteriota and Thermoproteota, where less than 50% were found for these organisms.

Terms classifying organisms by level of preference for any given parameter, such as *thermophile*, *acidophile*, *halophile*, etc. vary widely in scope of coverage when quantified. For example, temperature(s) distinguishing a hyperthermophilic archaeon from one which is thermophilic is not universally employed. Since agreed-upon standards are lacking, these descriptors are avoided, and in their place, parsing into categories based on specific value ranges is arbitrarily imposed. The exception is a requirement for presence of O₂ in culture to support cell growth. *Aerobic* and *anaerobic* are quantitative because mutually exclusive, and *facultative* has been defined above, as has the term *both* in this context.

Table 1 summarizes Table S2 by counting the number of species in each category. When phylum distinctions are used in conjunction with values for each variable, it is seen that most Archaea are properly assigned:

- 39 of 43 Thermoproteota satisfy the hyperthermophile or thermophile designation usually employed for them;
- 34 of 38 Methanobacteriota are anaerobic, the natural consequence of being methane producers by reducing oxygen-containing substrates such as acetate or formate;
- 36 of 60 Halobacteriota demand sodium chloride at concentrations exceeding 1.0 M, and as shown in Table S2 (not Table 1), seven of the remaining twenty-four species for which numbers were compiled prefer salt amounts not to exceed 0.1 M (5.8 g/L);
- 24 of 43 Archaea from phylum Thermoproteota would also be construed as acidophiles;

- 18 of 69 halophiles (Halobacteriota) and 25 of 38 methanogens (Methanobacteriota) would also be defined as thermophiles;
- 92 of 150 Archaea, at minimum, are classified as anaerobic.

Table 1. Cell Culture Growth Conditions Summary.

	All Archaea	Halobacteriota	Methanobacteriota	Thermoproteota
		N = 69	N = 38	N = 43
Temperature				
> 80°C	52	5	16	31
50–79°C	30	13	9	8
25–49°C	66	50	13	3
< 25°C	2	1		1
pH				
> 9.0	7	7		
6.0–9.0	99	54	26	19
< 6.0	37	5	8	24
NaCl				
> 3.0 M	31	31		
1.0–3.0 M	5	5		
< 1.0 M	59	24	17	18
Oxygen				
Aerobic	47	33	2	12
Anaerobic	92	35	34	23
Both	4		2	2
Facultative	7	1		6

The inevitable existence of taxonomic violators, no matter how few or numerous, is due solely to the fact that systematic classification utilizes multiple criteria (Gong et al., 2020). First and foremost is genotype grounded on primary sequence alignment for 16S rRNA, despite claims of being an unreliable indicator for halophilic Archaea, and possibly more generally (Boucher et al., 2004). Second is phenotype, usually determined by cell morphological shape based on width and diameter in addition to ether lipid composition in its membrane. In third place, at best, are growth conditions in culture. The assertion in the Introduction that there is no possible pristine taxonomy seems well-supported.

Supplementary Table 3 presents the raw data upon which all subsequent information is grounded: sequences of encoded synthetase and amidotransferase genes arranged alphabetically by genus + species, with separate tables covering each associated amino acid. They are written with constant-width courier font in order to conveniently enable residue numbering using the conventional arrangement with N-terminus on the extreme left and C-terminus on the extreme right. Where multiple genes are found for a single aaRS or amidotransferase, they are recorded on successive rows and labeled ‘1’, ‘2’, ‘3’. Multiple genes for any enzyme may contain strings with: (1) limited numbers of mutations (never more than eight); (2) shorter fragments of full-length sequences; (3) alternative and apparently unrelated full-length strings annotated by UniProtKB as that

synthetase. The alternative SerRS2 version (Kim et al., 1998) is located immediately after SerRS, and amidotransferases with subunits A–E are presented after ValRS.

Supplementary Table 4 summarizes Table S3 data by concisely supplying information for species (strain name, phylum), UniProtKB ID, type of synthetase or amidotransferase, sequence length in number of amino acids. In total, 3626 sequences are encoded by 150 Archaea.

Four synthetase sequences are missing from UniProtKB’s archive, and their absence confirmed by unsuccessfully searching NCBI’s database: GluRS for *Thermoproteus tenax*; IleRS for *Saccharolobus islandicus*; SerRS for *Ignisphaera aggregans*; ValRS for *Natrinema limicola*. In addition, a sequence for amidotransferase subunit A from *Hyperthermus butylicus* was not found. Possibilities for absence are: (1) not encoded in genomes; (2) not sequenced by those generating that organism’s other enzyme strings; (3) not annotated, or annotated incorrectly such that search requests failed to find them (Bastian et al., 2015).

Nine IleRS strings for *Saccharolobus islandicus* strains are available from UniProtKB and NCBI, but for unknown reasons the selected strain (Lassen #1) is not among them. Although the likelihood is high the missing sequence would be a match for at least one of the others available, it is not a certainty; caution dictated no substitute be used. There is no ready explanation for unavailability of the other enzymes beyond the generalized options expressed.

A synthetase initially thought absent from UniProtKB, HisRS from *Caldisphaera lagunensis*, led to discovery of a disturbing feature. Several months after the first download of HisRS genes, a new search was performed in order to confirm its exclusion because the database is updated continuously. On this occasion, an appropriate string was found, and its existence supported by one from the NCBI repository. These sequences differed though the amino acid compositions were identical: a thirty-six residue segment (LIVDSIKSLGFESGFSRLNDRRLLSGIFEQELNIK) was located immediately after a longer block of residues (SPLAVYRIIDKLDKIGIDNVKKELLEQINNEEIVNKKIIEVISLSGKPEEILENLYSKYGR) for UniProtKB, but immediately preceded it in the NCBI version (Supplementary Table 5).

Rather than consult still another source to attempt to resolve the discrepancy, Clustal Omega was employed to align the contrary sequences against all other Archaea included in this study. The result was unambiguous: the NCBI sequence fit with all other HisRS and a long series of gaps (– – –) was added by Clustal’s alignment algorithm against the UniProtKB string; Table S5 has details. As one-third partner in INSDC, NCBI is a primary repository for proteins, whereas UniProtKB is a secondary archive getting information from the Consortium (Goudey et al., 2022). How the sequence became rearranged is unknown; a study involving three representative bacteria found numerous disagreements between NCBI’s Gen Bank and UniProtKB’s swiss-prot databases (Karp et al., 2001).

According to UniProtKB, LeuRS for *Cenarchaeum symbiosum* encodes in two pieces of 373 and 578 amino acids. When these strings are included individually in a multi-sequence alignment with LeuRS from other species, their relative shortness (40% and 62%, respectively, of the average length of 933 amino acids) causes Clustal Omega to insert many gaps to compensate. It was observed that their combined length was near the average, and inspection made it obvious they really were one sequence: one piece contained HIGH tetrapeptide and the other KMSKS pentapeptide. A repeat multiple sequence alignment with a united *C. symbiosum* markedly improved the result: fewer gaps led to forty-five invariant amino acids. Whether the *C. symbiosum* genome contains LeuRS pieces in different genes, or this partitioning was an inadvertent error introduced somewhere in the chain of transmission between sequencing and archiving, is unclear. No other split enzymes were encountered.

Synthetases have one fundamental function: bind ATP, amino acid, cognate tRNA (not necessarily simultaneously or in that priority) in order to unite the second pair by ester formation between the amino acid’s carboxylate and the 2’-OH (or 3’-OH) of the terminal ribose in a given tRNA. Some synthetases display large length variation among its archaeal members, while others are tightly compacted in length range, as signified by standard deviation from the average. Table 2 displays length ranges by enzyme type as well as the number of total sequences encoded for each. In

Table 2, $\pm$ values require division by three to obtain one standard deviation. If arranged by decreasing average length, and focusing only on aaRS, except for IleRS at the high end and TrpRS at the low end, all seven enzymes whose standard deviation is ≥ 40 residues are on the long-length side of the scale: longer length implies greater length variability. Of the seven, LeuRS, ValRS, MetRS, ArgRS are Class Ia synthetases, TrpRS is Class Ic, AlaRS and ThrRS are Class IIa enzymes (Chalioitis et al., 2017). Aside from ArgRS and ThrRS, the others are hydrophobic.

While there is no obvious explanation for extreme enzyme length range within a synthetase type, some variation in residue number between aaRS may be accounted for by addition of an editing domain to a core base region in order to reduce chances of misacylation (Beebe et al., 2004). This does not fully account for the observation: IleRS has an editing domain (Nureki et al., 1998; Fukunaga & Yokoyama, 2006), but a small standard deviation for amino acid number.

Unlike synthetases, the five amidotransferase subunits show little variety in residue number. This is consistent with the outcome anticipated, yet not attained, for synthetases because they also have single biochemical functions (Sheppard & Söll, 2008): (1) act as kinase to phosphorylate aspartyl-tRNA^{Asn} or glutamyl-tRNA^{Gln} yielding O-phosphoaspartyl-tRNA^{Asn} (B subunit) and O-phosphoglutamyl-tRNA^{Gln} (E subunit); (2) hydrolyze sidechain amides in asparagine (A subunit) or glutamine (D subunit) and deliver released ammonia molecules to activated carboxylates for substitution on their respective tRNA; (3) assist subunit A in function (C subunit).

Table 2. Synthetase and Amidotransferase Distribution by Type.

Enzyme	n	AA Range	Avg $\pm$ 3 StDev	Enzyme	n	AA Range	Avg $\pm$ 3 StDev
AlaRS	153	446-936	900 $\pm$ 195 = 705-1095	PheRS— α subunit	153	323-561	496 $\pm$ 72 = 424-568
ArgRS	152	542-770	600 $\pm$ 138 = 462-738	PheRS— β subunit	153	418-634	554 $\pm$ 84 = 470-638
AsnRS	43	419-439	431 $\pm$ 9 = 422-440	ProRS	155	380-512	480 $\pm$ 42 = 438-522
AspRS	46	417-476	433 $\pm$ 36 = 397-469	SerRS	126	398-516	452 $\pm$ 54 = 398-506
Asn_AspRS	113	372-609	440 $\pm$ 87 = 353-527	SerRS2	27	495-527	508 $\pm$ 27 = 481-535
CysRS	135	374-559	475 $\pm$ 78 = 397-553	ThrRS	164	384-710	603 $\pm$ 198 = 405-801
GluRS	150	532-634	570 $\pm$ 36 = 534-606	TrpRS	169	310-614	423 $\pm$ 192 = 231-615
GlyRS	150	445-652	576 $\pm$ 90 = 486-666	TyrRS	161	306-388	338 $\pm$ 66 = 272-404
HisRS	152	303-470	426 $\pm$ 54 = 372-480	ValRS	153	768-939	857 $\pm$ 123 = 734-980
IleRS	151	959-1143	1065 $\pm$ 90 = 975-1155	Amidotransferase subunit A	115	421-501	451 $\pm$ 66 = 385-517
LeuRS	182	395-1022	933 $\pm$ 159 = 774-1092	Amidotransferase subunit B	117	446-509	482 $\pm$ 48 = 434-530
LysRS	160	480-610	551 $\pm$ 78 = 473-629	Amidotransferase subunit C	92	71-111	92 $\pm$ 18 = 74-110
MetRS	152	498-772	692 $\pm$ 201 = 491-893	Amidotransferase subunit D	150	382-469	426 $\pm$ 48 = 378-474
				Amidotransferase subunit E	152	591-650	627 $\pm$ 33 = 594-660

In contrast to missing enzyme sequences, some found in UniProtKB were subsequently removed. Two amidotransferase subunit C strings are annotated as derived from computer prediction rather than by laboratory experiment. It is unknown why authentic information could not

be obtained since there was no difficulty for other related enzymes encoded in *Methanocaldococcus jannaschii* or *Methanothermobacter marburgensis* genomes. PheRS- β from *Halalkalicoccus jeotgali* possessed nine C-terminal amino acids in the second of two otherwise identical genes that had to be a contaminant from another organism. The first, designated *H. jeotgali1*, continues to full length (567 residues), but *H. jeotgali2* terminates at 418 residues (including the foreign component).

Beyond these three near-mandatory exclusions, another justification for removing sequences prior to finalizing multiple sequence alignments is predicated on Clustal Omega's liberal insertion of gaps to maximize invariance in the identities of amino acids at fixed locations. String length becomes a critical factor because especially short or long sequences distort that evaluation. Average length and standard deviation from that average were calculated to assess which strings might be intentionally excluded from alignments (Table 2). Statistically, 99.7% of datapoints plausibly construed as obeying normal Boltzmann distributions lie within three standard deviations of the average. It is reasonable that enzyme length should conform to a Boltzmann-type constraint, as there is no *a priori* reason to believe they should exhibit bias such that not sorting themselves equally about the average could be anticipated. Therefore, lengths not within three standard deviations were considered for exclusion. It is obvious this approach lends itself to potential gross manipulation of results if not applied judiciously; a statistical justification is insufficient. Alignments were conducted with and without the 'deviant' strings. At least a 100% increase in number of invariant amino acids for the 'without' test run had to be found before deciding permanent removal from further analysis was methodologically proper.

There was still a third factor to ponder in decisions to remove sequences from alignment without undermining validity of the output. The phylogenetic tree created by Clustal Omega is complemented by a distance matrix calculation. This value runs $0 \rightarrow 1$ carried out to five decimal places, and supposedly signifies how close nearest neighbor strings in each tree are to each other from an evolutionary point of view based on shared residues at each position. Two decimal places were deemed sufficient for the purpose of choosing possible candidates to delete from a phylum-based set, so a tree distance above 0.30 was initially (and arbitrarily) selected as a potential cutoff. After tentative removal of outliers, alignments were run again to observe changes in the assessed number of invariant amino acids. It became clear, after trials with different enzymes and organisms, that 0.30 was too low a bar: alteration in invariant residue number was too minimal to justify such an extreme approach. Empirically, tree outliers ≥ 0.43 caused dramatic adjustment to invariant amino acid number, defined again as a minimum 100% gain; it became the smallest acceptable criterion for permanent removal of the offending string.

In illustration: the original SerRS alignment covered 126 species because the remaining twenty-four encode an alternate version of the synthetase designated *SerRS2* (Kim et al., 1998). Three species, all from genus *Methanosarcina*, possess genes for both SerRS and SerRS2, meaning there were 153 strings total. In the initial comparison, twenty-three invariant amino acids were found when all SerRS species strings were aligned; an accompanying phylogenetic tree showed *Methanosphaerula palustris* as having a distance value of 0.50, signifying its sequence was very different from 125 others. A second comparison for these 125 increased the number of invariant residues 2.7-fold to sixty-one. Consequently, *M. palustris* was removed from the set.

It is recognized sequence removal decisions based upon outlier lengths or evolutionary distance measures could be legitimately questioned, with an associated charge of 'cherry picking' outcomes. There must exist convincing reasons to eliminate encoded sequences from alignments, and it is hoped the 100% improvement criterion provides a satisfactory rationale. Nonetheless, this procedural fix has an unsettling implication: researchers possess ability to manipulate uncovered data. Such control may alter conclusions, especially if subsequent effort is made to generalize results by assertions of universal conformity beyond the organisms studied. Even if this kind of extrapolation is not pursued, and it is not here so proposed, statements concerning invariant amino acids in a set of proteins alleged to provide a given biochemical function has consequences for assessment of mechanism utilized by that protein or enzyme among the collection of lifeforms investigated.

Rejected enzymes are shown in Table 3, along with justification for removal and consequences for undertaking such action. Thirteen of 3626 sequences (0.36%) were not considered for further analysis of results.

Table 3. Removed Sequences; H = Halobacteriota; M = Methanobacteriota; T =Thermoproteota.

Enzyme	Organism	H/M/T	Justification	Invariant AA Change
AlaRS	N magadii1	H	446 AA, 49.6% of avg	29 → 127
AlaRS	M alvus	M	590 AA, 65.6% of avg	49 → 115
AlaRS	M A intestinalis	M	593 AA, 65.9% of avg	49 → 115
Asn_AspRS	M alvus	M	606 AA, 137.7% of avg	64 → 135
Asn_AspRS	M A intestinalis	M	609 AA, 138.4% of avg	64 → 135
Asn_AspRS	K cryptofilum	T	tree outlier (0.43)	42 → 85
LeuRS	H tiamatea3	H	395 AA, 42.8% of avg	17 → 45
PheRS—α subunit	C symbiosum	T	tree outlier (0.50)	8 → 34
PheRS—β subunit	H jeotgali2	H	foreign C-terminus (9 AA)	43 → 53
PheRS—β subunit	C symbiosum	T	tree outlier (0.50)	8 → 30
SerRS	M palustris	H	tree outlier (0.50)	23 → 61
Amidotransferase subunit C	M jannaschii	M	computer predicted	0 → 4
Amidotransferase subunit C	M marburgensis	M	computer predicted	0 → 4

The first-pass effort in determining invariant amino acid numbers often led to fewer than ten such residues when alignments were performed in a one-batch operation on all acceptable downloaded strings for each synthetase type. Sheppard & Söll (2008) demonstrated Crenarchaeota and Euryarchaeota formed isolated clusters in the resultant phylogenetic trees of archaeal amidotransferase sequences, and it was thought synthetases might perform similarly. Since, from the time their work was done, division into two phyla has been superseded, the new taxonomic categorization was applied. Tables 1 and S2 data strongly support their analysis. Acidity was not used as a classifier because inclusion would overdetermine implied boundaries: acidophiles, neutrophiles, alkaliphiles are sprinkled throughout the three phyla. The first run of sequence comparisons on all archaeal sequences using Clustal Omega became a baseline reference for subsequent trial runs, which always pertained to Halobacteriota, Methanobacteriota, Thermoproteota.

In Table 2, n > 150 indicate multiple genes. Its occurrence is due to: (1) mutations of one or more residues when sequences are internally compared; (2) one sequence being a fragment of another due to deletion of residues either at the N- or C- terminus; (3) UniProtKB-annotated alternative sequences. The third option dominates: distinction consists of diversity in amino acid composition and order. Allocation of multiple genes, and distribution as to enzyme type, sometimes occur in patterns related to genera:

- two *Nitrososphaera* encode two of four CysRS multiples, and possess single genes for all others;
- three *Acidianus* encode three of twenty-eight LeuRS multiples, and possess single genes for all others;

- eight *Methanosarcina* encode eight of ten LysRS multiples, and possess single genes for all others;
- two *Methanocella* encode two of fourteen ThrRS multiples, and possess single genes for all others;
- six *Pyrobaculum* encode six of eleven TyrRS multiples, and possess single genes for all others.

Some species (not genera) are prone to multiple gene production for more than one enzyme type. *Picrophilus oshimae* is unique, thriving comfortably in hyperacidic ($\text{pH} \leq 2$) surroundings (Schleper et al., 1996). It seems plausible its creation of multiple genes for eighteen synthetases plus amidotransferase subunit E ties directly to this environmental preference. Each repeated gene is a full-length copy with 1–8 residue mutations, but otherwise identical every time. Aside from these paired mutation strings, species from Methanobacteriota encode multiple genes once: *Pyrococcus furiosus* produced a mutation duplicate for AlaRS. The other two phyla are more prone to generate multiple genes either as fragments of encoded full-length genes or as strings of different composition, depending upon enzyme type. Chaliotis et al (2017) claimed synthetase duplicates and fragments are involved in tRNA-dependent amino acid biosynthesis, and it would be a worthwhile project to see if this proposal held over the variety of species examined here. A complete accounting (enzyme type, species, phylum, multiplicity category) is recorded in Supplementary Table 6.

Multiplicity in synthetases in conjunction with noticing a similar phenomenon in archaeal tRNA (Laibelman, 2022), led to an inquiry as to whether the same species were involved in generating these two molecular types crucial for the translation process. Comparing tables of relevant synthetase enzymes with equally pertinent tRNA isodecoders showed an *absence* of correlation. *Aeropyrum camini* encodes dual tRNA^{Arg(GCG)}, but no multiple synthetases. *Caldisphaera lagunensis* contains nonidentical ThrRS, but no more than one tRNA for an isoacceptor. *Halalkalicoccus jeotgali* has two tRNA^{Arg(UCG)} and tRNA^{Thr(UGU)}, but its multiple synthetase copy is for the PheRS- β subunit, and exists as a fragment, not a full-length molecule.

Illustration could continue, but the implication is clear: whatever causes replication of genes in Archaea, identically sequenced or not in each case, is not an organism's global response to environmental stimuli. Causality for tRNA does not carry over to aaRS, or vice versa. Perhaps this outcome is obvious from the simple fact that multiplicity in any given species does not extend to every amino acid-connected molecule meaningful for translation, except for synthetases in *P. oshimae*, although this archaeon encodes nonidentical multiple copies for tRNA^{Asn(GUU)} alone. With respect to production of nonidentical multiple gene copies in Archaea, tRNA and their correlative synthetases did *not* co-evolve. A lack of co-evolution was also the conclusion drawn from absence of correlation in phylogenetic trees for LeuRS and tRNA^{Leu} sequences (Andam et al., 2012).

Regardless of whether attention is focused on single genes, multiple genes, or combinations of both, quantitative results from sequence alignments for synthetase and amidotransferase enzymes encoded by Halobacteriota, Methanobacteriota, Thermoproteota are distinct in terms of number and identity of invariant amino acids. Table 4 is unambiguous in its conclusiveness about quantity. Cells with arrows ($\rightarrow$) indicate changes incurred when certain species are removed from alignment; it mirrors information from Table 3.

Alignments for 'All Archaea' collections yield fewer invariant residues than obtains from aligning Halobacteriota, Methanobacteriota, Thermoproteota sets individually, and crucially, these quantities differ among the three phyla. This outcome supports distinction when cell culture growth factors are compared in Tables 1, S2. It implies diverse environmental living conditions impact gene pools, which is not exactly an innovative conclusion. More emphatically, as ancient in evolutionary development as these enzymes are said to be (Woese et al., 2000; O'Donoghue & Luthey-Schulten, 2003), nurture was instrumental in enhancing natural genetic divergence beyond the conventional random mutation channel.

Segregation by phylum did not lead to large numbers of invariant amino acids for LysRS, unlike what was found for all other types of synthetases and amidotransferases. This result is informative. It is the only aaRS proclaimed to exist in Class I and II versions (Ibba et al., 1997a), which would certainly influence the outcome if both classes were mixed indiscriminately in attempts at

maximizing alignment of individual species. According to Levensgood et al. (2004), almost all Archaea are Class I with a limited number of species being members of Class II.

At the time those statements were made, small numbers of archaeal species were known of either type. It has been difficult to uncover, given the intervening years and discovery of numerous organisms in remote or unconventional geographic areas, which additional Archaea should join the minority category beyond an original characterization as “few”. Class II definitely includes: (1) *Saccharolobus solfataricus* (Ibba et al.,1997b); (2) *Pyrobaculum aerophilum* (Woese et al., 2000); (3–5) *Aciduliprofundum boonei*, *Archaeoglobus A sulfaticallidus*, *Thermogladius calderae* (<https://www.aars.online>); (6–13) all members of genus *Methanosarcina* (Mahapatra et al., 2007). It cannot be assumed others within these genera share this property; absent proof, they could just as probably encode the Class I type. *Archaeoglobus fulgidus* is specifically included within phylogenetic trees for Class I LysRS (Ambrogelly et al., 2002) in contrast to *Archaeoglobus A sulfaticallidus*.

Table 4. Numbers of Invariant Amino Acids in Synthetases and Amidotransferases.

Synthetase	All Archaea	Halobacteriota	Methanobacteriota	Thermoproteota
	Invariant AA	Invariant AA	Invariant AA	Invariant AA
AlaRS	7	29 → 127	49 → 115	92
ArgRS	17	43	49	36
AsnRS	83	3 encoded	137	98
AspRS	46	0 encoded	101	52
Asn_AspRS	17	75	64 → 135	42 → 85
CysRS	24	36	71	47
GluRS	22	75	63	45
GlyRS	44	107	91	66
HisRS	22	38	43	36
IleRS	60	182	140	89
LeuRS	0	17 → 45	82	45
LysRS	1	4	4	2
Class I	0			
Class II	93			
MetRS	22	100	64	37
PheRS—α subunit	6	70 → 77	53	8 → 34
PheRS—β subunit	1	43 → 53	38	8 → 30
ProRS	19	30	78	45
SerRS	12	23 → 61	24	63
SerRS2	59	208	141	0 encoded
ThrRS	3	30	32	3
TrpRS	6	29	37	18
TyrRS	21	44	37	34
ValRS	46	119	110	85
Amidotransferase	All Archaea	Halobacteriota	Methanobacteriota	Thermoproteota
	Invariant AA	Invariant AA	Invariant AA	Invariant AA
Subunit A	38	99	73	78

Subunit B	38	86	87	69
Subunit C	0	4	3 encoded	2
Subunit D	42	76	59	57
Subunit E	57	125	84	89

Direct Comparison	All Archaea Invariant AA	Halobacteriota Invariant AA	Methanobacteriota Invariant AA	Thermoproteota Invariant AA
AspRS + Asn_AspRS	24		86	46
SerRS + SerRS2		18	16	

Alignment of these thirteen Class II LysRS found ninety-three invariant amino acids. However, alignment of the remaining 137 Archaea as if they are Class I members produced zero consensus residues, which proves the membership list for Class II is incomplete. At least some, perhaps many, isolated and sequenced over the last twenty years must be transferred from Class I to Class II. Movement should elevate invariant residue number for the former while not diminishing greatly the latter’s output. Discovering which species are properly designated Class II, such as possibly other *Pyrobaculum* and/or *Saccharolobus*, will require detailed research.

The real issue of far more importance than numbers visualized in Table 4 is the identities of those invariant residues. Since Clustal Omega adds gap spaces as frequently as the algorithm deems necessary in order to maximize conserved amino acids at each position, relative location within a sequence is lost. Residues adjacent in one sequence may become disconnected after the algorithm aligns multiple strings. A partial solution eliminates mutable amino acids as well as gap spaces, leaving only invariant residues in the order in which they were originally placed as a way to express the consensus outcome for a single phylum. It is desirable, however, to be able to compare consensus results across phyla in order to perceive a more generalized invariance. This approach is *not* equivalent to the ‘All Archaea’ alignment because the number of gap spaces introduced by Clustal Omega depends directly on the number of sequences it is trying to align simultaneously. It is for this reason Table 4 shows higher totals for the separate phyla than for ‘All Archaea’.

Using small clusters of those remaining amino acids that are still adjacent (as few as two, as many as seven) in consensus sequences for each phylum permits alignment across all phyla to be performed manually. The resulting condensed format still eliminates all absolute location information (which is why this approach is only a partial solution), but makes the identities of those conserved easier to determine when large numbers of strings are compared at one time. Moreover, the loss of absolute position data is not as great a liability as might initially appear: folded enzymes frequently unite separated regions viewed one-dimensionally into a close encounter for three-dimensionally functional molecules. In illustration, 152 histidyl-synthetase sequences containing 303–470 amino acids in each string (Table 2) generates consensus condensed versions for the three phyla as shown (Halobacteriota on the first line, Methanobacteriota on the second line, Thermoproteota on the third line):

HisRS H/M/T
000000000111111111222222222233333333344444444455555555
123456789012345678901234567890123456789012345678901234567
EPE SG FDRPETRPR E QGRRFQGEDKRGYYGVF GGG Y G
GRD PEL G FDRP TRPRYEEPQGRR QGED RGYV VFEGGG Y GGRGQ
D E K GDTR P R EPQ RRFQG DKRGYYG EGGGRYDLGGR

Some residues are invariant in one or two phyla, highlighting the fact they cannot be treated as a single large all-archaeal enzyme. Where three phyla display invariance, then universality is functionally significant. For histidine synthetase, they are: E^{CH7}, G^{CH11}, P^{CH18}, R^{CH23}, Q^{CH28}, R^{CH30}, R^{CH31}, Q^{CH33}, G^{CH34}, D^{CH36}, R^{CH38}, G^{CH39}, Y^{CH40}, Y^{CH41}, G^{CH46}, G^{CH47}, G^{CH48}, Y^{CH50}, G^{CH53}, with one-letter symbols for

each amino acid followed in superscript by a lowercase ‘c’ for consensus plus (in this case) capital ‘H’ signifying HisRS plus condensed version position number.

The compacted consensus format can be readily extended to compare invariant residues across multiple enzyme types. For example, ArgRS and CysRS are Class I synthetases, and the literature alleges that all synthetases in this group display invariant HIGH and KMSKS peptides within their full sequence (Eriani et al., 1990). Ten years later, this statement was repeated in the review by Woese et al., (2000) and twenty years after that, still adhered to in an updated review penned by Gomez & Ibba (2020). If this thesis is intended literally, it is patently false; if allowance is made for mutation in one or more positions, then interpretive expansion weakens any impact the proposition possesses. The consensus condensed version for archaeal phyla proves the belief invalid.

ArgRS (H/M/T)

00000000011111111122222222233333333344444444455555555566666666677777777788888
1234567890123456789012345678901234567890123456789012345678901234567890123
PN N PHGRN D GDGQ Y E EY RD A LP G M TRG A F QYRILA NFYVRLGM
D NEHTSNPPHGRNGDR DGQ YE E R DGT YDYHK G G S RG RKFWLF YR F Y R G
NE SNP HGRNG D QKD Y ES R G YDY QQHYV M RGD YR N Y G

CysRS (H/M/T)

00000000011111111122222222233333333344444444455555555566666666677777777788888
12345678901234567890123456789012345678901234567890123456789012345678901234
CG HG R DLG NTD Y VYGL DFL P PW D GG D HH WH KMS SNR
CGT DHGHRFD VNTDDDKII Y YF GLKDF WKWSPGGRPWHIEY DHGGGDLFPHEQ WHGKMSKSNR I
CGTYDHGH D NTDD K LAYGY G KD W G PWH G HGG L PHHE AWH KM KSNRYD

Concentrating on ArgRS and CysRS, of the six sequences, HxGy appears (H^{cR12} , G^{cR13} , H^{cC6} , G^{cC7}) in all, only two possess a second unchanging histidine (H^{cC8}), and none contain conserved isoleucine. When looking at individual rather than consensus sequences, HxGy tetrapeptide may be expanded to show x is chosen from I / L / M / V and y from A / H / R / S. These permutations probably represent distinctions without biological significance for enzyme function, but that fact does not contravene repudiation of the HIGH invariance dogma. KMSKS is recognizable within CysRS (condensed residues 69–73), but ArgRS only hints at its existence (M^{cR56} , S^{cR57}). An unexpected invariant amino acid (R^{cR59}) in place of the second serine should be noted because it does not match expectations from *E. coli*, upon which the original claim was staked. In contrast to HIGH, KMSKS pentapeptide is part of a loop region claimed to be critical for stabilizing the transition state in aaRS/aa/ATP complexes leading to acylamino adenylate formation (First & Fersht, 1995). Its omission in ArgRS is relevant.

The remaining Class I synthetases, except for LysRS for which no consensus residues were found (Table 4), also display deviance from the ‘universal’ pattern: *no consensus sequence* encodes HIGH in full, and *no consensus sequence* except the CysRS Methanobacteriota set just shown encodes KMSKS in full. This failure encompasses 3000 archaeal aaRS (3626 – 626 amidotransferases; Table 2, column ‘n’), A full account by phylum is presented in Supplementary Table 7. That these Archaea differ in this regard not only from *E. coli* but amongst themselves is a prime example of *adaptive radiation* substantiated by the breadth of physical characteristics displayed in Table 1.

With respect to Class II synthetases, Carter’s review (Carter Jr, 1993) mentioned three patterns extrapolated from a limited amount of data, yet declared characteristic for these enzymes. All information derives from *E. coli* primary sequences and/or crystal structures of bacterial complexes for aaRS plus translation-relevant ligands. The shorthand used below comes from Carter: + means

charged (Arg, Asp, Glu, His, Lys), φ represents hydrophobic (Ile, Leu, Met, Phe, Trp, Tyr, Val), x stands for any other (Ala, Asn, Cys, Gln, Gly, Pro, Ser, Thr).

- motif 1: +G(F/Y)xx(V/L/I)xxP $\varphi\varphi$ $\rightarrow$ Permutations = $5 \times 1 \times 2 \times 8 \times 8 \times 3 \times 8 \times 8 \times 1 \times 7 \times 7 = 6,021,120$
- motif 2: + $\varphi\varphi\varphi\varphi$ xxxFRxE $\rightarrow$ Permutations = $5 \times 7 \times 7 \times 8 \times 7 \times 8 \times 8 \times 8 \times 1 \times 1 \times 8 \times 1 = 56,197,120$
- motif 3: φ G φ G φ G φ ER $\varphi\varphi\varphi\varphi$ $\rightarrow$ Permutations = $7 \times 1 \times 7 \times 1 \times 7 \times 1 \times 7 \times 7 \times 1 \times 1 \times 7 \times 7 \times 7 = 40,353,607$

To give some perspective, according to Internet sources, the most common full name in the world is Zhang Wei possessed by ~290,000 Chinese; some version of Muhammad (variable spelling) is said to be the most universal given name with an estimated 150 million people using it; those of Chinese heritage also claim the most common surnames with Li/Lee or Zhang possessed by over 100 million individuals. If all 290,000 Zhang Wei (much less all Muhammad or Li/Lee) displayed identical signatures, would they be presumed to possess the same behavioral traits?

The key question, of course, is: how do differences in sequence affect translation? With respect to supposed Class II synthetase motifs, could possible permutations for each pattern be alleged to produce aminoacylation kinetics values for k_{cat} and/or K_M sufficiently close among themselves to declare plausible both overall functional and detailed mechanistic interchangeability? The question, intended seriously, is actually rhetorical since Table S7 makes the issue moot: consensus sequences from three archaeal phyla do not adhere to these suggested patterns. In many cases, no motifs can be discerned at all. If motif 2 is abbreviated to FRxE, it makes the most convincing case because it or some version definitively appears nineteen times out of thirty-three (eleven aaRS $\times$ three phyla). Motif 3 appears seventeen times, if license is taken on how variable GxGxGxxER might be realized.

Amino acids in proteins are invariant across many species for one or more reasons: (1) induce secondary or tertiary structure for the molecule in isolation; (2) stabilize a bioactive 3D conformation; (3) facilitate function by direct interaction with other agents in vivo. Phylum-based differentials in identity of conserved residues might be insignificant artifacts. Alternatively, they lead to variation in binding affinity for tRNA, amino acid charged to its cognate tRNA, or ATP needed to facilitate aminoacylation, thereby directly affecting catalytic rates for that organism to survive and prosper in its natural environment. It is logical to hypothesize that amino acids found invariant in all three phyla can be considered candidates for all three options, with particular emphasis on the precise role outlined for translation-related processes. In contrast, perhaps unchanging residues in ≤ 2 phyla are relegated to acting in those first two capacities developed specifically for adaptation to their internal biochemistry and native habitat.

Variation in conserved amino acids among Halobacteriota, Methanobacteriota, Thermoproteota is detailed in Supplementary Table 8 using the condensed format. A count for number of unchanging residues in both synthetases and amidotransferases found in all three phyla, as opposed to just one or two, yields the data in Supplementary Table 9; 838 residues total. The numbers in parentheses indicate their distribution:

- high frequency $\rightarrow$ Gly (149), Arg (114), Pro (85), Glu (77), Asp (68)
- intermediate frequency $\rightarrow$ Tyr (47), His (38), Gln (33)
- low frequency $\rightarrow$ Phe (29), Ser (28), Trp (27), Thr (26), Lys (24), Leu (21), Asn (20)
- rarely $\rightarrow$ Cys (18), Ala (13), Val (10), Met (8), Ile (3)

Although any residue can interact with added substrates through hydrogen bonding activity using backbone NH or C=O functionality (Kaiser et al., 2018), glycine has a unique attribute: it can extend a chain to enhance positioning of other residues without incurring spatial consequences demanded by fitting a sidechain into an otherwise crowded region. Arginine as a source of positive charge to counter negative charge on ATP phosphate or incoming amino acid carboxylate has a crucial advantage over similarly basic lysine, namely multiple hydrogen-bonding capability due to its guanidinium moiety. Proline, like glycine, is unique: it forces a protein chain to alter spatial direction leading to preferred final 3D shape. The two acids serve roles opposite to that of arginine by opposing otherwise unbalanced positive charge incurred by addition of externally-sourced metal

cations as well as by producing salt bridges with indigenous Arg, His, Lys to stabilize bioactive conformations.

Salt bridge establishment is critically important for sodium chloride-loving halophiles due to the presence of an overabundance of cations, as well as to thermophiles because thermodynamic stabilization via ionic bonds is much stronger energetically than hydrogen bonding, meaning they are more resistant to temperature stresses. Consequently, when halophiles and thermophiles are also acidophiles (Tables 1, S2), successful salt bridge formation is substantively mitigated, if not eliminated altogether, because sidechain carboxylates in aspartic and glutamic acids are protonated to their conjugate forms. Competing effects are complications for aminoacylation kinetics not usually addressed in conventional experiments (Francklyn et al., 2008).

Among rare invariant amino acids common to these Archaea, cysteine is the most intriguing: like proline, it can potentially have a major impact on overall protein shape, but it requires the presence of a suitably-positioned cysteine partner to form an internal disulfide bond. Using Table S8, one can evaluate whether an even number of cysteines are present in consensus sequences for any studied enzyme. This, of course, does not prove existence of a disulfide linkage; it simply demonstrates that the first criterion (even number) has been met.

In principle, Table S8 can be used to generate a Table S9 analog for invariant residues encoded in only one or two of the three phyla, but this has not been done. Experimental structure/function investigations are needed to determine their exact role; how it differs from their all-phyla cousins; why they are conserved in some, but not all, Archaea. One speculative idea: identity differentials for synthetases might enable other reagents to bind in order to produce pre-translationally modified amino acids subsequently incorporated into proteins during translation (Laibelman, 2024).

In combination, Tables 4, S8, S9 convey a problem: some aaRS, despite producing reasonable quantities of invariant amino acids in each separate phylum (Table 4), yield unacceptably low consensus numbers when integrating these results (Tables S8, S9). The issue was raised with an earlier discussion referencing LysRS, but single digit values are also obtained for LeuRS (eight consensus invariants across all three phyla), ThrRS (two), TrpRS (seven). Table S6 shows these four encode the most multiple genes possessing compositions and sequences different from single gene Archaea within the same phylum. This connection is not a coincidence: heterogeneity in strings necessarily implies fewer invariant residues upon alignment trials.

There are reports of divergent gene forms for LeuRS within Halobacteriota (Andam et al., 2012; Fang et al., 2014; Weitzel et al., 2020), and of wider-ranging taxonomic scope for archaeal ThrRS (Beebe et al., 2004; Korencic et al., 2004). Weitzel et al. stated that one version lacked tRNA aminoacylation activity, yet could still bind it and produce leucyl adenylate. They speculated the activated tRNA might be useful in chemical modification of other amino acids. As possible precedent for this suggestion, Wan et al. (2014) showed the joining of two lysine amino acids were responsible for the biosynthesis of pyrrolysine. Crystal structures of ThrRS from several archaeal organisms with two versions show one form possessing catalytic activity for aminoacylation while the other is used exclusively to edit tRNA erroneously charged with noncognate amino acids (Shimizu et al., 2009).

Organisms specified in these various publications for multiple forms of LeuRS and ThrRS were subjected to alignment tests, with high numbers of unchanging amino acids obtained for the former, but lower for the latter to the point where practically no invariant residues were produced for Thermoproteota ThrRS strings (Table 4). Regardless of alignment results, the problem is the same as for lysyl-synthetase: highlighted species in these documents do not cover anywhere close to the full complement of Archaea known and characterized in culture, and for whom sequences are now available from database archives. Without new research, analysis of invariant residues within and between phyla for these aaRS will limit complete understanding of the effects of gene multiplicity on proofreading function or aminoacylation mechanism(s).

Perusal of Table 4 also reveals a paucity of unchanging residues for amidotransferase subunit C (≤ 4). Unlike synthetases, this is not a troublesome outcome given the brevity in length (71–111 amino acids) and its functional role rendering assistance to subunit A in producing and/or transporting

ammonia (Nakamura et al., 2006). As long as a subunit C sequence permits binding to a related subunit A molecule at a minimally effective level evoking successful accomplishment of their joint goal, then a genetic match to other C-type subunits is irrelevant. The same remark could theoretically be uttered for synthetases, i.e., only a match to its cognate tRNA is vital for function, but the fact that all aaRS must also bind identical ATP molecules means there must be greater sequence and 3D structure held in common across all types.

Table 4 contains a category called *direct comparison*, under which are found two sets of alignment trials. The first establishes commonalities and differences between a prototypical selective AspRS version and its nondiscriminatory counterpart Asn_AspRS; the second contrasts conventional seryl-synthetase with alternate SerRS2. Although the latter possesses just twenty-seven member species for certain (Table 2), it may be viewed as a representative analog of what is sought for diverse leucyl, lysyl, threonyl, tryptophanyl synthetase types because none have, as yet, twenty-seven known archaeal organisms in their membership lists.

There are noteworthy observations (Table S3) to make before direct comparison between AspRS and Asn_AspRS is undertaken:

- no Halobacteriota encode AspRS; all utilize the nondiscriminatory gene. Justification for this fact is not obvious: why should preference for high salt concentration compromise binding selectivity between aspartic acid and asparagine for charging cognate tRNA?
- if species possess genes for AsnRS and AspRS, they do not need nondiscriminatory Asn_AspRS enzyme. *Korarchaeum cryptofilum*, *Nitrososphaera gargensis*, *Nitrososphaera viennensis* apparently overcompensate by encoding the nonselective type as well.
- the logical corollary to the second note should hold: if genomes lack both AsnRS and AspRS, they must utilize Asn_AspRS. This thesis does in fact apply to all Archaea studied.
- absence of AsnRS should automatically cause genomes to include a gene for nondiscriminatory Asn_AspRS if possessing AspRS. This is valid for *Cenarchaeum symbiosum*, *Methanomassiliicoccus A intestinalis*, *Methanomethylophilus alvus*, *Nitrosopumilus maritimus*, but not for *Aeropyrum camini*, *Hyperthermus butylicus*, *Pyrolobus fumarii*. How they are able to charge tRNA^{Asn} is a mystery.
- absence of AspRS should automatically cause genomes to include a gene for nondiscriminatory Asn_AspRS if possessing AsnRS. This is valid for *Methanocella conradii*, *Methanocella paludicola*, *Methanocella. A arvoryzae*.

To directly compare all archaeal sequences from AspRS and Asn_AspRS, several steps had to be performed. First, separate alignments were conducted for each synthetase. From Table 3, strings from *M. A intestinalis*, *M. alvus*, *K. cryptofilum* were removed for cause from Asn_AspRS, and from the first note, only two alignments are possible for AspRS. Second, all gaps inserted by Clustal Omega were eliminated, as were nonconserved residues; the resulting condensed format left five invariable residue strings. Third, the pair from AspRS (Methanobacteriota, Thermoproteota) were aligned manually, as were those from the three phyla for Asn_AspRS; these representations can be viewed in Table S8. Fourth, this information was reorganized by relating conserved Methanobacteriota sequences in condensed format for the two aaRS to each other, and performing the same action for the corresponding Thermoproteota strings. Fifth, these sets were manually realigned again in order to identify common invariant amino acids in each phylum. Sixth, these sets of consensus residues were restructured into the condensed format for easy depiction, which is shown below. All stages of the direct comparison process are contained in Supplementary Table 10.

AspRS + Asn_AspRS consensus M/T
0000000001111111112222222222333333333344444444445555555555666666666677777777778888888888999
1234567890123456789012345678901234567890123456789012345678901234567890123456789012
GWFRDQG EGELPVTRDRF PKEGGL FYFALQSPQLYKGEFRAETRHESDEY GDPP FYFDLESGRHLFYAFGM
PHGGRNREFPRDRP

GW GGE P RFTPE G A Y ALQ PQ R E T H E E D D FFYDLE GR L G PPHGGR RD P

As anticipated based on function, many invariants are found, with more in methanogens than in (hyper)thermophiles. The final condensed format reveals forty of ninety-two are conserved in both types of synthetases. Class II's motif 2 FRxE is visible as R^{c44} and E^{c46}, while motif 3 GxGxGxxER can be seen as G^{c81}, G^{c82}, R^{c83}. How invariant residues found in only one phylum relate to kinetics, thermodynamics, or mechanism of aminoacylation has not been explored, but would make an informative study.

From a different perspective, one would suppose all organisms using Asn_AspRS would encode amidotransferase subunits A–C in order to produce asparaginyI-tRNA^{Asn}. Table 2's 'n' column indicates numbers of sequences = 113, 115, 117, 92 for Asn_AspRS, subunit A, subunit B, subunit C, respectively. It is vital to recognize that encodings do not equal species: multiple nonselective synthetase genes exist for two species and a duplicate gene for subunit B is present in one species (Table S6). Reflecting this state, the actual number of species involved are 111, 115, 116, 92, respectively.

The notes for direct comparison between AspRS and Asn_AspRS mentioned *K. cryptofilum*, *N. gargensis*, *N. viennensis* contain copies of the extraneous nondiscriminatory synthetase even though it is not needed because all three also possess selective versions (AsnRS, AspRS). It turns out that the two *Nitrososphaera* species retained genes for amidotransferase subunits A–C, presumably fully functional, as well, but *K. cryptofilum* no longer contains associated subunits, so its Asn_AspRS sequence is not viable. The last point is moot, however, because Table 3 shows that same Asn_AspRS gene product was removed from alignment analysis due to a high level of difference in composition from other synthetase sequences of that type, as indicated by the phylogenetic tree generated when it was included. The count is now 110, 115, 116, 92, respectively.

Hyperthermus butylicus is missing a subunit A gene according to UniProtKB. Along with *A. camini* and *P. fumarii*, these three lack selective AsnRS or indiscriminate Asn_AspRS as well, and “how they are able to charge tRNA^{Asn} is a mystery” has also been expressed. That riddle can now be both deepened and somewhat clarified: *H. butylicus* possesses genes for subunits B and C, while subunits A–C are found in *A. camini* and *P. fumarii*. Thus, they have the ingredients to form asparaginyI-tRNA^{Asn} without the requisite synthetase to catalyze it, or so it seems. The count is 110, 113, 113, 89, respectively

Just as *K. cryptofilum*, *N. gargensis*, *N. viennensis* do not need Asn_AspRS or amidotransferases because they possess selective AsnRS and AspRS, the same is the case for *Acidolobus saccharovorans*, *Caldisphaera lagunensis*, *Fervidicoccus fontis*. Genes for subunits A and B are in their genomes, but not those for Asn_AspRS or subunit C. Perhaps these sequences constitute remnants from a past evolutionary time period when they were necessary, but, though still present, are no longer functionally expressed. The count is now 110, 110, 110, 89.

An appeal to gene loss also reconciles the remaining gap of twenty-one species lacking subunit C while possessing the other components. The claimed purpose for subunit C is to assist subunit A in NH₃ transfer from asparagine to aspartyl-tRNA^{Asn} (Nakamura et al., 2006; Sheppard & Söll, 2008). Subunit C attaches to subunit A and has no direct interaction with subunit B or tRNA (Curnow et al., 1997). Except for Thermoproteota *Aeropyrum pernix*, *Cenarchaeum symbiosum*, *Ignisphaera aggregans*, the remaining eighteen devoid of subunit C are all from Methanobacteriota phylum. Such a large number suggests some Archaea have evolved to a state for which they do not need subunit C. Mechanistic details for ammonia transfer in Methanobacteriota may have diverged from what they are in Halobacteriota.

With respect to amidotransferase subunits, without regard for participation by synthetase, attempts to uncover invariant residues in alignments between subunits A and D, or subunits B and E, met with failure. Despite near-identical biochemistry—amide hydrolysis of asparagine and glutamine, respectively, for the first pair; phosphorylation of sidechain carboxylates in aspartate and glutamate, respectively, for the second—there appeared to be no common conserved amino acids in the same relative chain position. This held true for all three phyla. It seems unlikely that difference

in substrate (Asp *versus* Glu) on which these chemical transformations are intended is sufficient to cause thoroughgoing alteration in subunit sequences along their entire length.

Validating this hypothesis of similarity in some respect, crystal structures depict a very different situation. Subunits D and E from *Pyrococcus abyssi* co-crystallized as a heterotetramer of composition $\alpha_2\beta_2$ (Schmitt et al., 2005). When folded into an enzymatically active state, residues surrounding the region seen to bind ATP and the terminus of Glu-tRNA^{Gln} (subunit E) or Asp-tRNA^{Asn} (subunit B) are the same in 3D space though distantly placed in their respective primary sequences: in other words, structural conservation without sequence homology.

In contrast to this lengthy exposition, exploring the relationship between SerRS and SerRS2 holds no surprises (Table S10). Seryl-synthetase sequences, minus *Methanosphaerula palustris* (Table 3), and SerRS2 were aligned. Halobacteriota and Methanobacteriota phyla only are in play, since SerRS2 does not encode strings from any species belonging to Thermoproteota. As shown in Table 4, eighteen residues in the first phylum and sixteen in the second are invariant. This outcome was supported by the accompanying phylogenetic tree displaying separate clusters for these two enzymes with no overlap. It is observed that species from genus *Methanosarcina* lack uniformity in their genomes:

- genes for SerRS and SerRS2 → *M. barkeri* B, *M. horonobensis*, *M. vacuolate*;
- gene for SerRS only → *M. acetivorans*, *M. lacustris*, *M. mazei*, *M. siciliae*;
- gene for SerRS2 only → *M. thermophila*.

Why this distribution exists in a single genus is unknown; the sole exceptional cell culture growth factor is a ten-degree higher optimal temperature parameter for *M. thermophila* (Table S2). Using Clustal Omega, multi-sequence comparison leads, in condensed format notation, to finding nine amino acids in common between the phyla: P^{c4}, P^{c7}, G^{c12}, R^{c13}, F^{c15}, E^{c16}, S^{c23}, R^{c24}, P^{c25}. This outcome contrasts greatly to that derived from a direct comparison between selective AspRS and nondiscriminatory Asn_AspRS where numerous invariant amino acids are recorded across archaeal phyla.

SerRS + SerRS2 consensus H/M
0000000001111111122222
1234567890123456789012345
P PTEP HGR FEG DEPESRP
GLP PREG GRVFE Y SRP

Part II. Transfer RNA

A prior investigation (Laibelman, 2022) concentrated on demonstrating ways in which archaeal tRNA was composed of unique sequences for isodecoders within each amino acid-associated isoacceptor group. One conclusion: the standard genetic code does not represent redundant (degenerate) isoacceptors. It is crucial to turn to the converse topic: how sequences from tRNA contain invariant nucleotides, and what this conservation signifies with respect to aminoacylation kinetics and mechanism(s).

Of the 186 archaeal organisms studied before, many referenced strains of the same species. After purging these excesses to leave one member from each, and matching one-for-one species researched for examination of synthetases and amidotransferases, 131 Archaea provided information for the three kinds of molecules. Table S4 denotes this collection by incorporating an asterisk (*) beside their name; the other nineteen lack corresponding tRNA sequence data. On hand are 6063 tRNA strings divided into:

- Halobacteriota → 56 species, 2789 tRNA;
- Methanobacteriota → 37 species, 1541 tRNA;
- Thermoproteota → 38 species, 1733 tRNA.

Since the length in number of nucleotides of tRNA sequences (~70–90) is drastically less than for enzymes counted in amino acids (303–1143, Table 2), the raw data is organized a little differently. Instead of arrangement by type (AlaRS, ArgRS, etc.), Supplementary Table 11 classifies them by phylum first and by amino acid second with, for example, all tRNA^{Ala} isoacceptors for each phylum-related species (listed alphabetically) grouped together, followed by all tRNA^{Arg} isoacceptors, and so on.

Some species encode strings both shorter and longer than typical lengths for their type. Transfer RNA for leucine and serine usually possess 81–91 nucleotides, while the eighteen other tRNA associated with canonical amino acids habitually contain 70–80 nucleotides. Included Archaea with tRNA lengths outside these boundaries are, with a single exception, obtained from a second gene paired with a normal-length version for the same anticodon. *Natrinema pellirubrum* is the sole deviant from this pattern; it has a single novel tRNA^{Gly(ACC)} isoacceptor of eighty-two nucleotides.

Sequences longer than expected for amino acid type could be: (1) authentic biomolecules with in vivo activity; (2) transcription errors induced by read-through of Stop codons; (3) processing errors caused by still-incorporated introns (Popow et al., 2012; Fujishima & Kanai, 2014), though GtRNAdb from which they were retrieved alleges downloads are ‘mature’ tRNA. Sequences shorter in length than anticipated are more difficult to explain; possibilities include: (1) authentic biomolecules with in vivo activity; (2) post-transcriptional fragments caused by kinetic or thermodynamic instability within organisms; (3) researcher-produced error failing to sequence the full gene length; (4) archive error during information retrieval or file corruption during storage.

In total, there are thirty-five sequences outside this 70–91 nt range (out of 6063, 0.58%): eighteen Halobacteriota, five Methanobacteriota, twelve Thermoproteota; fourteen < 70 nt and twenty-one > 91 nt. *Sulfolobus acidocaldarius* supplied four strings, *Archaeoglobus A sulfaticallidus* three, *Natronobacterium gregoryi* and *Pyrobaculum oguniense* two each; the remaining two dozen were provided by twenty-four individual species. A trend does emerge upon consideration of amino acid type and specific isoacceptor: 21 of 35 (60%) derive from leucine (ten), serine (seven), glycine (four); ten amino acids yield 40% of the total; six of leucine’s ten coincide with CAA anticodon and five of serine’s seven have anticodon CGA. What these tendencies reveal is unknown. Designations ‘1’, ‘2’ have no significance other than indicating the order in which multiple copies were processed after download. Table 5 provides the complete roster.

Inclusion of abnormal length tRNA affects number and identity of invariant nucleotides during multi-sequence alignment, and their removal evoked a disparity of consequences: 1–13 additional invariant nucleotides per tRNA consensus sequence were recovered. These results by themselves neither vindicate nor repudiate any suggestion that atypical length sequences should, or should not, be ignored. The single legitimate approach to determining their inclusion in multi-sequence alignments is conveyed by the phrase *authentic biomolecules*. Aminoacylation kinetics data alone resolves the issue. Can these unusual tRNA bind to their cognate synthetase with K_M values approximately equal to their normal-length counterparts from the same species? Upon binding, can they also be charged with k_{cat} values nearly equal to their normal-length counterparts from the same species?

Multiple sequence alignments using Clustal Omega were performed with and without those from Table 5. In Table 6, ‘n’ is the number of strings for each tRNA type; arrows (→) indicate with → without Table 5 sequence inclusions.

Table 5. Short and Long Length tRNA.

Halobacteriota short sequences	Methanobacteriota short sequences
Archaeoglobus A sulfaticallidus Leu(CAA)2 [55 nt]	Methanocaldococcus vulcanius Ser(GCU)2 [70 nt]
Geoglobus acetivorans Leu(UAA)2 [77 nt]	Thermococcus γ-tolerans Ser(CGA)2 [53 nt]
Halorhabdus tiamatea Gln(UUG)2 [64 nt]	Thermococcus A litoralis Leu(CAA)2 [76 nt]
Methanosarcina lacustris Ser(CGA)2 [57 nt]	Methanobacteriota long sequences

Methanothrix soehngensis Leu(UAG)2 [64 nt]	Pyrococcus yayanosii Gln(UUG)2 [90 nt]
Natronobacterium gregoryi Ala(CGC)2 [66 nt]	Thermococcus kodakarensis Arg(UCU)2 [83 nt]
Natronobacterium gregoryi Ser(CGA)2 [61 nt]	Thermoproteota short sequences
Halobacteriota long sequences	Fervidicoccus fontis Leu(CAA)2 [70 nt]
Archaeoglobus A sulfatallidus Val(UAC)2 [108 nt]	Sulfolobus acidocaldarius Asp(GUC)2 [68 nt]
Archaeoglobus A sulfatallidus Arg [GCG]2 [85 nt]	Sulfolobus acidocaldarius Leu(CAA)3 [68 nt]
Haloferax mediterranei Thr(GGU)2 [88 nt]	Sulfolobus acidocaldarius Ser(CGA)1 [74 nt]
Halogeometricum borinquense Glu(UUC)2 [90 nt]	Thermoproteota long sequences
Halomicrobium mukohataei Trp(CCA)2 [98 nt]	Aeropyrum pernix Leu(CAA)2 [128 nt]
Halostagnicola larsenii Thr(GGU)2 [102 nt]	Desulfurococcus amylolyticus Gly(UCC)2 [127 nt]
Halovivax ruber Gly(GCC)2 [84 nt]	Metallosphaera sedula Ser(CGA)2 [96 nt]
Methanococcoides methylutens A Leu(UAG)2 [127 nt]	Pyrobaculum aerophilum Met(CAU)2 [109 nt]
Natrialba magadii Gly(GCC)2 [84 nt]	Pyrobaculum oguniense Leu(CAG)1 [125 nt]
Natrinema pellirubrum Gly(ACC) [82 nt]	Pyrobaculum oguniense Val(UAC)1 [108 nt]
Natronomonas pharaonis Pro(UGG)2 [105 nt]	Saccharolobus solfataricus Ser(GGA)1 [109 nt]
	Sulfolobus acidocaldarius Leu(CAA)2 [97 nt]

Table 6. Numbers of Invariant Nucleotides in tRNA.

tRNA	All Archaea		Halobacteriota		Methanobacteriota		Thermoproteota	
	n	Invariant nt	n	Invariant nt	n	Invariant nt	n	Invariant nt
Alanine	434	9 → 11	212	20 → 31	107	25	115	19
Arginine	625	4 → 5	274	16	158	10 → 11	193	11
Asparagine	145	15	65	38	41	24	39	27
Aspartic Acid	157	15 → 22	75	40	43	28	39	32 → 38
Cysteine	197	11	119	20	40	25	38	21
Glutamic Acid	269	15 → 18	123	22 → 26	70	31	76	44
Glutamine	258	8 → 13	118	13 → 26	66	18 → 24	74	31
Glycine	401	6 → 12	191	16 → 19	94	23	116	17 → 21
Histidine	131	16	56	24	37	33	38	26
Isoleucine	143	15	66	26	38	26	39	38
Leucine	657	0 → 3	304	2 → 15	154	9 → 13	199	6 → 17
Lysine	257	21	115	26	67	32	75	33
Methionine	283	11 → 13	126	19	78	30	79	18 → 19
Methionine, elongator	135	12 → 15	58	22	37	44	40	25 → 27
Methionine, initiator	148	36 → 38	68	52	41	45	39	42 → 44
Phenylalanine	145	19	68	29	39	29	38	44
Proline	365	14 → 13	168	17 → 19	84	24	113	23
Serine	525	6 → 8	228	7 → 13	143	9 → 14	154	12 → 23
Threonine	402	3 → 3	186	8 → 9	100	22	116	25
Tryptophan	133	16 → 21	57	37 → 49	38	25	38	39
Tyrosine	135	16	59	40	38	24	38	38
Valine	401	4 → 8	179	16 → 23	106	12	116	10 → 23

6063 2789 1541 1733

Table 6 includes all invariant nucleobases within the anticodon triplet. Since this varies from zero for all tRNA^{Ser} up to three for several tRNA isoacceptors, an accurate tally requires subtraction of anticodon triplets from totals shown in order to put all tRNA on an equal footing. Nonetheless, as was true for Table 4 data, invariant numbers of nucleotides, like invariant numbers of amino acids, are unequal across phyla. The ‘all Archaea’ tabulation yielding quantities less than those from each phylum supports the notion that preferred conditions promoting cell growth in culture probably impacts innate genetic characteristics, as does the reverse causal direction.

Since lysine synthetase is said to possess both Class I and Class II types, it might be supposed that the corresponding tRNA^{Lys} would likewise demonstrate a distribution between classes through sequence differences. There is, here, no evidence for this proposal. On the contrary, the three phyla together display high levels of invariant nucleotides: twenty-one positions are conserved, whereas for tRNA^{Val} associated with class I ValRS just eight unchanging nucleotide positions exist, and for tRNA^{Ala} linked to class II AlaRS only eleven such locations. Whatever evolutionary properties caused lysyl-synthetase division into two classes, those same parameters did not induce a similar cleavage among tRNA molecules. This constitutes another piece of evidence that tRNA and their synthetases did not co-evolve.

Fewer than twenty invariant nucleotides are present in consensus sequences for all three phyla for tRNA^{Arg} and tRNA^{Leu}. LeuRS had low numbers, where it was attributed to an incomplete species tabulation leading to an inability to adequately divide all known species into types said to exist (Andam et al., 2012; Fang et al., 2014; Weitzel et al., 2020). Also comparatively heterogeneous in composition among their respective species are tRNA^{Ser} from Halobacteriota and Methanobacteriota, tRNA^{Thr} from Halobacteriota, tRNA^{Val} from Methanobacteriota. Besides sequence inhomogeneity for taxonomic phyla pertaining to types of aaRS and tRNA, this characteristic trait extends to ribosomal proteins (Brochier et al., 2005; Brochier-Armanet et al., 2008).

Separate alignments for elongator methionine (eMet) tRNA and initiator methionine (iMet) tRNA display large differences in invariant nucleotide number between them for species from Halobacteriota and Thermoproteota phyla, but not for Methanobacteriota. In addition, the quantity of conserved nucleotides for all archaeal tRNA^{iMet} strings is comparatively huge; no other tRNA isoacceptor group comes close to that total of thirty-eight.

Numbers are interesting and worthy of speculation regarding possible biological explanation, but identity is of far-reaching consequence for lifeforms. Given the brevity of tRNA length, few gaps are added by the algorithm during alignments. The entire length of consensus strings for invariant nucleotides can be retained, meaning abridgment into the condensed format introduced for synthetases and amidotransferases is unnecessary, allowing exact position information to be determined. Supplementary Table 12 provides consensus tRNA sequences for these phyla containing identity and position location information pertaining to invariant nucleotides. Upon inspection of Table contents, the first notable aspect is the variation in length not only *between* types of tRNA, which is obvious since tRNA^{Leu} and tRNA^{Ser} are known to exceed the other eighteen canonical amino acid isoacceptors in this feature, but *within* each type: Halobacteriota, Methanobacteriota, Thermoproteota species do not possess the same number of nucleotides in their strings for a given tRNA type. The most economical way to detect this is to focus on the anticodon triplet. One example from Table S12 suffices to make the point, with the anticodon triplet written in red. Unique for these studied Archaea, tRNA^{Asn} from Halobacteriota possess a uniform seventy-three nucleotides for every species; tRNA^{Asn} from Methanobacteriota are 71–77 nt and tRNA^{Asn} from Thermoproteota are 70–78 nt (Table S11).

tRNA Asn H/M/T
000000000111111111222222222233333333334444444444555555555566666666667777777777
12345678901234567890123456789012345678901234567890123456789012345678
GCC UA CU AG UGGUAGAG GUUA G C AGGUUCGA CCU GGCG
G U A GU A C G CUGUUAAC GUUC A C CG

C UAGCU AG GG AGC G CUGUUA GG GU C C

This example offers more detail. The second row (Methanobacteriota) begins to diverge from the first row (Halobacteriota) at position 23→24 because adenosine shifts one column rightward. Movement is carried beyond the anticodon triplet through the remainder of the string, observable at positions 53→54 and 72→73. Thermoproteota species (third row) display different positional adjustments; they diverge from Halobacteriota beginning at 18→19, undergoing a second shift at 23→25, resulting in a total adjustment of two nucleotide positions at the anticodon triplet. This change is carried through to positions 53→55, then a third alteration occurs for the last cytidine as seen by movement at 62→65. If desired, similar analyses can be undertaken for other tRNA using Tables S11 and S12. All shifts are traceable to specific nucleotide insertions or deletions, depending on perspective; it is conceivable environmental adaptation is responsible for genetic alterations.

Inspection of Table S12 reveals it is a partial truth that for Archaea “U⁸, A¹⁴, G¹⁸, G¹⁹, A²¹, U³³, G⁵³, T⁵⁴, Ψ⁵⁵, C⁵⁶, A⁵⁸, C⁶¹, C⁷⁴, C⁷⁵, and A⁷⁶ are the most conserved nucleotides within tRNA” (Biela et al., 2023). Here, T⁵⁴ and Ψ⁵⁵ are treated as uridines because it is uncertain these transformations transpire in all, or many, of these 131 species. Taking the list from Biela *et al.* in order, with tRNA^{eMet} and tRNA^{iMet} counted separately, there are sixty-three possible consensus sequences. After consideration of positional shifts for each nucleotide, Table 7 is obtained.

Table 7. Number of Consensus Sequences with Missing “Conserved” Nucleotides.

Nucleotide	H	M	T	Nucleotide	H	M	T	Nucleotide	H	M	T
U ⁸	4	19	0	U ³³	5	6	9	A ⁵⁸	2	0	3
G ¹⁰	13	7	4	G ⁵³	8	1	0	C ⁶¹	9	3	4
A ¹⁴	2	0	2	T ⁵⁴	1	1	4	C ⁷⁴	21	21	21
G ¹⁸	11	9	10	Ψ ⁵⁵	5	11	8	C ⁷⁵	21	21	21
G ¹⁹	8	9	11	C ⁵⁶	1	0	1	A ⁷⁶	21	21	21
A ²¹	9	16	18								

Many of these nucleotides are alleged to be vital for internal folding of tRNA into its ultimate 3D *quasi*-L shape (Tamaki et al., 2018); *quasi* because the angle between anticodon triplet at one end and 3'-acceptor nucleotide at the other end deviates from perfect perpendicularity depending, in large part, on length and nucleobase composition in the variable loop. Compared to conformation-critical nucleotides mentioned by Biela et al., Tamaki et al. added G¹⁰, removed A²¹, ignored 3'-CCA; Table 7 includes all.

By simple arithmetic, seventy-eight violators in Halobacteriota, eighty-two in Methanobacteriota, seventy-four in Thermoproteota are found, discounting 3'-CCA. The approximate L-shaped configuration must be maintained in order to guarantee complementary base-pairing with mRNA at the ribosomal A-site, so their absence in a tRNA sequence from any species must lead to one conclusion: unless Clustal Omega has performed poorly in alignments, substitute nucleobases suffice. Nucleotide insertions and/or deletions causing shifts in relative positioning between them, even if the 'critical' nucleotides are present, does not induce a disastrous impact on the translation process. At worst, the L-shape varies within acceptable limits. This, in turn, suggests there is a degree of flexibility in the spatial co-existence of rRNA, mRNA, tRNA, ribosomal proteins.

One cannot fail to notice the absence of 3'-CCA in consensus sequences, as it was the subject of prolonged discussion (Laibelman, 2022, 2024). The only repetitive emphasis needed in these pages is that ~25% of archaeal tRNA strings naturally encode 3'-CCA. Since a variety of lengths are observed (Tables S11, S12), it must be true that they cannot appear as invariant nucleotides in consensus sequences.

Direct comparisons between two types of aaRS were mentioned earlier (Tables 4, S10), as well as failed efforts for amidotransferase subunits. With respect to tRNA, one direct comparison (Table S10) holds major biological significance: tRNA^{eMet} and tRNA^{iMet} bind to the same MetRS; properly charged by the same methionine; utilize the same ATP to effect aminoacylation. Their distinction is

comprised solely by post-aminoacylation transfer to agents responsible for transportation to ribosomes in order to complete the translation process. Clustal Omega output (Table S12, reproduced here) reveals many similarities and differences across phyla (anticodon in red).

There is one extra nucleotide in consensus eMet compared to iMet sequences, with the anticodon triplet appearing between positions 36–41. Inserted nucleotides are pronounced in Halobacteriota eMet tRNA, creating a rightward shift at 26→27 relative to Methanobacteriota and Thermoproteota, plus multiple insertions at 34→37, 46→50, 50→55 leading to five total additional residues. Methanobacteriota and Thermoproteota keep in positional lockstep from start to finish. A single displacement exists for iMet tRNA occurring at 25→26 in Methanobacteriota relative to the other two phyla. Table 6 quantifies the presence of many more invariant nucleotides in tRNA^{iMet} easily visualized: they congregate mainly, not exclusively, around anticodon positions. Collapsing invariant nucleotides for all phyla and readjusting positional shifts to force uniformity yields the following series:

eMet tRNA H/M/T									
000000000111111111222222222233333333334444444444555555555566666666667777777777									
1234567890123456789012345678901234567890123456789012345678901234567890									
G	U	G	A	GG	GC	CUCAUA	A	C	GUC C CA
GCCG	G	GCU	AG	GG	AGCG	GCUCAUAA	C	AG	C GGGUUC A CCC C CGGCA
G	U	GC	AGC		G G G	G CCAU A C		C	G G UC A C C CA
iMet tRNA H/M/T									
000000000111111111222222222233333333334444444444555555555566666666667777777777									
123456789012345678901234567890123456789012345678901234567890123456789									
AGCG	G	U	GG	UAG	AGG	AU	CCG	CGGGCUCAUAACCCG	AGA C GUUC AAUC C CGCUA
AGC	GG	GGG	AG	GG		CCCCG	GGGCUCAUAACCC	AG C	GUUC AAUC C GCUA
AGCG	GU	GG	AG	C	GG		C G	GGGCUCAUAACCC	AG C GUUC A UC C CGCUA
eMet tRNA common H/M/T									
000000000111111111222222222233333333334444444444555555555566666666667777777777									
1234567890123456789012345678901234567890123456789012345678901234567890									
G		G	A		G	C	CAU		C G UC C CA
iMet tRNA common H/M/T									
000000000111111111222222222233333333334444444444555555555566666666667777777777									
123456789012345678901234567890123456789012345678901234567890123456789									
AGC		GG	AG	GG		C	G	GGGCUCAUAACCC	AG C GUUC A UC GCUA

tRNA eMet + tRNA iMet common H/M/T									
000000000111111111222222222233333333334444444444555555555566666666667777777777									
1234567890123456789012345678901234567890123456789012345678901234567890									
	G	A				C	CAU		C G UC C A

In tRNA^{eMet}, applying the condensed consensus formulation introduced for synthetases to identify specific invariants, and ignoring the anticodon, G^{c10}/A^{c14}/G^{c55}/U^{c57}/C^{c58}/C^{c63} are acknowledged as vital for internal molecular conformation in 3D, as long as a two-nucleotide shift post-anticodon is accepted, just as Biela et al. and Tamaki et al. expressed (Table 7). Transfer RNA of iMet type encodes all of them as well, reinforcing their significance for molecular folding, and supplements them with G^{c19}/G^{c20}/U^{c35}/U^{c56}/A^{c60} from Table 7. In other words, tRNA^{iMet} maintains invariance for all nucleotides claimed as crucial except U⁸, A²¹, 3'-CCA. As first in line for translation, this tRNA must get into

proper binding mode as precisely as possible for interaction at the ribosome, which requires strict adherence to the *quasi*-L shape needed. If it fails to do so, translation is problematic at the outset. Possessing invariance for these eleven nucleotides across all archaeal phyla amounts to a guarantee of conformity.

Returning to tRNA^{eMet}, still ignoring the anticodon, conserved bases G^{c1}/G^{c26}/C^{c34}/C^{c50}/C^{c74}/A^{c75} do not as yet have a functional role attributed to them. Of these, C^{c34}/C^{c50}/A^{c75} are also in the consensus tRNA^{iMet} string, meaning they supply the same attribute for the translation process, just as is true for those nucleotides engaged in molecular self-folding. There are two obvious options: binding to MetRS; binding to appropriately-positioned residues in their respective elongation/initiation factors. The mode of binding might be salt bridge formation to tRNA phosphate oxygens, but hydrogen-bonding to nucleobases cannot be dismissed as an option without further information to adjudicate between possibilities.

Still needing explanation for tRNA^{eMet} invariance are G^{c1}/G^{c26}/C^{c74}. None are conserved within these tRNA^{iMet}. It is posited that they interact with residues in the elongation factor to ensure transport of the correct tRNA at the correct time for protein incorporation. This thesis is almost certainly true for G^{c1} because it suffices to distinguish eMet from iMet types. Elongation factor binding to charged tRNA^{eMet} is less crucial because while there is only one first amino acid in a protein, methionine may appear multiple times throughout a chain. Aside from the anticodon triplet itself, few bases are needed to interact with an elongation factor to ensure its proper linkage

A similar stepwise analysis can be undertaken to explain invariance in tRNA^{iMet}, but probably not needed: its significance for translation is overt. Any initiation factor conveying a charged tRNA to the starting location at the ribosome must be certain it binds to tRNA^{iMet} and no other molecule. One way to obtain this guarantee is for every targeted agent to possess a large number of invariant nucleotides making misidentification as close to impossible as can be designed, although it is inevitable that errors will rarely eventuate. This 'large number' are seen to cluster around the anticodon such that a misreading resulting in linkage to another tRNA, even tRNA^{eMet}, cannot be easily accomplished. Clustered GGG, AA, CCC does the job. The leading adenosine A^{c1} alone ensures an absence of confusion with G^{c1} of tRNA^{eMet}. A handful of others detailed in the condensed consensus sequence are sprinkled throughout the string in order to further contribute to correct identification by an initiation factor and/or by rRNA.

Part III. Aminoacylation Mechanisms

Aminoacylation of tRNA as necessary precursor for translation of mRNA transcripts into protein is accomplished by interaction of amino acid, ATP, and most often Mg²⁺ in the immediate environment of an appropriate tRNA synthetase. Virtually every article expresses this process as the product of a two-step sequence: amino acid plus ATP form aminoacyl adenylate and pyrophosphate; aminoacyl adenylate plus hydroxyl of 3'-terminal ribose within tRNA form aminoacyl-tRNA and AMP. In actuality, this overview is simplistic and constitutes only one possibility out of a spectrum of options at the molecular level. Stripped of biochemical trappings, aminoacylation is nothing but an esterification reaction learned in introductory organic chemistry class: R₁CO₂H + R₂OH ⇌ R₁CO₂R₂ + H₂O. Generality is revealed by usage of R₁ and R₂: small methyl groups or large polynucleotide chains makes no ultimate difference. Complexity remains hidden. As a pH-dependent reaction, there are three esterifications from a mechanistic perspective:

- at low pH: R₁CO₂H + R₂OH ⇌ R₁CO₂R₂ + H₂O
- at intermediate pH: R₁CO₂⁻ + R₂OH ⇌ R₁CO₂R₂ + OH⁻
- at high pH: R₁CO₂⁻ + R₂O⁻ + H₂O ⇌ R₁CO₂R₂ + 2OH⁻

The third equation reveals a necessity for intervention of water; it is an invisible component in the other two reactions. Complexity still remains hidden. This summary description is grounded on a detailed mechanism comprised of elementary steps involving nucleophilic substitution and proton transfers. With respect to aminoacylation, a side reaction is viable: acylate transfer from an aminoacyl adenylate to ribose hydroxyl. The adenylate moiety is a mixed anhydride (O=C-O-P=O) serving as a

vehicle for delivery of amino acid (as carboxylate) to alcohol: $R_1CO_2(P)AMP + R_2OH \rightleftharpoons R_1CO_2R_2 + AMP(P)-OH$, where (P) is intended to convey covalent binding of phosphorous atom within AMP.

Since H^+ are indistinguishable, migrations between donor and acceptor can be treated identically: $D-H + A \rightleftharpoons D + H-A$, where donor D and acceptor A can be charged ions or uncharged atoms. Kinetic rates for exchange depend not only on the chemical identity of D and A, but on solvent viscosity, medium pH, reaction temperature; thus, a range of values can be anticipated.

Liepinsh & Otting (1996) used a NMR technique to explore proton exchange rates between water and the sidechains of Ser, Thr, Tyr, Arg, Lys over pH = 0.5–8.5 and T = 4–36°C, finding the slowest to be $\sim 700\text{ s}^{-1}$ and the speediest at $\sim 10^4\text{ s}^{-1}$. However, the NMR technique has an effective timescale range for measurement peaking at $\sim 10^{-4}\text{ s}$ though used for processes occurring as rapidly as $\sim 10^{-7}\text{ s}$ according to the authors. This means the larger rate cited is pushing the envelope boundary for reliability.

Pines et al. (1997) looked at allegedly diffusion-controlled proton transfer between isomeric α - or β - naphthols as donor with acetate as acceptor over a pH range of eight log units. By direct measurement, they obtained an average rate of $3 \times 10^{11}\text{ s}^{-1}$, and mention within their report that another study found a second order rate of $2.9 \times 10^9\text{ M}^{-1}\text{ s}^{-1}$ at 1M base strength, equivalent to a transfer reaction lifetime of ~ 350 picoseconds.

The classical Grotthuss mechanism based on movement of a proton between two amphoteric H_2O requires 1–2 ps (Grotthuss, 1806; Hassanali et al., 2013). If this model is of general applicability, then a picosecond scale lifetime implies a proton transfer rate upper bound limit at $\sim 10^{12}\text{ s}^{-1}$. The Pines et al. experiment is in accord with this extremum.

None of these rates must accommodate multi-component media contained within biological cells, particularly in regard to enzyme-catalyzed tRNA aminoacylation kinetic events. *Fick's Law* formulates a first order differential equation for an amount of material passing through an area per unit time under diffusion-controlled conditions called *diffusion flux* J. It is expressed as $J = D(dC/dx)$, with D a diffusion coefficient constant, C the concentration of material expressed in molar units, x the one-dimensional path length of the medium. For water movement, D is known with accuracy as equal to $9.3 \times 10^{-5}\text{ cm}^2/\text{s}$.

At 37°C, pure water has a density of 0.99333 g/cm^3 (<https://www.internetchemistry.com>). For a study on the effects of hydration levels on proton movement, the total solvent accessible surface area of a tRNA molecule was estimated as $\sim 13,400\text{ \AA}^2$, with surface proton donor sites occupying 67% on average, or $\sim 8978\text{ \AA}^2$; the rest are nonexchangeable hydrogens on nucleobases or ribose rings (Roh et al., 2009). According to Biro (2012), the one-dimensional length of tRNA is $\sim 7.6\text{ nm}$ from acceptor end to anticodon triplet end. Proceeding stepwise, the calculation is:

$$\begin{aligned} (0.99333\text{ g/cm}^3) (6.022 \times 10^{23}\text{ molec}/18.01\text{g}) &\approx 3.32 \times 10^{22}\text{ molec/cm}^3 \text{ (H}_2\text{O concentration C)} \\ 3.32 \times 10^{22}\text{ molec/cm}^3 / 7.6 \times 10^{-7}\text{ cm} &\approx 4.37 \times 10^{28}\text{ molec/cm}^4 \text{ (C/x = maximum dC/dx)} \\ (9.3 \times 10^{-5}\text{ cm}^2/\text{s}) (4.37 \times 10^{28}\text{ molec/cm}^4) &\approx 4.06 \times 10^{24}\text{ molec/cm}^2\text{-s} \text{ (J = D dC/dx, diffusion flux)} \\ (4.06 \times 10^{24}\text{ molec/cm}^2\text{-s}) (8.978 \times 10^{-13}\text{ cm}^2) &\approx 3.65 \times 10^{12}\text{ molec/s} \text{ (diffusion rate of H}_2\text{O over tRNA)} \end{aligned}$$

According to Grotthuss, two H_2O equals one transfer: $[H_3O^+]_1 + [H_2O]_2 \rightleftharpoons [H_2O]_1 + [H_3O^+]_2$. This means the calculated diffusion rate for water must be reduced by 50% to reflect the H^+ diffusion rate, so a maximum of $\sim 1.8 \times 10^{12}$ proton migrations occur per second between donor and acceptor located on the surface of a typical tRNA molecule at 37°C. This outcome agrees not only with Grotthuss, but with the Pines et al. experimental data as well.

Three factors have been ignored in this idealized calculation: (1) effect of temperature; (2) effect of pH; (3) effect of salt concentration present. These are precisely the same three parameters examined for optimal cell growth conditions (Tables 1, S2). As temperature rises, water density decreases due to greater thermal motion; as salt concentrations increase, water density increases because greater mass is present in an unchanged spatial volume. For simplicity, pretend these two effects magically cancel each other such that the idealized solution is a viable representation of authentic cellular states.

As for pH, it is irrelevant: although a tiny fraction of hydronium ions exists in bulk water, there is no interest in number of protons transferred, only the kinetic rate for movement. Depiction of a 'typical' tRNA solvent accessible surface area and the proportion of its surface nucleotides harboring migratory protons will vary for each specific case depending on composition and length; the true exchange rate will need adjustment from the idealized calculation given.

This extended discussion proves something intuitively obvious: H^+ transfer, independent of donor and acceptor identity, can never be the rate-determining elementary step for aminoacylation kinetics. This leaves as options for potential RDS:

- synthetase binding k_{on}/k_{off} for: (i) tRNA; (ii) amino acid; (iii) ATP;
- complex $k_{formation}/k_{decomposition}$: (i) aaRS/aa/ATP; (ii) aaRS/aa/ATP/tRNA; (iii) aaRS/aa-(P)AMP/PP_i; (iv) aaRS/aa-(P)AMP/tRNA/PP_i; (v) aaRS/aa-tRNA/AMP; (vi) aaRS/aa-tRNA;
- at low pH, S_N2 attack by: (i) aa-CO₂H on (P_α)ATP to form aa-CO₂(P)AMP; (ii) HO-(P_α)ATP on aa-CO₂H to form aa-CO₂(P)AMP; (iii) ribose 2'-OH (3'-OH) on aa-CO₂(P)AMP to form aa-tRNA;
- at intermediate pH, S_N2 attack by: (i) aa-CO₂⁻ on P_α-ATP to form aa-CO₂(P)AMP; (ii) ribose 2'-OH (3'-OH) on aa-CO₂(P)AMP to form aa-tRNA;
- at high pH, S_N2 attack by: (i) aa-CO₂⁻ on P_α-ATP to form aa-CO₂(P)AMP; (ii) ribose 2'-O⁻ (3'-O⁻) on aa-CO₂(P)AMP to form aa-tRNA;
- departure of aa-tRNA from aaRS (i) with or (ii) without assistance of initiation (elongation) factor.

This accounting supports the conceptualization of aminoacylation mechanisms as constituting a 'spectrum of options at the molecular level.' It is emphatically *not* the intention here to ascertain which possibility holds for any system, whether archaeal, bacterial, or eukaryal. To sort through the candidates for rate-determining step, an inquiry into conventional procedures used to investigate tRNA aminoacylation kinetics is necessary. Francklyn et al. (2008) offered a comprehensive review on the topic pertinent nearly two decades later, providing extensive details on exact protocols:

- assay mixture monitoring formation of E•AA~AMP at pH 7.78/25–37°C (thermophiles at ≥ 65°C);
- monitoring ATP consumption at pH 7.5/37°C;
- amino acid activation with aminoacyl-tRNA synthetase by PP_i exchange reaction at pH 8.0/37°C;
- complete aminoacylation reaction by steady state assay at pH 7.5/37°C.

These procedures, which the review authors maintain are standardized conditions, are demonstrably incompatible with most thermophilic, acidophilic, halophilic, anaerobic Archaea. Aminoacylation kinetics trials usually use *E. coli* or another readily available bacterium as enzyme source or host, hence conducted under conditions (Berg et al., 1961; Blanquet et al., 1975) inhospitable to extremophiles. All four growth factors employed to culture Archaea effectively, considered reasonable proxies for natural environmental living conditions, are violated by these methods (Tables 1, S2). Aerobic conditions are more convenient in laboratory settings, but contrary to the needs of 60% of covered archaeal organisms. Neutral pH ignores both acidophilic and alkaliphilic species. Added salts at 10–200 mM amounts are inadequate even for most non-halophiles, since the < 1.0 M category frequently means 300 mM or more, as shown in Table S2. Although temperature is sometimes attended to as a factor, a choice between 37°C and 65°C leaves much territory excluded.

Schwartz & Pan (2016) found during an aminoacylation kinetics study on *Aeropyrum pernix* that different amounts of mistranslation by tRNA^{Leu} isoacceptors were produced when conducted at 75°C or 90°C (latter is the optimal growth temperature, Table S2): lower temperature produced more substitution by methionine for leucine as charged amino acid. They speculated that the introduction of methionine was beneficial when the surrounding environment was colder because it allowed for greater sidechain mobility (greater entropy) when incorporated into proteins. Understanding this

type of interplay between nature and nurture is crucial for comprehensive formulation of the mechanics of translation. If experiments do not apply to Archaea, then inferences drawn on mechanistic details might legitimately be questioned for this entire domain of lifeforms.

The other premier source for information on mechanism comes from structural features obtained from crystallographic data, whose three principal contributors are bacterial: *E. coli*, *T. thermophilus*, *B. stearothermophilus*. The primary problem is not Bacteria *versus* Archaea, but the intrinsic pragmatic issue pertaining to constraints imposed by methodology. Crystal structures are literally ‘frozen in time’ images; they are the antithesis of dynamic alterations happening in real time. Molecules, biologically significant or otherwise, are not static, but are continuously undergoing vibrational, rotational, translational motions. This dynamism cannot be captured by X-ray bombardment of solid-state materials. Such criticism does not imply the data obtained is unimportant or irrelevant. It does, however, mean the information represents a boundary limit on possibilities to be considered plausible. Invocation of interactions between enzyme and substrates (amino acid, tRNA, ATP), or between H₂O and these biomolecules, are not entirely reliable because bond distances and angles vary not just instantaneously but on an ongoing basis without cessation.

Despite intrinsic flaws in the two main sources of information on aminoacylation of tRNA, the most controversial and critical aspect for assigning a rate-determining elementary step is whether aminoacyl adenylate aa-CO₂(P)AMP is a transition state or an intermediate. The distinction is crucial, though often glossed over in the literature. A true intermediate is in a thermodynamic potential energy well such that the molecule has chemical metastability of sufficient lifetime to be isolatable, if only briefly. A true transition state lacks thermodynamic stability and decays rapidly to generate expected reaction product or unwanted byproduct. The decisive factor causing any chemical entity to be assigned to transition state or intermediate status is a thermodynamic property of the material, not its dependence on kinetic rates of formation.

It is possible the crux of the transition state *versus* intermediate debate has its origins in ongoing use of imprecise language. The beginning of this Section specified ‘virtually every article expresses this process as the product of a two-step sequence’. More emphatically, the written words are often verbatim from research paper to research paper: a cut-and-paste operation. Due to the citation reference system in place, this is not plagiarism, but accepted practice. Nonetheless, what author A intends to convey may not be the same as that of author B implementing the same verbiage. Does aminoacyl adenylate as product of step 1 really mean isolable molecule? Is it unstated that it actually refers to activated amino acid bound to synthetase and that the complex is the long-lived entity? The two interpretations are not interchangeable. Is it plausible that the word ‘step’ does not suggest a distinct stopping point between one chemical state and another chemical state in the same manner that steps exist on a staircase? To be, literally, meaningful, scientific language must be as precise as possible.

If amino acid and ATP are placed in water, aminoacyl adenylate and pyrophosphate would never spontaneously ($\Delta G_{\text{rxn}} < 0$) appear. “Absent biological catalysts, amino acid activation by ATP occurs at a rate $\approx 8.3 \times 10^{-9} \text{ M}^{-1} \text{ s}^{-1}$ at pH 9.7/39°C. This represents, by many orders of magnitude, the highest activation energy barrier of any reaction required for protein synthesis” (Pham et al., 2010). From the Arrhenius activation energy formula $k = Ae^{-E_a/RT}$, with $10^9 \text{ M}^{-1} \text{ s}^{-1}$ taken as average second order reaction collision frequency pre-exponential value A (typically 10^8 – $10^{10} \text{ M}^{-1} \text{ s}^{-1}$), this rate translates to $E_a \approx 26.3 \text{ Kcal/mol}$. This information has no direct bearing on the issue at hand, though its indirect significance means formation of any true intermediate must have $\Delta G_{\text{formation}}$ less than this value.

It is tempting to presuppose synthetase automatically compensates for an energetically disfavored thermodynamic transformation, but that supposition is incorrect: in the presence of aaRS, a substantial barrier is still predicted for reactivity. A quantum mechanics/molecular mechanics computer simulation for adenylation of aspartic acid with ATP situated at the catalytically active site in a fully solvated AspRS from *E. coli* held at room temperature ($\sim 25^\circ\text{C}$) gave $\Delta G_{\text{rxn}} = 23.3$ – 23.5 Kcal/mol . This in silico result is in reasonable agreement with the laboratory-based outcome

articulated by Pham et al. despite a temperature differential and assumed pre-exponential A value. Using Gibbs free energy equation $\Delta G = -RT \ln K_{eq}$, the computed calculation yields $K_{eq} \approx 5.8 \times 10^{-18}$ for $Asp + ATP \rightleftharpoons Asp-AMP + PP_i$ (Dutta & Chandra, 2022). Though reaction is not thermodynamically favored even in the presence of aaRS (i.e., not spontaneous), to the extent these two results can be placed side-by-side, since $\Delta G_{rxn} \approx 23.4$ Kcal/mol is smaller than $E_a \approx 26.3$ Kcal/mol, then it is chemically feasible for Asp-AMP to be a true intermediate, but not a certainty.

Better, yet still not definitive, proof for the reality of aa-AMP as an intermediate was provided by Zhang et al. (2006), who isolated the CysRS/Cys-AMP complex from *E. coli* and measured $t_{1/2} = 23$ min at room temperature. Attachment to the enzyme probably enhanced the adenylate's innate stability, and it is unknown whether Cys-AMP by itself is isolable with a measurable half-life. The difficulty in offering convincing evidence of status as intermediate is exemplified by the results from Dong et al. (2010). These researchers went at the issue from the opposite direction Zhang et al. did. They synthesized aminoacyl adenylates in a laboratory in order to generate crystal structures of aaRS/aa-AMP complexes. They then showed these complexes would react with tRNA to effect aminoacyl transfer. What this result does *not* mean is sequence (1) $aaRS + aa-AMP \rightleftharpoons aaRS/aa-AMP$; (2) $aaRS/aa-AMP + tRNA \rightarrow aa-tRNA + AMP + aaRS$ must occur within living cells in real time. Pre-formed complexes are unable to differentiate in vivo interaction as an energetically disfavored transition state from existence in a local energy minimum well characteristic of a molecule with a brief, but measurable, lifetime as a true intermediate.

Since aminoacylation is a pH-dependent process, aminoacyl adenylate intermediates would have to demonstrate sufficient innate stability to resist pH-induced pressure to react on a timescale faster than one half-life. At least a portion of such stabilization would need to derive from salt bridge formation while bound to synthetase. Absent assistance, can aminoacyl adenylates survive for the time necessary to isolate and quantify it as an intermediate? Berg et al. (1961) stressed innate instability of aminoacyl adenylate physiologically due to rapid hydrolytic destruction regenerating amino acid along with AMP.

To have a chance at success, thermodynamic stability must correlate with formation or collapse as the kinetic rate-determining step in the overall aminoacylation reaction. Traditionally, kinetic rates cannot be inferred from thermodynamic values, or vice versa, except for activation energy through the Arrhenius formula. Yet, a relationship has been developed by the National Aeronautics and Space Administration (NASA) in the US for *gas phase* reactions: $r = D [e^{(-1/RT)(\partial G/\partial X)} - 1]$ where r is kinetic rate constant, $D = 3 \times 10^8 \text{ sec}^{-1}$, X is extent of conversion to products, $\partial G/\partial X$ is Gibbs free energy gradient (Marek, 1995). The formula is untried on complex systems such as exist in biological cells.

Under $pH \leq 5$ (Bas et al., 2008), conditions preferred by acidophilic Archaea (Tables 1, S2), the charging amino acid exists as CO_2H rather than CO_2^- . Arg, His, Lys sidechains are protonated also, which would prevent them acting as bases to abstract the acidic hydrogen and enhance the charging amino acid's nucleophilicity. Similarly, low pH makes protonation of phosphate oxygens more likely, so they also would have diminished capacity to increase S_N2 attack rates. Ribose hydroxyls would be in their neutral state, but possibly hydrogen-bonded and weakly nucleophilic. Loftfield & Eigner (1969) discovered that added base (imidazole, hydroxylamine) stimulated tRNA-charging reactions. The single component present able to improve charging amino acid nucleophilicity is H_2O acting as a general base even at low pH by way of ultra-fast proton transfers. In combination, the low pH state makes it more likely any of the three S_N2 attack options listed could be viable as a rate-determining step candidate, implying it becomes more likely for $aa-CO_2(P)AMP$ to exist as a real intermediate. The question: would it be an intermediate by itself or an intermediate only in the form $aaRS/aa-CO_2(P)AMP$?

As preferred reaction pH moves towards neutrality, the incoming amino acid is in its CO_2^- form. Except for histidine, cationic aaRS residues would still be protonated and able to serve as salt bridges to reduce nucleophilic attack on P_α in ATP. Metal cations (Mg^{2+} , K^+ , Na^+) can chelate both with $aa-CO_2^-$ and phosphate oxyanions, also serving to reduce nucleophilicity. Stabilization through hydrogen bonding would act as neither boon nor deterrent due to ongoing formation and cleavage

concomitant with molecular vibrations, rotations, translations. Ribose hydroxyls would still be in their -OH form, so relatively poor nucleophiles dependent on lone pair electrons for reaction activation. The only component present able to improve charging amino acid nucleophilicity is H₂O acting as a general base. A lack of rate enhancement around pH 7 leads to the same situation as the low pH condition: either listed option would be a viable RDS candidate, which implies it becomes more likely for aa-CO₂(P)AMP to exist as a real intermediate. The question: would it be an intermediate by itself or an intermediate only in the form aaRS/aa-CO₂(P)AMP?

Under pH > 9 conditions preferred by some alkaliphiles (Tables 1, S2), the situation changes. The amino acid is now aa-CO₂⁻, as is PO⁻ from ATP (and tRNA if present). Arg and Lys synthetase residues have smaller fractions in their protonated state, and those that are have more choices for salt bridge partners. Metal cations still chelate available anions within range, but greater amounts of hydroxide present in bulk water means cations have less need to seek out aa-CO₂⁻. Furthermore, OH⁻ can now act as base to abstract a proton from ribose 2'-OH (3'-OH), making them more capable of S_N2 approach towards acyl adenylate. What these chemical states indicate is quicker aa-CO₂(P)AMP formation along with faster degradation to produce aa-tRNA. In other words, higher pH is more likely to include acyl adenylate as a transition state than as an intermediate whether enzyme-bound or not, but especially if still joined to enzyme because the tRNA nucleophile is also bound nearby.

As detailed, the only components present able to improve charging amino acid nucleophilicity are H₂O and PO⁻ acting as general bases. Dutta & Chandra (2023) calculated phosphate possesses ΔG = 52.6 Kcal/mol barrier if acting as general base, but this value is reduced to 39.7 Kcal/mol when H₂O-mediated H⁺ transfer is allowed. The thermodynamic picture improves when exothermic conversion of ATP ⇌ AMP + PP_i is factored in because ΔG ≈ -11 Kcal/mol under standard state parameters (1M, 25°C), and more negative (difficult to measure accurately) under physiologic conditions where millimolar concentrations or greater of cations are present. Nonetheless, it is clear that activation is endothermic, and consistent with the aspartic acid system discussed where ΔG = 23.3–23.5 Kcal/mol; at worst, 39.7 Kcal/mol + less than -11 Kcal/mol ≤ 28.7 Kcal/mol.

If reasoning on the possibilities for chemical involvement in promoting acyl adenylate formation and breakdown are correct, kinetic rates at the activation stage should not only reflect this thermodynamic picture, but, within limits, be approximately the same for all aaRS types. An *E. coli* TyrRS system using 40 μM Tyr, 12 μM tRNA^{Tyr}, 0.5 mM ATP at 25°C and pH 7.78 gave aminoacylation rate ≈ 40 s⁻¹ (Fersht & Jakes, 1975). An *E. coli* CysRS system using 0.5 mM Cys, 5 μM tRNA^{Cys}, 6.25 mM ATP at unstated T/pH conditions gave aminoacylation rate = 15.2 s⁻¹ (Zhang et al., 2006). An *E. coli* ThrRS system (abstract with no details given) gave aminoacylation rate ≈ 29 s⁻¹ (Bovee et al., 2003). All three determinations were made under single turnover, not steady-state, kinetics conditions.

Given the unfavorable thermodynamics for overall amino acid activation, production of a mixed anhydride *could* constitute the rate-determining portion of the overall process and become a legitimate intermediate. An alternative to what is universally invoked in contemporary literature is that aminoacyl adenylate formation represents a definite transition state in a pseudo-concerted process. Covalent bond breaking and bond making need not be exactly equal at every moment to qualify as simultaneous. This conception has a long history (Berg et al., 1961; Loftfield & Eigner, 1969; Fersht & Jakes, 1975), but is sometimes ignored or not considered by researchers. Perona et al. (1993) declared aminoacylation of tRNA^{Gln} occurs by a concerted reaction in *E. coli*, but this has no direct bearing on archaeal reactivity since none encode GlnRS. Charging tRNA^{Gln} via an indirect pathway involves amidotransferase subunits, which complicates assignment of rate-determining step because additional elementary steps must be kinetically evaluated.

In a more-or-less concerted reaction, to prevent hydrolysis of aminoacyl adenylate by bulk water, a tRNA's terminal ribose 2'-OH or 3'-OH, an unimportant distinction mechanistically, must rapidly attack the activated carboxyl carbon. The nucleophile exists in its neutral form at acidic pH but could develop a transitory oxyanion at neutral or basic pH via general base catalysis. Nucleophilic attack forms the usual tetrahedral transition state (not intermediate) at the former carbonyl carbon of the amino acid to generate a bond to tRNA. A series of diffusion-controlled proton transfers causes

quick collapse, releasing aminoacyl-tRNA plus AMP plus pyrophosphate. At the reaction's end, H₂O is (re)generated if cellular pH ≤ 8, or OH⁻ is produced at higher levels of basicity.

A basic amino acid in its neutral form within synthetase could abstract H⁺ from ribose-OH with $\Delta G = 10.7$ Kcal/mol (Dutta & Chandra, 2023). A large thermodynamic change in barrier height, compared to others they calculated, is enthalpy driven because aa-NH₂ > H₂O >> PO⁻ in basicity towards ROH, as shown by Loftfield & Eigner (1969). Entropy changes in the wrong direction since tRNA and candidate amino acid acting as base are already in proximity in/near the active site. Avoiding involvement of free-floating H₂O produces a more organized state prior to reaction. They also calculated that 3'-O⁻ nucleophilic attack forming the expected tetrahedral transition state comes with $\Delta G = 24.8$ Kcal/mol. Near-identical free energy changes for adenylation (23.4 Kcal/mol) and definite transition state formation upon nucleophilic attack at carboxyl carbon (24.8 Kcal/mol) suggests aa-CO₂(P)AMP represents a transition state rather than an intermediate.

Another clue to existence of a *quasi*-concerted aminoacylation reaction is a kinetic need for tRNA presence in the aaRS/AA/ATP complex in order for amino acid activation to proceed. ArgRS, GluRS, Class I LysRS are all known to fit this requirement in some (probably bacterial) organisms (Cvetesic & Gruic-Sovulj, 2017). Evidence also stems from recognizing the existence of two synthetase Classes. Although greatly limited in scope of covered species, consensus has gradually been achieved pertaining to the rate-determining step: it is said to be different for each, with removal from enzyme for Class I, but amino acid activation for Class II types (Zhang et al., 2006).

They based their declaration on work with CysRS from *E. coli*, discovering the rate of dissociation of Cys-tRNA^{Cys} from synthetase relative to rate of aa-CO₂(P)AMP formation was slower, suggesting final release from this Class I synthetase was rate-limiting. They hypothesized EF-Tu may have participated in product removal. Pyrkosz et al. (2010) were more quantitative: in *T. thermophilus*, binding affinity of tRNA^{Glu} for EF-Tu was 300× stronger than for GluRS. If division by Class is upheld through exploration of more systems—especially archaeal over an assortment of temperature, pH, salinity, aerobic, anaerobic variables—then it must be said that RDS release of aminoacylated tRNA from synthetase is consistent with a *quasi*-concerted mechanism.

Theoretically, concerted reactions producing charged tRNA do not demand aminoacyl adenylate formation and collapse. Aminoacylation rate would be slower if ATP were absent; the E_a barrier would be greater because there is an absence of pyrophosphate production to serve as driving force. Still, the overall thermodynamic picture is unchanged: $\Delta G_{\text{rxn}} = \Delta G_{\text{aminoacyl-tRNA}} - (\Delta G_{\text{tRNA}} + \Delta G_{\text{aa}})$ where ΔG refers to free energy exhibited by each molecule when bound to synthetase. Introduction of nonstandard amino acids into proteins has been accomplished without involving ATP (Wang et al., 2009).

For every combination of aaRS, tRNA, amino acid, the transition state *versus* intermediate debate lends itself to experimental determination by isotopic labeling. Natural amino acids can be prepared by Strecker synthesis (RCHO + NH₃ + CN⁻), with H₂O¹⁸ used for hydrolysis of intermediate aminonitrile RCH(NH₂)CN to RCH(NH₂)C¹⁸O₂H, where R is any amino acid sidechain (Strecker, 1854; Masamba, 2021). The original Strecker synthesis procedure yields enantiomers in 50:50 ratio (racemic mixture), but chromatographic chiral resolution gives optically pure material if performed under conditions where the isotopic label is not lost. Alternatively, a variety of asymmetric syntheses can be undertaken requiring no subsequent resolution steps (Cai & Xie, 2014).

If aa-C¹⁸O₂(P)AMP is an actual intermediate, 100% of the isotopic label will be retained during its formation. Aliquots taken over time from the reaction, followed by rapid isolation and measurement of radioactivity provides conclusive evidence, even if t_{1/2} is on the order of minutes. An identical situation holds if complexation with aaRS is obligatory for isolation. Whether transition state or intermediate, 50% of the label must be ultimately lost through production of H₂O¹⁸ or ¹⁸OH⁻, again demonstrable through aliquot removal, isolation of charged tRNA, radioactivity measurement.

At each time point t₁, t₂ ... t_n, amount of change in label location d(¹⁸O) must be equal for the aliquots due to the mass conservation law: ¹⁸O_{init} = ¹⁸O_{fin}. Label is present in either aa-C¹⁸O₂(P)AMP or aa-C¹⁸O₂-tRNA or H₂O¹⁸ or ¹⁸OH⁻. and the sum of label amounts found in water or hydroxide ion

must equal amount found in charged tRNA. Comparison of time-dependent measures of radioactivity enables classification as intermediate or transition state: rate of label location change $d(^{18}\text{O})/dt$ differs for the two options. This kinetics experiment must be conducted in a stopped-flow instrument in order to achieve the short timescales needed for analysis.

The rate declines slowly if aminoacyl adenylate is an intermediate due to concentration buildup of $\text{aa-C}^{18}\text{O}_2(\text{P})\text{AMP}$ signifying ^{18}O retention between successive times. This trend is mirror-image matched for $\text{aa-C}^{18}\text{O}_2\text{-tRNA}$ by slow increases in label accumulation over time. Plotting ^{18}O level vs. time for the two aliquots on the same graph will show $d(^{18}\text{O})/dt$ as linear for both, with slopes < -1 (adenylate) and $< +1$ (tRNA); the plot appears as a horizontally-elongated X. For aminoacyl adenylate as transition state, $\text{aa-C}^{18}\text{O}_2(\text{P})\text{AMP}$ is converted to $\text{aa-C}^{18}\text{O}_2\text{-tRNA}$ as rapidly as formed. Loss of label from adenylate is continuous and rate of loss constant over time. When graphed, $d(^{18}\text{O})/dt$ is linear for aminoacyl adenylate with slope = -1 , while for aminoacyl-tRNA slope = $+1$, presenting a perfect X image.

Whether intermediate or transition state, the intersection point in the graph corresponds to a time when 25% of ^{18}O label has shifted from aminoacyl adenylate to charged tRNA; that is, when the reaction is half-complete. For an intermediate, $t_{25\%}$ comes later in time than it would for a transition state. At the end time, 50% of initial ^{18}O label amount moves from $\text{aa-C}^{18}\text{O}_2(\text{P})\text{AMP}$ to $\text{aa-C}^{18}\text{O}_2\text{-tRNA}$, or equivalently from $\text{aa-C}^{18}\text{O}_2(\text{P})\text{AMP}$ to H_2O^{18} and $^{18}\text{OH}^-$ taken in combination.

Turning to a different but related topic, geometric details pertaining to average values for all RNA nucleotides are well-established: bond lengths $\text{N}^1/\text{N}^9\text{-C}_1'$ (1.471Å), $\text{C}_1'\text{-C}_2'$ (1.528Å), $\text{C}_2'\text{-O}_2'$ (1.413Å); bond angles $\text{N}^1/\text{N}^9\text{-C}_1'\text{-C}_2'$ (113.4°), $\text{C}_1'\text{-C}_2'\text{-O}_2'$ (110.6°) according to Gelbin et al. (1996). Since purine/pyrimidine nucleobase is covalently bound in the exo configuration at ribose C_1' , whereas $\text{C}_2'/\text{C}_3'\text{-OH}$ are on the endo side, the former is distant from the site of catalytic action for tRNA aminoacylation transpiring at the latter and plays no role, though it could be involved in π -stacking interactions or hydrogen bonding with synthetase amino acids (Kaiser et al., 2018). An inherent geometric barrier to involvement implies identity of the nucleobase is unimportant, not requiring terminal adenine despite the dogmatic position taken in the literature.

Not only does genomic primary sequence data support this conclusion, but kinetic evidence also. In a study on the effect of mutating 3'-CCA for aminoacylation, Zhou and coworkers (2011) obtained k_{obs} values for AMP formation using $\text{LeuRS/tRNA}^{\text{Leu(GAG)}}$ from *E. coli* here summarized:

- $3'\text{-CCA} \rightarrow 5.59 \text{ s}^{-1}$; $3'\text{-CCU} \rightarrow 0.24 \text{ s}^{-1}$; $3'\text{-CCG} \rightarrow 0.20 \text{ s}^{-1}$; $3'\text{-CCC} \rightarrow 0.48 \text{ s}^{-1}$; $3'\text{-CC} \rightarrow 0.71 \text{ s}^{-1}$
- $3'\text{-CAA} \rightarrow 8.34 \text{ s}^{-1}$; $3'\text{-CUA} \rightarrow 6.26 \text{ s}^{-1}$; $3'\text{-CGA} \rightarrow 0.64 \text{ s}^{-1}$
- $3'\text{-ACA} \rightarrow 7.48 \text{ s}^{-1}$; $3'\text{-UCA} \rightarrow 8.87 \text{ s}^{-1}$; $3'\text{-GCA} \rightarrow 2.65 \text{ s}^{-1}$

Their comment is uncompromising: "several CCA-mutated tRNA are efficiently aminoacylated, and LeuRS even aminoacylates tRNA^{Leu} lacking a terminal adenosine, showing remarkable plasticity." It will be recalled that when it was revealed (Table 7) supposedly necessarily conserved tRNA nucleotides were, in fact, not conserved in Archaea, a comment was made, now repeated: at worst, the L-shape varies within acceptable limits. This, in turn, suggests there is a degree of flexibility in the spatial co-existence of rRNA, mRNA, tRNA, ribosomal proteins. The work by Zhou et al. points in the same direction, only inserting 'plasticity' for 'flexibility.'

As a central process in biology, the understanding of translation needs a thorough review and, in selected areas, an overhaul, especially when combined with broad-based genomic information contrary to current doctrine, and a severe paucity of kinetic data accumulated under conditions of temperature, pH, salinity, presence or absence of O_2 favorable to the majority of Archaea. It is probable every organism (Archaea, Bacteria, Eukarya) will be kinetically and thermodynamically unique in one or more respects based on its evolutionary history in the environment in which it is most suited to live.

Supplemental Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Supplementary Table 1. Archaea Studied.

Supplementary Table 2. Cell Culture Growth Conditions. Supplementary Table 3. Synthetase and Amidotransferase Sequences. Supplementary Table 4. Synthetase and Amidotransferase Sequence Summary. Supplementary Table 5. HisRS Alignment for 2 *Caldisphaera lagunensis* Sequences. Supplementary Table 6. Synthetase and Amidotransferase Multiple Genes. Supplementary Table 7. Synthetase Class I and II Motifs. Supplementary Table 8. Synthetase and Amidotransferase Sequence Comparisons—Condensed Format. Supplementary Table 9. Number of Invariant Amino Acids in Three Phyla. Supplementary Table 10. Direct Sequence Comparisons. Supplementary Table 11. tRNA Sequences by Phylum. Supplementary Table 12. Consensus Invariant Nucleotides in tRNA Sequences.

References

- Ambrogelly A, Korencic D, Ibba M. 2002. Functional Annotation of Class I Lysyl-tRNA Synthetase Phylogeny Indicates a Limited Role for Gene Transfer. *J Bacteriol*, 184: 4594–600.
- Andam CP, Harlow TJ, Papke RT, Gogarten JP. 2012. Ancient Origin of the Divergent Forms of Leucyl-tRNA Synthetases in the Halobacteriales. *BMC Evol Biol*, 12: 85.
- Bas DC, Rogers DM, Jensen JH. 2008. Very Fast Prediction and Rationalization of pK_a Values for Protein-Ligand Complexes. *Proteins*, 73: 765–83.
- Bastian FB, Chibucos MC, Gaudet P, Giglio M, Holliday GL, Huang H, Lewis SE, Niknejad A, Orchard S, Poux S, Skunca N, Robinson-Rechavi M. 2015. The Confidence Information Ontology: A Step Towards a Standard for Asserting Confidence in Annotations. *Database*, article ID: bav043.
- Beebe K, Merriman E, Ribas de Pouplana L, Schimmel P. 2004. A Domain for Editing by an Archaeobacterial tRNA Synthetase. *Proc Natl Acad Sci USA*, 101: 5958–63.
- Berg P, Bergmann FH, Ofengand EJ, Dieckmann M. 1961. The Enzymic Synthesis of Amino Acyl Derivatives of Ribonucleic Acid I. The Mechanism of Leucyl-, Valyl-, Isoleucyl-, Methionyl Ribonucleic Acid Formation. *J Biol Chem*, 236: 1726–34.
- Biela A, Hammermeister A, Kaczmarczyk I, Walczak M, Koziej L, Lin T-Y, Glatt S. 2023. The Diverse Structural Modes of tRNA Binding and Recognition. *J Biol Chem*, 299: 104966.
- Biro JC. 2012. The Concept of RNA-assisted Protein Folding: The Role of tRNA. *Theor Biol Med Model*, 9: 10.
- Blanquet S, Fayat G, Poiret M, Waller J-P. 1975. The Mechanism of Action of Methionyl-tRNA Synthetase from *Escherichia coli*: Inhibition by Adenosine and 8-Amino-adenosine of the Amino-Acid Activation Reaction. *Eur J Biochem*, 51: 567–71.
- Boucher Y, Douady CJ, Sharma AK, Kamekura M, Doolittle WF. 2004. Intragenomic Heterogeneity and Intergenomic Recombination Among Haloarchaeal rRNA Genes. *J Bacteriol*, 186: 3980–90.
- Bovee ML, Pierce MA, Francklyn CS. 2003. Induced Fit and Kinetic Mechanism of Adenylation Catalyzed by *Escherichia coli* Threonyl-tRNA Synthetase. *Biochem*, 42: 15102–13.
- Brochier C, Forterre P, Gribaldo S. 2005. An Emerging Phylogenetic Core of Archaea: Phylogenies of Transcription and Translation Machineries Converge Following Addition of New Genome Sequences. *BMC Evol Biol*, 5: 36.
- Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. 2008. Mesophilic Crenarchaeota: Proposal for a Third Archaeal Phylum, the Thaumarchaeota. *Nature Rev Microbiol*, 6: 245–52.
- Cai X-H & Xie B. 2014. Recent Advances on Asymmetric Strecker Reactions. *Arkivoc*, 2014, 1: 205–48.
- Carter Jr. CW. 1993. Cognition, Mechanism, and Evolutionary Relationships in Aminoacyl-tRNA Synthetases. *Annu Rev Biochem*, 62: 715–48.

- Chalioitis A, Vlastaridis P, Mossialos D, Ibba M, Becker HD, Stathopoulos C, Amoutzias GD. 2017. The Complex Evolutionary History of Aminoacyl-tRNA Synthetases. *Nucl Acids Res*, 45: 1059–68.
- Curnow AW, Hong K, Yuan R, Kim S, Martins O, Winkler W, Henkin TM, Söll D. 1997. Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc Natl Acad Sci USA*, 94: 11819–26.
- Cvetesic N & Gruic-Sovulj I. 2017. Synthetic and Editing Reactions of Aminoacyl-tRNA Synthetases Using Cognate and Non-cognate Amino Acid Substrates. *Methods*, 113: 13–26.
- de Grotthuss CJT. 1806. Sur la Décomposition de L'eau et des Corps qu'elle tient en Dissolution à l'aide de L'électricité Galvanique. *Ann Chim*, 58: 54–73.
- Dong X, Zhou M, Zhong C, Yang B, Shen N, Ding J. 2010. Crystal Structure of *Pyrococcus horikoshii* Tryptophanyl-tRNA Synthetase and Structure-based Phylogenetic Analysis Suggest an Archaeal Origin of Tryptophanyl-tRNA Synthetase. *Nucl Acids Res*, 38: 1401–12.
- Dutta S & Chandra A. 2022. Free Energy Landscape of the Adenylation Reaction of the Aminoacylation Process at the Active Site of Aspartyl tRNA Synthetase. *J Phys Chem B*, 126: 5821–31.
- Dutta S & Chandra A. 2023. A Multiple Proton Transfer Mechanism for the Charging Step of the Aminoacylation Reaction at the Active Site of Aspartyl tRNA Synthetase. *J Chem Inform Mod*, 63: 1819–32.
- Eriani G, Delarue M, Poch O, Gangloff J, Moras D. 1990. Partition of tRNA Synthetases into Two Classes Based on Mutually Exclusive Sets of Sequence Motifs. *Nature*, 347: 203–6.
- Fang Z-P, Wang M, Ruan Z-R, Tan M, Liu R-J, Zhou M, Zhou X-L, Wang E-D. 2014. Coexistence of Bacterial Leucyl-tRNA Synthetases with Archaeal tRNA Binding Domains that Distinguish tRNA^{Leu} in the Archaeal Mode. *Nucl Acids Res*, 42: 5109–24.
- Fersht AR & Jakes R. 1975. Demonstration of Two Reaction Pathways for the Aminoacylation of tRNA. Application of the Pulsed Quenched Flow Technique. *Biochem*, 14: 3350–6.
- First EA & Fersht AR. 1995. Analysis of the role of the KMSKS loop in the catalytic mechanism of the tyrosyl-tRNA synthetase using multimutant cycles. *Biochem*, 34: 5030–43.
- Francklyn CS, First EA, Perona JJ, Hou Y-M. 2008. Methods for Kinetic and Thermodynamic Analysis of Aminoacyl-tRNA Synthetases. *Methods*. 44: 100–18.
- Fujishima K & Kanai A. 2014. tRNA Gene Diversity in the Three Domains of Life. *Front Genet*, 5: 142.
- Fukunaga R & Yokoyama S. 2006. Structural basis for substrate recognition by the editing domain of isoleucyl-tRNA synthetase. *J Mol Biol*, 359: 901–12.
- Gelbin A, Schneider B, Clowney L, Hsieh S-H, Olson WK, Berman HM. 1996. Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constituents. *J Amer Chem Soc*, 118: 519–29.
- Gomez MAR & Ibba M. 2020. Aminoacyl-tRNA synthetases. *RNA*, 26: 910–36.
- Gong P, Lei P, Wang S, Zeng A, Lou H. 2020. Post-Translational Modifications Aid Archaeal Survival. *Biomol*, 10: 584.
- Goudey B, Geard N, Verspoor K, Zobel J. 2022. Propagation, Detection and Correction of Errors Using the Sequence Database Network. *Brief in Bioinform*, 23: 1–12.
- Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. 2010. A New Bioinformatics Analysis Tools Framework at EMBL-EBI. *Nucl Acids Res*, 38 (Suppl): W695–9.
- Guerra S, Castiello U, Bonato B, Dadda M. Handedness in Animals and Plants. *Biol (Basel)*, 13: 502.
- Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, Madupu R, Robinson J, et al. 2010. The Complete Genome Sequence of *Haloferax volcanii* DS2, a Model Archaeon. *PLoS ONE*, 5: e9605.
- Hassanali A, Giberti F, Cuny J, Parrinello M. 2013. Proton Transfer Through Water Gossamer. *Proc Nat Acad Sci USA*, 110: 13723–8.

- Ibba M, Morgan S, Curnow AW, Pridmore DR, Söll D, et al. 1997a. A Euryarchaeal Lysyl-tRNA Synthetase: Resemblance to Class I Synthetases. *Science*, 278: 1119–22.
- Ibba M, Bono JL, Rosa PA, Söll D. 1997b. Archaeal-type Lysyl-tRNA Synthetase in the Lyme Disease Spirochete *Borrelia burgdorferi*. *Proc Natl Acad Sci USA*, 94: 14383–8.
- Jebbar M, Franzetti B, Girard E, Oger P. 2015. Microbial Diversity and Adaptation to High Hydrostatic Pressure in Deep-Sea Hydrothermal Vents. *Extremophiles*, 19: 721–40.
- Kaiser F, Bittrich S, Salentin S, Leberecht C, Haupt VJ, Krautwurst S, et al. 2018. Backbone Brackets and Arginine Tweezers delineate Class I and Class II Aminoacyl-tRNA Synthetases. *PLoS Comput Biol*, 14: e1006101.
- Karp PD, Paley S, Zhu J. 2001. Database Verification Studies of SWISS-PROT and GenBank. *Bioinform*, 17: 526–32.
- Karsch-Mizrachi I, Takagi T, Cochrane, G. 2018. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res*, 46: D48–D51.
- Kim H-S, Vothknecht UC, Hedderich R, Celic I, Söll D. 1998. Sequence Divergence of Seryl-tRNA Synthetases in Archaea. *J Bacteriol*, 180: 6446–9.
- Klenk H-P, Clayton RA, Tomb J-F, White O, Nelson KE, Ketchum KA, Dodson RJ, et al. 1997. The Complete Genome Sequence of the Hyperthermophilic, Sulphate-reducing Archaeon *Archaeoglobus fulgidus*. *Nature*, 390: 384.
- Korencic D, Ahel I, Schelert J, Sacher M, Ruan B, Stathopoulos C, Blum P, Ibba M, Söll D. 2004. A Freestanding Proofreading Domain is Required for Protein Synthesis Quality Control in Archaea. *Proc Nat Acad Sci USA*, 101: 10260–5.
- Laibelman AM. 2022. The Standard Genetic Code Lacks Redundancy for Amino Acid Codons. *BioRxiv.org*: 2022/519817.
- Laibelman AM. 2024. Lessons from Eukaryal tRNA: Fungi are not Mammals. *Preprints.org*: 202407.1791.v1.
- Levengood J, Ataide SF, Roy H, Ibba M. 2004. Divergence in Noncognate Amino Acid Recognition between Class I and Class II Lysyl-tRNA Synthetases. *J Biol Chem*, 279: 17707–14.
- Liepinsh E & Otting G. 1996. Proton Exchange Rates from Amino Acid Sidechains: Implications for Image Contrast. *Magn Reson Med*, 35: 30–42.
- Lofffield RB & Eigner EA. 1969. Mechanism of Action of Amino Acid Transfer Ribonucleic Acid Ligases. *J Biol Chem*, 244: 1746–54.
- Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, Lopez R, Butcher S. 2024. The EMBL-EBI Job Dispatcher Sequence Analysis Tools Framework in 2024. *Nucl Acids Res*, 52: W521–5.
- Mahapatra A, Srinivasan G, Richter KB, Meyer A, Lienard T, Zhang JK, Zhao G, Kang PT, Chan M, Gottschalk G, Metcalf WW, Krzycki JA. 2007. Class I and Class II Lysyl-tRNA Synthetase Mutants and the Genetic Encoding of Pyrrolysine in *Methanosarcina* spp. *Molec Microbiol*, 64: 1306–18.
- Mardanov AV, Gumerov VM, Beletsky AV, Prokofeva MI, Bonch-Osmolovskaya EA, Ravin NV, Skryabin KG. 2011. Complete Genome Sequence of the Thermoacidophilic Crenarchaeon *Thermoproteus uzoniensis* 768-20. *J Bacteriol*, 193: 3156–7.
- Marek CJ. 1995. Calculation of Kinetic Rate Constants from Thermodynamic Data. <https://ntrs.nasa.gov>
- Masamba W. 2021. Petasis vs. Strecker Amino Acid Synthesis: Convergence, Divergence and Opportunities in Organic Synthesis. *Molec*, 26: 1707.
- Nakamura A, Yao M, Chimnarong S, Sakai N, Tanaka I. 2006. Ammonia Channel Couples Glutaminase with Transamidase Reactions in GatCAB. *Science*, 312: 1954–8.

- Nureki O, Vassilyev DG, Tateno M, Shimada A, Nakama T, Fukai S, Konno M, Hendrickson TL, Schimmel P, Yokoyama S. 1998. Enzyme Structure with Two Catalytic Sites for Double-sieve Selection of Substrate. *Science*, 280: 578–82.
- O'Donoghue P & Luthey-Schulten Z. 2003. On the Evolution of Structure in Aminoacyl-tRNA Synthetases. *Microbiol + Molec Biol Rev*, 67: 550–73.
- Paris Z, Fleming IMC, Alfonzo JD. 2012. Determinants of tRNA Editing and Modification: Avoiding Conundrums, Affecting Function. *Semin Cell Dev Biol*, 23: 269–74.
- Perona JJ, Rould MA, Steitz TA. 1993. Structural Basis for Transfer RNA Aminoacylation by *Escherichia coli* Glutamyl-tRNA Synthetase. *Biochem*, 32: 8758–71.
- Pham Y, Kuhlman B, Butterfoss GL, Hu H, Weinreb V, Carter Jr CW. 2010. Tryptophanyl-tRNA Synthetase Urzyme: A Model to Recapitulate Molecular Evolution and Investigate Intramolecular Complementation. *J Biol Chem*, 285: 38590–601.
- Pines E, Magnes B-Z, Lang MJ, Fleming GR. 1997. Direct Measurement of Intrinsic Proton Transfer Rates in Diffusion-Controlled Reactions. *Chem Phys Lett*, 281: 413–20.
- Popow J, Schleiffer A, Martinez J. 2012. Diversity and Roles of tRNA Ligases. *Cell Mol Life Sci*, 69: 2657–70.
- Pyrkosz AB, Eargle J, Sethi A, Luthey-Schulten Z. 2010. Exit Strategies for Charged tRNA from GluRS. *J Mol Biol*, 397: 1350–71.
- Rogers LJ. 2023. Knowledge of Lateralized Brain Function Can Contribute to Animal Welfare. *Front Vet Sci*, 10: 1242906.
- Roh JH, Briber RM, Damjanovic A, Thirumalai D, Woodson SA, Sokolov AP. 2009. Dynamics of tRNA at Different Levels of Hydration. *Biophys J*, 96: 2755–62.
- Schleper C, Pühler G, Klenk H-P, Zillig W. 1996. *Picrophilus oshimae* and *Picrophilus torridus*. Two Species of Hyperacidophilic, Thermophilic, Heterotrophic, Aerobic Archaea. *Int J System Evol Microbiol*, 46: 814–6.
- Schmitt E, Panvert M, Blanquet S, Mechulam Y. 2005. Structural Basis for tRNA-Dependent Amidotransferase Function. *Struct*, 13: 1421–33.
- Schwartz MH, Pan T. 2016. Temperature Dependent Mistranslation in a Hyperthermophile Adapts Proteins to Lower Temperature. *Nucl Acids Res*, 44: 294–303.
- Sheppard K & Söll D. 2008. On the Evolution of the tRNA-dependent Amidotransferases, GatCAB and GatDE. *J Mol Biol*, 377: 831–44.
- Shimizu S, Czarina E, Juan M, Sato Y, Miyashita Y, Hoque MM, Suzuki K, Sagara T, Tsunoda M, Sekiguchi T, Dock-Bregeon A-C, Moras D, Takénaka A. 2009. Two Complementary Enzymes for Threonylation of tRNA in Crenarchaeota: Crystal Structure of *Aeropyrum pernix* Threonyl-tRNA Synthetase Lacking a cis-Editing Domain. *J Molec Biol*, 394: 286–96.
- Strecker A. 1854. Ueber einen neuen aus Aldehyd-Ammoniak-Blausäure entstehenden Körper. *Annalen der Chemie und Pharmacie*, 91: 349–51.
- Tamaki S, Tomita M, Suzuki H, Kanai A. 2018. Systematic Analysis of the Binding Surfaces between tRNA and Their Respective Aminoacyl tRNA Synthetase Based on Structural and Evolutionary Data. *Front Genet*, 8: 227.
- Wan W, Tharp JM, Liiu WR. 2014. Pyrrollysyl-tRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochim Biophys Acta*, 1844: 1059–70.
- Wang O, Parrish AR, Wang L. 2009. Expanding the Genetic Code for Biological Systems. *Chem Biol*, 16: 323–36.
- Weitzel CS, Li L, Zhang C, Eilts KK, Bretz NM, Gatten AL, Whitaker RJ, Martinis SA. 2020. Duplication of Leucyl-tRNA Synthetase in an Archaeal Extremophile may Play a Role in Adaptation to Variable Environmental Conditions. *J Biol Chem*, 295: 4563–76.

- Woese CR & Fox GE. 1977. Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms. *Proc Nat Acad Sci USA*, 74: 5088–90.
- Woese CR, Kandler O, Wheelis ML. 1990. Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya. *Proc Nat Acad Sci USA*, 87: 4576–9.
- Woese CR, Olsen GJ, Ibba M, Söll D. 2000. Aminoacyl-tRNA Synthetases, the Genetic Code, and the Evolutionary Process. *Microbiol Molec Biol Rev*, 64: 202–36.
- Wolff P, Villette C, Zumsteg J, Heintz D, Antoine L, Chane-Woon-Ming B, Droogmans L, Grosjean H, Westhof E. 2020. Comparative Patterns of Modified Nucleotides in Individual tRNA Species from a Mesophilic and Two Thermophilic Archaea. *RNA*, 26: 1957–75.
- Yoder JB, Clancey E, Des Roches S, Eastman JM, Gentry L, Godsoe, Hagey TJ, Jochimsen D, Oswald BP, Robertson J, Sarver BAJ, Schenk JJ, Spear SF, Harmon LJ. 2010. Ecological Opportunity and the Origin of Adaptive Radiations. *J Evol Biol*, 23: 1581–96.
- Zhang C-M, Perona JJ, Ryu K, Francklyn C, Hou Y-M. 2006. Distinct Kinetic Mechanisms of the Two Classes of Aminoacyl-tRNA Synthetases. *J Mol Biol*, 361: 300–11.
- Zhou X-L, Du D-H, Tan M, Lei H-Y, Ruan L-L, Eriani G, Wang E-D. 2011. Role of tRNA Amino Acid-accepting End in Aminoacylation and its Quality Control. *Nucl Acids Res*, 39: 8857–68.

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