Common evolutionary origins of the bacterial glycyl tRNA synthetase and alanyl tRNA synthetase

Claudia Alvarez-Carreño [1,2]; Marcelino Arciniega[3]; Lluis Ribas de Pouplana[4,5]; Anton S. Petrov[1,2]; Adriana Hernández-González[3]; Marco Igor Valencia-Sánchez[3]*; Loren Dean Williams[1,2,]*; and Alfredo Torres-Larios[3]

[1]NASA Center for the Origin of Life, Georgia Institute of Technology, Atlanta, GA 30332-0400, USA

[2]School of Chemistry and Biochemistry, Georgia Institute of Technology, 901 Atlantic Drive, Atlanta, GA 30332, USA

[3]Departamento de Bioquímica y Biología Estructural, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Circuito Exterior s/n, Ciudad Universitaria, Apartado postal 70-243, Mexico City, 04510, México.

[4]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Parc Científic de Barcelona, C/ Baldiri Reixac 10, 08028 Barcelona, Catalonia, Spain.

[5]Catalan Institution for Research and Advanced Studies, Barcelona, Catalonia 08010, Spain and 6 University Pompeu Fabra, Barcelona, Catalonia 08003, Spain.

* Corresponding authors:

Marco Igor Valencia-Sánchez

Email: igorval@comunidad.unam.mx

Loren Dean Williams

Email: loren.williams@chemistry.gatech.edu

Abstract

Glycyl tRNA aminoacylation is catalyzed by two different enzymes in different lineages of the Tree of Life. We have investigated the complex evolutionary history of one of these enzymes, the bacterial glycyl tRNA synthetase (bacGlyRS). Component domains of bacGlyRS show common ancestry with 1) the catalytic domain of class II tRNA synthetases; 2) the HD domain of the bacterial RNase Y; 3) the body and tail domains of the archaeal CCA-adding enzyme; 4) the anti-codon binding domain of the arginyl tRNA synthetase; and 5) a previously unrecognized domain that we call ATL (Ancient tRNA latch). The emergence of bacGlyRS in its extant form would have required the recruitment of domains from a pool of existing homologs, possibly at different times, by a process of molecular bricolage. A domain-by-domain analysis reveals that the catalytic and ATL domains in bacGlyRS are more closely related to domains in the universal alanyl tRNA synthetase (uniAlaRS) than to any other aminoacyl tRNA synthetase. The data suggest that bacGlyRS was present at the last bacterial common ancestor. Similarities between component domains of bacGlyRS and archaeal homologs, including the archaeal CCA-adding enzyme, suggest a complex evolutionary history. Here we explore a scenario in which bacGlyRS and uniAlaRS emerged from a shared core before the last universal common ancestor by a stepwise accumulation of protein domains during the dawn of biology.

Keywords

Translation, GlyRS, tRNA

Abbreviations and symbols

aaRS: aminoacyl tRNA synthetase
ABD: anti-codon binding domain
arcGlyRS: glycyl aminoacyl tRNA synthetase of archaea and eukarya, and some bacteria
ATL: Ancient tRNA latch domain
bacGlyRS: bacterial glycyl aminoacyl tRNA synthetase
CPR: Candidate Phyla Radiation
GlyRS: glycyl aminoacyl tRNA synthetase
LysRS: Lysyl aminoacyl tRNA synthetase
mRNA: messenger RNA

57 MSA: multiple-sequence alignment
58 RNase Y: ribonuclease Y
59 tRNA: transfer RNA
60 uniAlaRS: universal alanyl tRNA synthetase
61 uniArgRS: universal arginyl-tRNA synthetase
62

63 Introduction

64 Translation of mRNAs into coded polypeptides is universal to life (1). mRNA decoding
65 depends on the ribosome and a group of enzymes called aminoacyl tRNA synthetases (aaRSs) (2-
66 5). aaRSs determine the genetic code by recognizing specific tRNAs and linking them to their
67 cognate amino acids. aaRSs fall into two classes. The catalytic domain of Class I aaRSs is a
68 Rossmann fold while the catalytic domain of Class II aaRSs is an α/β three-layered sandwich. Class
69 I and Class II catalytic domains appear to have independent origins in the deep evolutionary past
70 (6; 7; 4). The early evolution of aaRSs has been linked to the origins of the genetic code and to
71 the emergence of the first proteins (8-12).

72 Throughout the phylogenetic tree, a given tRNA is typically aminoacylated by a single type
73 of aaRS. However, there are exceptions to this rule. There are two distinct Lysyl aaRSs (LysRSs)
74 (13). The class II LysRS is more abundant in bacteria, while class I LysRS is found predominantly
75 in archaea (14).

76 There are also two distinct GlyRSs; a GlyRS is found in archaea and eukarya, and in some
77 bacteria, called here arcGlyRS, and a bacterial GlyRS (bacGlyRS) found only in bacteria (15-17).
78 The two versions of GlyRS are globally different except for their catalytic domains (18). bacGlyRS
79 and arcGlyRS use different strategies to recognize $tRNA^{Gly}$. Here we ask why are there two GlyRSs.
80 How did they originate? Which GlyRS appeared first? Why is the distribution of the two GlyRSs
81 over the phylogenetic tree so patchy?

82 The bricolage hypothesis suggests that complex proteins can arise and evolve through
83 exaptation of pre-existing domains or modules, which can be combined to produce new functions
84 (19). In this hypothesis, domains can be thought of as evolutionary building blocks that provide
85 information on ancestry. arcGlyRS is relatively small, with a catalytic domain and a tRNA
86 recognition domain. bacGlyRS is larger and more complex. We and others (20; 17) have
87 hypothesized that the component domains of bacGlyRS were serially exapted through bricolage.

88 If the bricolage hypothesis is correct, homologies of individual domains can be used to
89 help reconstruct the complex histories of multidomain proteins such as bacGlyRS and can help
90 us understand ancestral relationships. Accurate histories require accurate domain delineations.

91 A domain is an independent and stable three-dimensional structure (21-23) and is a common unit
92 of evolution (24; 25). Domains are encoded by genes. Duplicated genes, called paralogs (26), can
93 combine to form a variety of distinct multi-domain proteins.

94 bacGlyRS contains α and β subunits. The combined data are interpreted here to indicate
95 that there are six component domains in the β subunit of bacGlyRS rather than five domains as
96 proposed previously (27). The domain of bacGlyRS that escaped detection in previous work is
97 extremely informative about bacGlyRS ancestry and shows homology to a domain found in the
98 universal alanyl tRNA synthetase (uniAlaRS), a class II aaRS. Sequence and structural comparisons
99 suggest that bacGlyRS is more closely related to uniAlaRS than to arcGlyRS (28; 29; 20).

100 Here, we show that bacGlyRS domains are homologous to universal, bacteria-specific or
101 archaea-specific proteins. The results suggest that bacGlyRS evolved by a stepwise bricolage. Six
102 component domains of bacGlyRS show common ancestry with six domains of four proteins that
103 interact with RNAs other than $tRNA^{Gly}$. Component domains of bacGlyRS show sequence and
104 structural similarities with 1) the catalytic and ATL domains (see below) of uniAlaRS; 2) the anti-
105 codon binding domain (ABD) domain of the universal arginyl-tRNA synthetase (uniArgRS); 3) the
106 body and tail domains of the archaeal CCA-adding enzyme; and 4) the HD domain of the bacterial
107 RNase Y (in agreement with previous observations of structural similarity (27)). This conservation
108 of RNA as substrate throughout a series of exaptation steps suggests an accessible mechanism
109 by which RNA-associated proteins are repurposed for new RNA-associated functions.

110

Results

bacGlyRS has a complex evolutionary history.

The phylogenetic distribution of bacGlyRS appears irregular and discontinuous ('patchy') over the extant twigs of the bacterial tree. We examined a broad range of archaeal and bacterial proteomes, some of which are only partial (Figure 1). Bacterial species that do not have bacGlyRS have arcGlyRS. Bacteria containing both bacGlyRS and arcGlyRS are extremely rare. Of the 542 sampled proteomes, arcGlyRS (GenBank: PSL06957.1) and bacGlyRS (GenBank: PSL02860.1) are found together only once, in the bacterium *Haloactinopolyspora alba* (NCBI RefSeq assembly: GCA_003014555.1) isolated from the Dead Sea sediment. Many bacterial phyla contain multiple branches containing either bacGlyRS or arcGlyRS. Firmicutes, Actinobacteria, Chloroflexi, Proteobacteria and DST, PVC, and FCB groups each contain species with bacGlyRS and species with arcGlyRS. Candidate Phyla Radiation (CPR) uniformly contains arcGlyRS; Cyanobacteria uniformly contains bacGlyRS. We currently lack a compelling mechanistic model that explains the global or local distribution of bacGlyRS and arcRNA over the bacterial tree.

bacGlyRS sequences are present in early diverging bacteria

The phylogenetic distribution of bacGlyRS suggests that it was present at the Last Bacterial Common Ancestor (LBCA). bacGlyRS is found in early diverging bacterial phyla (30) such as Firmicutes, Acidobacteria, and Proteobacteria (Figure 1). bacGlyRS is not found in the Candidate Phyla Radiation phylum. The presence of arcGlyRS in CPR proteomes may be due to horizontal gene transfer and not vertical inheritance. The CPR phylum contains putative symbiotic and parasitic bacteria with reduced genomes, large gene losses and extensive lateral gene transfer (31).

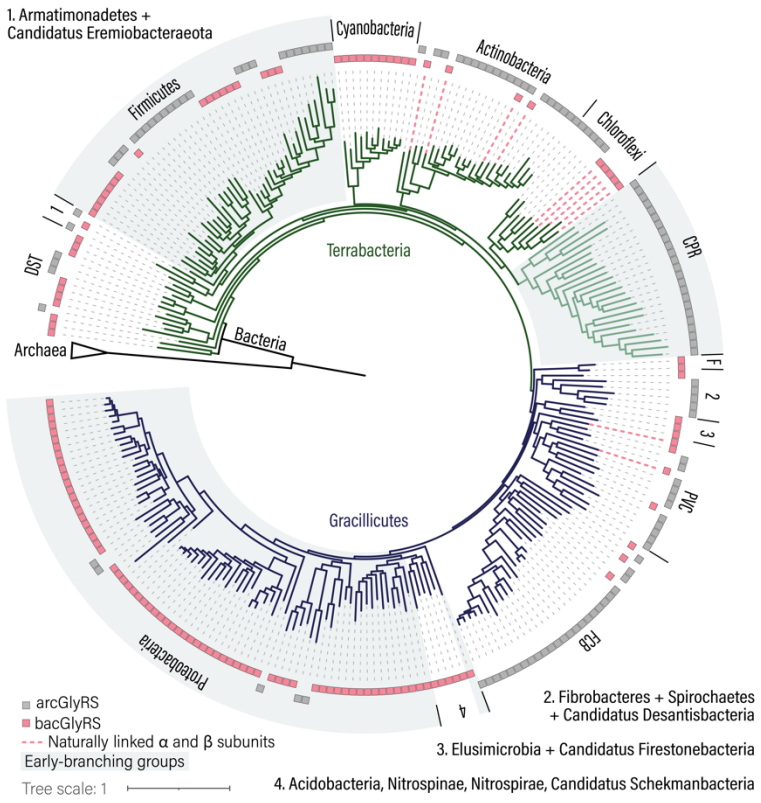

Figure 1. Phylogenetic distribution of the two versions of Glycyl tRNA ligase in Bacteria. The presence of bacGlyRS and arcGlyRS is mapped into a phylogeny of bacterial species. Pink dotted lines: bacGlyRS with naturally linked α and β subunits form heterodimers instead of heterotetramers. The topology of the tree was obtained from (32). DST: Deinococcus-Thermus, Synergistetes, Thermotogae, Caldiserica, Coprothermobacterota; CPR: Candidatus Phyla Radiation; PVC: Planctomycetes, Verrucomicrobia, Chlamydiae, Kiritimatiellaeota, Lentisphaerae, Candidatus Desantisbacteria, Candidatus Omnitrophica; FCB: Fibrobacteres, Chlorobi, Bacteroidetes, Gemmatimonadetes, Candidatus Cloacimonetes, Candidatus Fermentibacteria, Candidatus Glassbacteria; F. Fusobacteria.

Protein domains in bacGlyRS are related to RNA-modifying enzymes

Each domain of bacGlyRS has an independent evolutionary history. Six domains in bacGlyRS display strong sequence similarity to domains of other proteins (Figure 2). We determined which proteins are the closest sequence matches to each of the six domains of bacGlyRS (Figure 2D). The catalytic domain shows sequence similarity to the catalytic domain of uniAlaRS (e-value 1.1×10^{-07}). The body and tail domains show sequence similarity to domains of the archaeal CCA-adding enzyme (e-value 1.4×10^{-08}). The HD domain shows sequence similarity to a domain of ribonuclease Y (RNase Y) (e-value 5.7×10^{-08}), and the ABD shows sequence similarity to a domain of uniArgRS (e-value 2.9×10^{-17}).

142 We identified a previously unrecognized domain in bacGlyRS that we call Ancient tRNA
143 latch (ATL) domain. Our determiniation that ATL is a distinct domain follows previous
144 identification of this domain in uniAlaRS (33). If ATL is a distinct domain in uniAlaRS, then it is also
145 a domain in bacGlyRS. The ATL domains of uniAlaRS and bacGlyRS are similar to each other in
146 sequence and structure (Figure 3A). Previously, the ATL domain was described incorrectly as a
147 dependent element of the HD domain (27; 17; 34); we infer here from sequence and structural
148 considerations that the ATL is independent of other elements of bacGlyRS. The ATL domain of
149 uniAlaRS contains a C-terminal α-helical decoration that is absent from the ATL domain of
150 bacGlyRS.

151 In addition, we observe that the function of the ATL domain has been mischaracterized.
152 The ATL domain in uniAlaRS is called the "putative anticodon-binding domain" in some
153 classification databases (SCOPe ID a.203.1.1, (35), and ECOD ID 613.1.1.1 (36)). However, neither
154 the ATL domain of uniAlaRS nor the ATL domain of bacGlyRS contacts the anti-codon stem-loop
155 of tRNA (37; 38). Both interact extensively with the amino acid acceptor stem. This
156 mischaracterization of the ATL domain structure and function has obscured its evolutionary
157 origins and conserved molecular function.

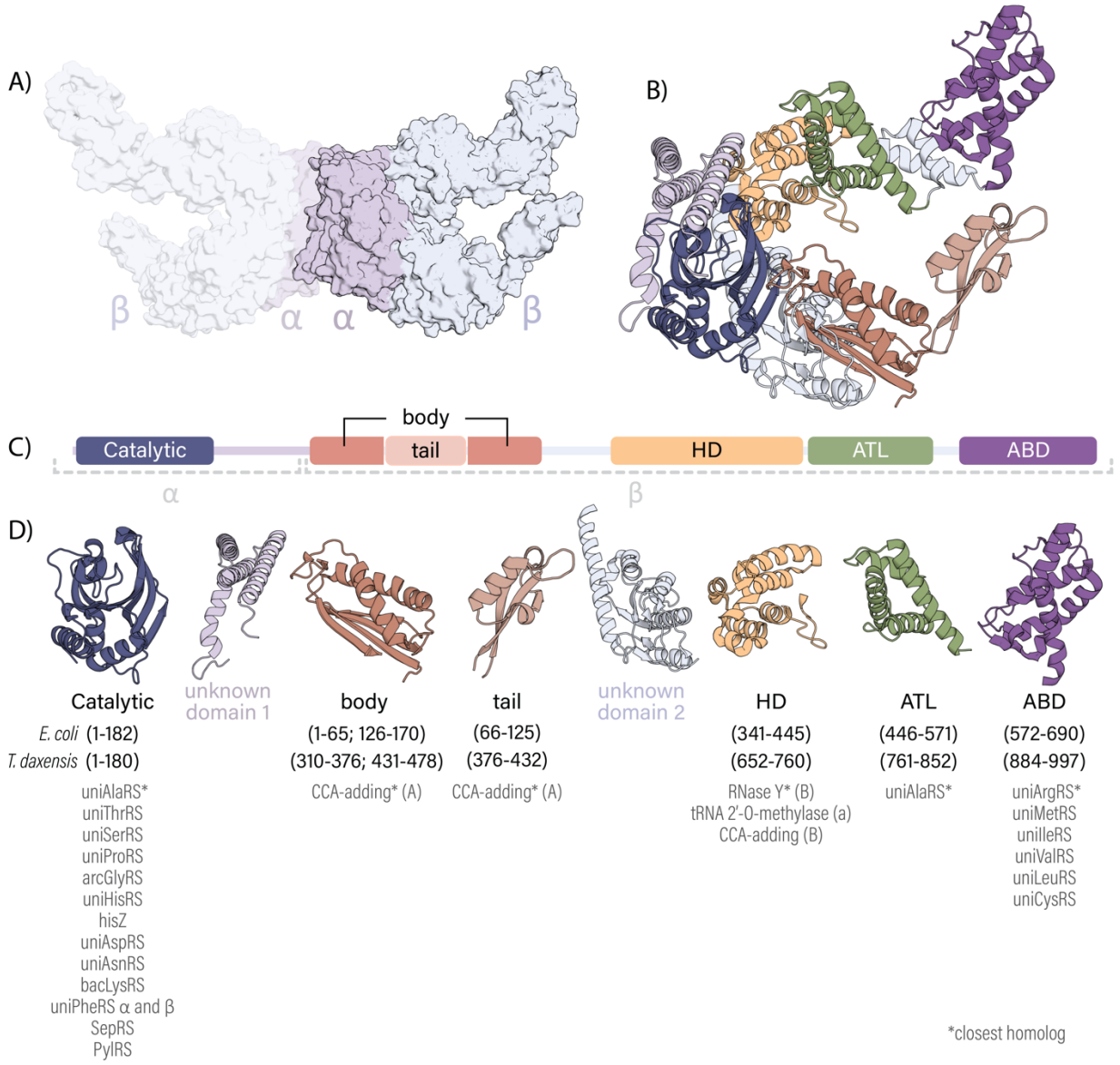

Figure 2. Domain organization of bacGlyRS. Structure of bacGlyRS of *Thermoanaerotrix daxensis* with naturally linked α and β-subunits (PDB: 7LU4). (A) Surface representation of bacGlyRS indicating the region that corresponds to the α- (light purple) and β- (light blue) subunits in other bacteria. (B) A ribbon representation of bacGlyRS shows structural units in different colors: dark blue, catalytic domain; light purple, unknown domain 1; dark orange, body domain; light orange, tail domain; light blue, unknown domain 2; yellow, HD domain; green, ATL domain; and dark purple, ABD domain. (C) Domains and structural units of bacGlyRS. (D) Proteins containing domains with high-scoring sequence similarity by bidirectional BLAST.

158 bacGlyRS and uniAlaRS homology is observed in key functional domains

159 The primary functions of aaRSs are to recognize tRNAs and amino acids and to covalently
160 join them together. The catalytic and ATL domains, which carry out these essential functions are
161 homologous between bacGlyRS and uniAlaRS.

162 The ATL domains of uniAlaRS and bacGlyRS recognize tRNAs by variations of a common
163 mechanism. $tRNA^{Gly}$ and $tRNA^{Ala}$ are recognized by uniAlaRS and bacGlyRS by interactions of ATL
164 with the acceptor stem (27; 17; 34). The tRNA acceptor stem is the first seven base pairs of tRNA,
165 the unpaired nucleotide at position 73, and the 3'-CCA tail. In species containing bacGlyRS, the
166 acceptor stems of $tRNA^{Gly}$ and $tRNA^{Ala}$ differ in three pairs and in position 73. These three base
167 pairs and position 73 are signatures (39), meaning that they are highly conserved within each
168 group, $tRNA^{Gly}$ or $tRNA^{Ala}$, but differ between the two groups. The specific signatures of $tRNA^{Gly}$
169 are base pairs C2:G71, G3:C70, G4:C69, and U73. The signatures of $tRNA^{Ala}$ are G2:C71, G3:U70,
170 C4:G69, and A73. In uniAlaRS, the signatures include the wobble base pair G3:U70, which is the
171 main identity determinant for $tRNA^{Ala}$ aminoacylation by uniAlaRS (40; 33; 37).

172 The ATL domain has adapted to distinguish between features of the amino acid acceptor
173 stems of $tRNA^{Gly}$ and $tRNA^{Ala}$ (Figure 3). We have identified four distinct ATL adaptations, allowing
174 a common fold to recognize different tRNAs. To discriminate between the two different tRNAs,
175 the ATL domain (i) slightly shifts global position relative to the tRNA, (ii) shifts local position of
176 the secondary structural elements by conformational change, and (iii) acquires α-helical
177 decorations.

178 The ATL domain undergoes modest global and local changes in position. After direct
179 superimpositions of $tRNA^{Gly}$ in complex with bacGlyRS (PDB 7YSE) and $tRNA^{Ala}$ in complex with
180 uniAlaRS (PDB 3WQY), the distance between the centers of mass of ^{bacGly}ATL and ^{Ala}ATL indicates
181 changes in the global pose of the ATL domains and reveals a global shift of 4.1 Å (Figure 3A). After
182 direct superimpositions of ^{bacGly}ATL and ^{Ala}ATL, the root mean square deviation (RMSD) between
183 the superimposed backbone atoms quantitates local changes in conformation and is 3.1 Å (Figure
184 3A). To describe local positional shifts, we have reindexed the ATL residues so that a common
185 index (subscripted here) indicates the same column in the sequence alignment. $^{bacGly}ATL_{10}$
186 indicates the amino acid at the 10th alignment position of ATL domain of bacGlyRS while $^{Ala}ATL_{10}$
187 indicates the amino acid of the ATL domain of uniAlaRS at the same alignment position (Figure
188 3F). Differences in local positioning of amino acids within ATL are seen by differences in ATL
189 indexes of a common interaction. For example, local positional shifts are seen in $^{bacGly}ATL_{29}$ and
190 $^{Ala}ATL_{28}$ interaction with the G1 base and the discriminator positions of the tRNAs (U73 in $tRNA^{Gly}$

191 and A73 in $tRNA^{Ala}$) and by $^{bacGly}ATL_{35}$ and $^{Ala}ATL_{34}$ interaction with C72 and the discriminator.
192 Base pair G1:C72 is conserved in both $tRNA^{Gly}$ and $tRNA^{Ala}$. By contrast, there is no shift at index
193 26. Both $^{bacGly}ATL_{26}$ and $^{Ala}ATL_{26}$ interact with position 70, which is a C in $tRNA^{Gly}$ and a U in $tRNA^{Ala}$.
194 The ATL domain interacts with the base of the discriminator position (U73 in $tRNA^{Gly}$, and A73 in
195 $tRNA^{Ala}$).

196 The ^{bacGly}ATL interaction with tRNA appears intrinsically less stringent than that of the
197 ^{Ala}ATL domain. The ^{bacGly}ATL domain contacts nucleotides of only one strand in the region of C68,
198 C69 and C70. It does not contact the opposing strand (G3, G4 or G5, Figure 3D). By contrast, ^{Ala}ATL
199 targets both strands of $tRNA^{Ala}$, interacting with both bases of base-pairs G3:U70 and C4:G69 in
200 part by using the ^{Ala}ATL-specific C-terminal decoration (Figure 3E).

201 The catalytic domain undergoes a radical change in global position. In uniAlaRS, the
202 catalytic and ATL domains are directly adjacent in sequence and structure. By contrast, in
203 bacGlyRS, the HD and BT domains are inserted between the catalytic and ATL domains. The HD
204 domain shifts and rotates, by 130°, the catalytic domain of bacGlyRS compared to that of
205 uniAlaRS (Figure 3B). To accommodate the change in the position of the catalytic domain of
206 bacGlyRS, the conformation of the four unpaired nucleotides of the acceptor arm of $tRNA^{Gly}$
207 bends (Figure 3C), such that the position of the terminal adenosine is shifted by ~ 26 Å in the
208 bacGlyRS complex compared to in the uniAlaRS complex.

209 The conformation of the CCA tail and the position of the terminal adenosine appear to be
210 unique in bacGlyRS compared to other aaRSs. The CCA tail of $tRNA^{Gly}$ bends away from the anti-
211 codon in the complex with bacGlyRS (34), shifting the terminal A toward the minor groove of the
212 tRNA. All other class II aaRSs bend the CCA tail in the opposite direction, towards the anticodon
213 (Figure S1) or, in the case of the uniAlaRS-tRNA complex (Figures 3A and 3C), maintain the relaxed
214 linear structure (41). The position of the terminal adenosine is shifted by ~ 30 Å in the tRNA
215 complex of bacGlyRS compared to in PheRS (Figure S1).

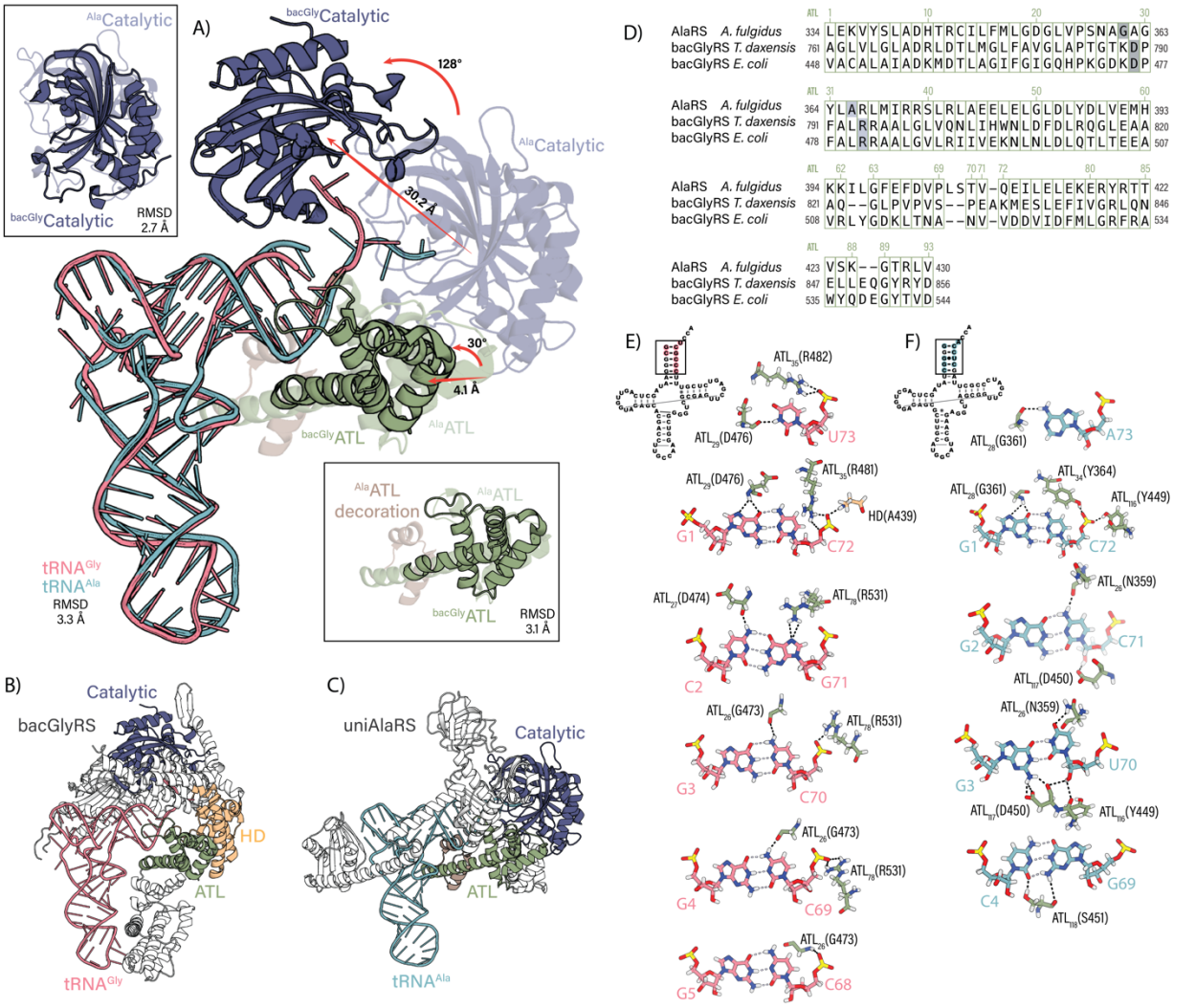

Figure 3. Structures of bacGlyRS and uniAlaRS showing regions of homology. (A) Comparison of the relative position of the catalytic and ATL domains of bacGlyRS (PDB: 7YSE) and uniAlaRS (PDB: 3WQY) with respect to their cognate tRNA molecules. Insets: Structure superimposition of the catalytic and ATL domains of bacGlyRS and uniAlaRS. (B) bacGlyRS of *Escherichia coli* in complex with tRNA. (C) uniAlaRS of *Archaeoglobus fulgidus* in complex with tRNA. (E) Hydrogen bonds between the ATL domain of bacGlyRS and its cognate tRNA. (F) Hydrogen bonds between the ATL domain of uniAlaRS and its cognate tRNA. bacGlyRS of *Escherichia coli* in complex with tRNA.

The catalytic and ATL domains of bacGlyRS resemble archaeal uniAlaRS homologs

We formulated a tentative model in which bacGlyRS originated in bacteria via duplication of a two-domain catalytic-ATL uniAlaRS module. This tentative model predicts that bacGlyRS component domains would be most similar to bacterial uniAlaRS homologs. Therefore, we investigated the similarity of bacGlyRS domains to archaeal and bacterial sequences.

The sequences of the catalytic and ALF domains in bacterial uniAlaRS and archaeal uniAlaRS are distinctive. A composite multiple-sequence alignment (MSA) containing both

223 bacterial and archaeal uniAlaRS sequences reveals lineage-specific signature positions and
224 insertions (Figure S2). Signature positions correspond to columns in a composite MSA that are
225 conserved within groups but are different between groups (39). Thus, differences between
226 bacterial and archaeal uniAlaRS sequences are not due to a simple compositional bias. This type
227 of pattern in a composite MSA is called a block structure and is observed in many universal aaRS
228 and other proteins of the translation system (3; 42).

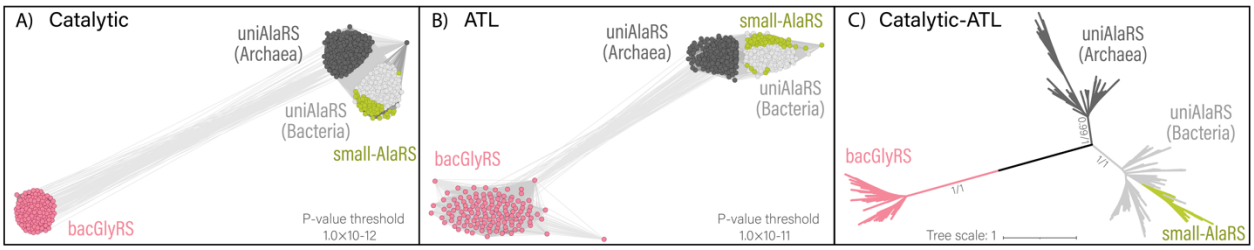

Figure 4. Sequence analysis of bacGlyRS protein domain homologs. Cluster map based on pairwise sequence similarities. Nodes represent sequences; connections between nodes indicate matches between pairs of sequences with a *P*-value lower than the threshold. (A) Pairwise sequence similarity between the catalytic domains of bacGlyRS and uniAlaRS. (B) Pairwise sequence similarity between the ATL domains of bacGlyRS and uniAlaRS. (C) Maximum-likelihood tree of catalytic and ATL domains including bacGlyRS and uniAlaRS sequences rooted at the midpoint. small-AlaRS are sequences from an anomalous subgroup of uniAaRS found in some archaea and bacteria.

229 The catalytic and ATL domains of bacGlyRS resemble archaeal homologs more closely
230 than bacterial homologs (Figures 4A and B). We compared the sequences of the catalytic and ATL
231 domains of bacGlyRS and uniAlaRS using the program CLANS (43; 44) and their profiles using the
232 program HH-align (45). Profiles are calculated from multiple-sequence alignments and reflect
233 position-specific sequence variability. A cluster map of sequences from the catalytic domains of
234 bacGlyRS and uniAlaRS at a P-value lower than 10^{-12} shows connections between bacGlyRSs and
235 archaeal uniAlaRSs. At this P-value threshold, the cluster map lacks connections between
236 bacGlyRSs and bacterial uniAlaRSs (Figure 4A). Similarly, a cluster map of sequences from the ATL
237 domains of bacGlyRS and uniAlaRS at a P-value threshold of 10^{-11} shows connections between
238 bacGlyRSs and archaeal uniAlaRSs. This cluster map lacks connections between bacGlyRSs and
239 bacterial uniAlaRSs (Figure 4B). Thus, the catalytic and ATL domains of bacGlyRS resemble
240 archaeal homologs and not bacterial homologs. Therefore, our tentative model in which
241 bacGlyRS originated by duplication of a bacterial catalytic-ATL uniAlaRS module is not supported
242 by the data (Figure 5D).

243 The catalytic and ATL domains represent an evolutionary and functional core

244 Canonical uniAlaRS has five component domains; (i) the catalytic domain, which
245 aminoacylates the tRNA, (ii) the ATL domain, which recognizes $tRNA^{Ala}$, (iii) a cradle loop barrel
246 domain, (iv) an editing domain and (v) a C-terminal domain (C-Ala) that binds to the tRNA elbow
247 formed by the D- and T- loops (46; 33) (Figures 3C and S1). We identified a subgroup of bacterial
248 and archaeal uniAlaRS sequences that have only three domains: catalytic, ATL, and editing
249 (Figures 4 and S1). uniAlaRSs in this subgroup lack the cradle loop barrel domain and the C-Ala
250 domain of canonical uniAlaRSs (Figure S2). We call this anomalous subgroup small-AlaRS. We
251 identified small-AlaRS in the proteomes of CPR bacteria, *Candidatus Thermoplasmatota*, and in
252 DPANN, and Asgard groups. small-AlaRS proteins have bacteria-specific insertions, suggesting
253 that sequences were transferred to some archaea via horizontal gene transfer. The maximum
254 likelihood tree of catalytic and ATL domain sequences supports a horizontal transfer hypothesis
255 of small-AlaRS to some archaea (Figure 4C and S3). Small-AlaRS sequences and recently
256 discovered viral mini-AlaRS sequences (47), which lack the cradle loop barrel, the editing and the
257 C-Ala domains of canonical cellular uniAlaRSs, support the hypothesis that a core composed of
258 the catalytic and ATL domains can be functional.

259 The combined data support a model in which bacGlyRS and uniAlaRS emerged before
260 LUCA from shared ancestors composed of independent catalytic and ATL domains (Figure 5A and
261 B). Other possibilities put the emergence of the bacGlyRS core after the divergence of LUCA. A
262 late emergence of bacGlyRS appears less parsimonious because it suggests that some bacteria
263 lost the ancestral GlyRS and replaced it with a protein that has no homologs in other bacteria.
264 Possible scenarios are 1) formation of the proto-bacGlyRS in the bacterial lineage independently
265 from the uniAlaRS catalytic-ATL core (Figure 5C); 2) formation of the proto-bacGlyRS by partial or
266 complete duplication of the uniAlaRS gene in the archaeal lineage, followed by a lateral gene
267 transfer of the proto-bacGlyRS from archaea to the bacterial lineage; and 3) formation of the
268 proto-bacGlyRS by partial or complete duplication of the uniAlaRS gene in the bacterial lineage
269 (Figure 5D). This last scenario is not supported by the data.

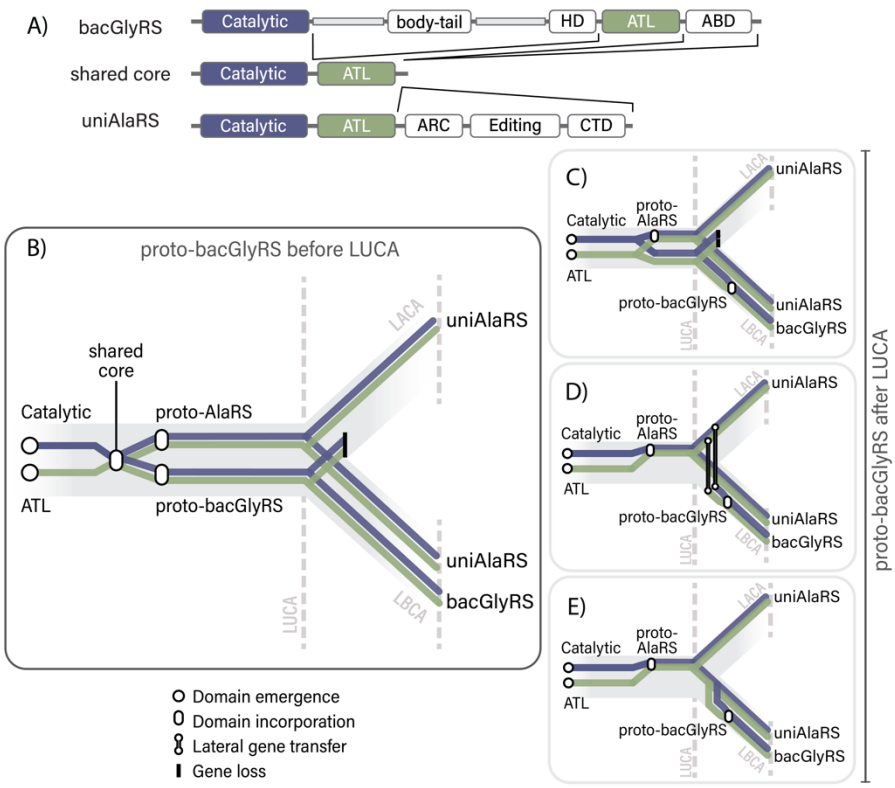

Figure 5. Emergence of bacGlyRS through molecular bricolage. (A) Comparative diagram of the multidomain arrangements of bacGlyRS and uniAlaRS. (B) Emergence of the common ancestor of bacGlyRS and uniAlaRS before the divergence of LUCA. Duplication of the common ancestor gives rise to the proto-uniAlaRS and proto-bacGlyRS. The ancestor of bacGlyRS is lost in archaea. (C) Emergence and duplication of the catalytic and ATL domains at LUCA. Before the divergence between archaea and bacteria, one copy of the catalytic domain and one copy of the ATL domain form the uniAlaRS core, the paralogs of the catalytic and ATL domains continue as single-domain proteins. Secondary loss of the copy of the catalytic and ATL domains in archaea. Formation of the bacGlyRS core in bacteria. (D and E) Evolution of the core of uniAlaRS before the divergence of LUCA. (D) The catalytic and ATL domains are horizontally transferred from the archaeal to the bacterial lineage and formation of the bacGlyRS core. (E) Duplication of the catalytic and ATL domains from uniAlaRS in bacteria and formation of the bacGlyRS core.

Discussion

The presence of bacGlyRS in a wide range of deeply rooted bacterial species suggests that it existed in the LBCA. The presence of bacGlyRS in Terrabacteria and Gracilicutes, the two major bacterial groups (Figure 1) supports this early origin. However, the distribution of bacGlyRS and arcGlyRS in the bacterial tree appears patchy and irregular. Currently, we do not have a model that fully explains the discontinuous distribution over evolutionary history. The dominance of arcGlyRS in CPR proteomes suggests that bacteria with reduced genomes prefer a smaller enzyme.

278 bacGlyRS is one of the largest aaRSs and is usually composed of two subunits, with a
279 catalytic domain in the α subunit and several recognition domains in the β subunit. In most
280 bacteria (Figure 1) bacGlyRS is a heterotetramer (4). In some bacteria, the α and β subunits are
281 linked in a single polypeptide chain (Figure 1); bacGlyRS with linked α and β subunits form
282 homodimers (Figure 2A).

283 A domain-by-domain analysis reveals evolutionary relationships between bacGlyRS
284 recognition domains and those of four different RNA-modifying proteins (Figure 2). The catalytic
285 domain was exapted (48) by bacGlyRS to perform aminoacylation. The ATL domain was exapted
286 to recognize the acceptor-stem of the $tRNA^{Gly}$. These two domains are closely related to domains
287 in the universal uniAlaRS. The HD domain was exapted apparently to interact with the acceptor
288 stem of $tRNA^{Gly}$ (17) and its presence influences the interaction of the catalalytic domain of
289 bacGlyRS and the CCA tail of $tRNA^{Gly}$ (discussed below). The HD domain in bacGlyRS is closely
290 related to a domain in the bacterial RNase Y. The body-tail domains were exapted by bacGlyRS
291 for interacting with the tRNA acceptor stem, these two domains perform an equivalent function
292 in the archaeal CCA-adding enzyme (49; 50). The ABD domain was exapted by bacGlyRS to
293 interact with the anti-codon loop (34) and is homologous to a domain in the class I uniArgRS.

294 Here we provide evidence that ATL is a distinct, independent domain of bacGlyRS rather
295 than a dependent component of the HD domain as previously proposed. This discovery sheds
296 light on the deep evolutionary history of aaRSs. Homology between the catalytic and ATL domains
297 of bacGlyRS and the corresponding domains of uniAlaRS suggests deep ancestry of their core
298 functions. It is likely that the two ancestral domains joined, and the resulting two-domain aaRS
299 gave rise to two separate enzymes, one for glycine and another for alanine. This model is
300 consistent with sequence and structural similarities between the core domains of bacGlyRS and
301 uniAlaRS.

302 The conservation of the pose of ATL domain in uniAlaRS and bacGlyRS, but not the pose
303 of the catalytic domain, is consistent with a model in which the ATL domain is ancestral to both
304 uniAlaRS and bacGlyRS (Fig. 5B). We have identified four adaptations of ATL, allowing a common
305 fold in a common pose to recognize different tRNAs. The ATL domain (i) slightly shifts global

306 position relative to the acceptor stem of the tRNA, (ii) shifts local positions of amino acid
307 backbone and sidechains by conformation change, (iii) changes sequence, and (iv) acquires a C-
308 terminal decoration in the uniAlaRS that is absent from the bacGlyRS.

309 bacGlyRS and uniAlaRS use different strategies to achieve final recognition of their
310 cognate tRNAs (Figure 3). ATL domains of both enzymes have similar interactions with the
311 discriminator base and the G1:C72 base pair which is conserved in $tRNA^{Gly}$ and $tRNA^{Ala}$. uniAlaRS
312 uses the ATL domain to interact intensively with the acceptor stem, employing the C-terminal
313 decoration to assist in targeting both strands of the acceptor step. The ATL domain in bacGlyRS
314 lacks C-terminal decoration and interacts less intensively with the acceptor stem. In the bacGlyRS
315 complex, the ATL domain only contacts nucleotides of one strand of the acceptor stem. To
316 compensate, bacGlyRS interacts with other regions of the tRNA. bacGlyRS but not uniAlaRS
317 exapted additional recognition domains through a process of bricolage.

318 The conformation of the CCA tail and the position of the terminal adenosine appear to be
319 unique in bacGlyRS compared to other aaRSs. In uniAlaRS, the catalytic and ATL domains are
320 directly adjacent in sequence and structure. $tRNA^{Ala}$ interacts with the CCA tail of uniAlaRS in a
321 relaxed state, which is reasonably similar to the structure in the free tRNA (51). In bacGlyRS,
322 insertion of the HD domain caused a translation and rotation of the catalytic domain relative to
323 the ATL domain (Figure 3B). To accommodate this translation/rotation, the acceptor-arm of
324 $tRNA^{Gly}$ in the bacGlyRS complex bends acutely in the direction of the minor groove (Figure 3C).

325 The sequences of bacGlyRS and archaeal uniAlaRS have distinctive features. The catalytic
326 and ATL domains of bacGlyRS more closely resemble archaeal homologs than bacterial homologs
327 (Figures 4A and B). The combined results suggest that bacGlyRS and uniAlaRS share a common
328 ancestor composed of the catalytic and ATL domains. The catalytic and ATL domains represent
329 an evolutionary and functional core. A model in which bacGlyRS originated by duplication of a
330 bacterial catalytic-ATL uniAlaRS module is not supported by the data (Figure 5D).

331 Here we propose a simple model, in which insertion of the body-tail and HD domains (or
332 possibly just the HD domain) into a gene closely related to the proto-AlaRS formed a novel
333 enzyme, a proto-bacGlyRS (Figure 5A). In this model, the interaction between ATL domain and

334 the tRNA remained relatively unchanged as the catalytic domain was displaced and the CCA tail
335 bends towards the minor groove. The proto-bacGlyRS enzyme underwent adaptations that
336 conferred a new amino acid specificity (Gly) and a new tRNA identity ($tRNA^{Gly}$). It is possible that
337 the new tRNA was better suited to the CCA tail orientation imposed by the HD insertion.

338 Our observations here reflect deep evolutionary history. Glycine and alanine appear to
339 be ancient amino acids, both prebiotically available and biologically important. They are among
340 the most abundant amino acids produced in spark discharge experiments (52) and are found in
341 chondrite asteroids (53). We propose that bacGlyRS and uniAlaRS emerged from a shared core
342 before LUCA (Figure 5A and B), from catalytic and ATL domains that had separate and distinct
343 origins. These prebiotic amino acids share the first and third codon positions, and the proteins
344 that aminoacylate $tRNA^{Gly}$ and $tRNA^{Ala}$ share sequence and structure relationships. It appears we
345 are seeing evidence of an ancestral protein with non-specific catalytic activity that may have been
346 able to link either glycine or alanine to an RNA, presumably an ancestor of tRNA. The deep
347 relationship between alanine and glycine is seen in their shared chemistry and biology.

348 # Methods

349 ## Sequence similarity search

350 A multiple sequence alignment (MSAs) of bacGlyRS sequences was calculated for a
351 representative set of archaea and bacteria. The MSA was trimmed to the domain boundaries as
352 defined by the structure analysis. The MSAs of the domains were converted to profile Hidden
353 Markov Models using the HHsuite version 3.3.0 (31). A profile-profile similarity search on the
354 Evolutionary Classification of Domains (ECOD) database was performed with HHsearch using the
355 profiles of bacGlyRS domains as queries. Target domains in ECOD yielding HHsearch probabilities
356 greater than 60% over more than 50 aligned columns were retrieved to serve as queries of a
357 second search on a sequence database containing the proteomes of the representative set of
358 bacterial and archaeal species. The second search was performed using jackhmmer from the
359 HMMER suite version 3.3.1 (31), with default parameters (five iterations). The ECOD database
360 only contains domains that are directly annotated on proteins with experimentally determined
361 structures. The second search aims to identify sequence similarity in proteins without

362 experimentally determined structures. Sequences passing the threshold values (E-value < $1\times10^{-}$
363 ^{3}, query coverage 80%) were retrieved and trimmed to the region of sequence similarity.

364 We broadly describe the phylogenetic distribution of the best hits according to the
365 number of sampled proteomes containing homologous sequences as follows: universal (presence
366 in > 50% of archaea and >50% of bacteria); archaea-specific (presence in >50% of archaea and
367 <10% of bacteria); bacteria-specific (presence in < 10% of archaea and <50% of bacteria); lineage-
368 specific within archaea (presence in <50% of archaea and <50% of bacteria); lineage-specific
369 within bacteria (presence in <10% of archaea and <50% of bacteria sampled); or of complex
370 distribution (presence in 10-50% of archaea and 10-50% of bacteria).

371 Cluster analysis

372 The query MSAs and trimmed target sequences retrieved from the second sequence
373 similarity search were clustered with CLANS (44) based on pairwise sequence similarities. To find
374 the best sequence match of each query domain, the cut-offs of the clustering were adjusted to
375 find the P-value for which the group of query sequences shows sequence relationships with only
376 one other group of orthologs. The subset of best matches was further clustered to reveal possible
377 differences between archaeal and bacterial sequences. Bacterial and archaeal uniAlaRS
378 sequences displayed differential similarity to bacGlyRS sequences on the cluster analysis. To
379 corroborate the clustering analysis results, additional analyses were performed at the level of
380 profiles.

381 Pairwise profile analysis: HHalign

382 An MSA containing orthologous sequences of uniAlaRS was further divided to include only
383 archaeal sequences or only bacterial sequences. The MSAs of archaeal and bacterial sequences
384 of uniAlaRS and the MSA of bacGlyRS were trimmed to the catalytic and ATL domains. The MSA
385 of the catalytic domain of bacGlyRS was aligned to the MSAs of the catalytic domains of archaeal
386 and bacterial uniAlaRS sequences using HHalign. The HHalign scores of these MSA-MSA
387 comparisons agree with the differential pairwise similarities observed in CLANS between
388 archaeal and bacterial sequences.

389 ML tree

390 The composite MSA of the catalytic domain containing sequences from bacGlyRS and
391 archaeal sequences from uniAlaRS was concatenated to the composite MSA of the ATL domain.
392 The concatenated MSA was trimmed with trimAl (54). A Maximum Likelihood tree was
393 calculated using the trimmed MSA using PhyML (55) and SMS model selection tool (56) on the
394 Montpellier Bioinformatics Platform. Branch support corresponds to a Bayesian-like
395 transformation of aLRT (aBayes) (57).

396 Data Availability

397 Sequence alignments and CLANS input files associated with this manuscript have been
398 deposited in the FigShare repository DOI: 10.6084/m9.figshare.23269007.

399 Acknowledgements

400 The authors dedicate this work to the memory of Professor ATL who passed away on September
401 9th, 2021. Professor ATL worked on the structural determination of the bacterial glycyl tRNA
402 synthetase and performed the initial structural analysis. The authors thank Jessica Bowman, Eric
403 Smith and Arturo Becerra for discussions. This work was funded by the National Aeronautics and
404 Space Administration grant 80NSSC18K1139. CAC was sponsored by the National Aeronautics
405 and Space Administration (NASA) through a contract with ORAU.

406 *Author contributions*: Conceptualization, ATL, MIVS and CAC. Methodology, CAC. Formal analysis
407 CAC, MA, AHG, ASP, MIVS and LDW. Writing – Original Draft, CAC and LDW. Writing – Review &
408 Editing, ASP, LRP and MIVS.

References

409 1. Bowman JC, Petrov AS, Frenkel-Pinter M, Penev PI, Williams LD (2020) Root of the Tree: The
410 Significance, Evolution, and Origins of the Ribosome. Chem Rev 120:4848-4878.
411 2. Burbaum JJ, Schimmel P (1991) Structural relationships and the classification of aminoacyl-
412 tRNA synthetases. J Biol Chem 266:16965-16968.
413 3. Woese CR, Olsen GJ, Ibba M, Söll D (2000) Aminoacyl-tRNA synthetases, the genetic code,
414 and the evolutionary process. Microbiol Mol Biol Rev 64:202-236.
415 4. Gomez MAR, Ibba M (2020) Aminoacyl-tRNA synthetases. RNA 26:910-936.

416 5. Giegé R, Eriani G (2023) The tRNA identity landscape for aminoacylation and beyond. Nucleic
417 Acids Res 51:1528-1570.
418 6. Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into
419 two classes based on mutually exclusive sets of sequence motifs. Nature 347:203-206.
420 7. Moras D (1992) Structural and functional relationships between aminoacyl-tRNA synthetases.
421 Trends Biochem Sci 17:159-164.
422 8. Hartman H (1995) Speculations on the evolution of the genetic code IV the evolution of the
423 aminoacyl-tRNA synthetases. Orig Life Evol Biosph 25:265-269.
424 9. Ribas de Pouplana L, Schimmel P (2001) Aminoacyl-tRNA synthetases: potential markers of
425 genetic code development. Trends Biochem Sci 26:591-596.
426 10. Fournier GP, Andam CP, Alm EJ, Gogarten JP (2011) Molecular evolution of aminoacyl tRNA
427 synthetase proteins in the early history of life. Orig Life Evol Biosph 41:621-632.
428 11. Koonin EV, Novozhilov AS (2017) Origin and evolution of the universal genetic code. Annu
429 Rev Genet 51:45-62.
430 12. Ribas de Pouplana L. The evolution of aminoacyl-tRNA synthetases: From dawn to LUCA.
431 (2020) The Enzymes. Elsevier, pp. 11-37.
432 13. Terada T, Nureki O, Ishitani R, Ambrogelly A, Ibba M, Söll D, Yokoyama S (2002) Functional
433 convergence of two lysyl-tRNA synthetases with unrelated topologies. Nat Struct Biol
434 9:257-262.
435 14. Tumbula D, Vothknecht UC, Kim H-s, Ibba M, Min B, Li T, Pelaschier J, Stathopoulos C,
436 Becker H, Söll D (1999) Archaeal aminoacyl-tRNA synthesis: diversity replaces dogma.
437 Genetics 152:1269-1276.
438 15. Ostrem DL, Berg P (1974) Glycyl transfer ribonucleic acid synthetase from Escherichia coli:
439 purification, properties, and substrate binding. Biochemistry 13:1338-1348.
440 16. Tang SN, Huang JF (2005) Evolution of different oligomeric glycyl-tRNA synthetases. FEBS
441 Lett 579:1441-1445.
442 17. Han L, Luo Z, Ju Y, Chen B, Zou T, Wang J, Xu J, Gu Q, Yang X-L, Schimmel P (2023) The
443 binding mode of orphan glycyl-tRNA synthetase with tRNA supports the synthetase
444 classification and reveals large domain movements. Science Advances 9:eadf1027.
445 18. Logan D, Mazauric M, Kern D, Moras D (1995) Crystal structure of glycyl-tRNA synthetase
446 from Thermus thermophilus. EMBO J 14:4156-4167.
447 19. Jacob F (1977) Evolution and tinkering. Science 196:1161-1166.
448 20. Dimas-Torres J, Rodríguez-Hernández A, Valencia-Sánchez M, Campos-Chávez E, Godínez-
449 López V, Rodríguez-Chamorro D, Grøtli M, Fleming C, Hernández-González A, Arciniega
450 M (2021) Bacterial glycyl tRNA synthetase offers glimpses of ancestral protein
451 topologies.
452 21. Rossman MG, Liljas A (1974) Letter: Recognition of structural domains in globular proteins. J
453 Mol Biol 85:177-181.
454 22. Levitt M, Chothia C (1976) Structural patterns in globular proteins. Nature 261:552-558.
455 23. Porter LL, Rose GD (2012) A thermodynamic definition of protein domains. Proc Natl Acad
456 Sci USA 109:9420-9425.
457 24. Ponting CP, Schultz J, Copley RR, Andrade MA, Bork P (2000) Evolution of domain families.
458 Advances in protein chemistry 54:185-244.
459 25. Ohno S. 2013. Evolution by gene duplication, Springer Science & Business Media.

460 26. Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst Zool 19:99-113.
461 27. Ju Y, Han L, Chen B, Luo Z, Gu Q, Xu J, Yang XL, Schimmel P, Zhou H (2021) X-shaped
462 structure of bacterial heterotetrameric tRNA synthetase suggests cryptic prokaryote
463 functions and a rationale for synthetase classifications. Nucleic Acids Res 49:10106-
464 10119.
465 28. Smith TF, Hartman H (2015) The evolution of Class II Aminoacyl-tRNA synthetases and the
466 first code. FEBS Lett 589:3499-3507.
467 29. Valencia-Sánchez MI, Rodríguez-Hernández A, Ferreira R, Santamaría-Suárez HA, Arciniega
468 M, Dock-Bregeon A-C, Moras D, Beinsteiner B, Mertens H, Svergun D, Brieba LG, Grøtli
469 M, Torres-Larios A (2016) Structural Insights into the Polyphyletic Origins of Glycyl tRNA
470 Synthetases. J Biol Chem 291:14430-14446.
471 30. Moody ERR, Mahendrarajah TA, Dombrowski N, Clark JW, Petitjean C, Offre P, Szöllősi GJ,
472 Spang A, Williams TA (2022) An estimate of the deepest branches of the tree of life from
473 ancient vertically evolving genes. eLife 11.
474 31. Achsel T, Stark H, Lührmann R (2001) The Sm domain is an ancient RNA-binding motif with
475 oligo(U) specificity. Proc Natl Acad Sci USA 98:3685-3689.
476 32. Moody ER, Mahendrarajah TA, Dombrowski N, Clark JW, Petitjean C, Offre P, Szöllősi GJ,
477 Spang A, Williams TA (2022) An estimate of the deepest branches of the tree of life from
478 ancient vertically evolving genes. eLife 11:e66695.
479 33. Naganuma M, Sekine S, Fukunaga R, Yokoyama S (2009) Unique protein architecture of
480 alanyl-tRNA synthetase for aminoacylation, editing, and dimerization. Proc Natl Acad Sci
481 USA 106:8489-8494.
482 34. Yu Z, Wu Z, Li Y, Hao Q, Cao X, Blaha GM, Lin J, Lu G (2023) Structural basis of a two-step
483 tRNA recognition mechanism for plastid glycyl-tRNA synthetase. Nucleic Acids Res
484 51:4000-4011.
485 35. Chandonia JM, Fox NK, Brenner SE (2017) SCOPe: Manual Curation and Artifact Removal in
486 the Structural Classification of Proteins - extended Database. J Mol Biol 429:348-355.
487 36. Schaeffer RD, Liao Y, Cheng H, Grishin NV (2017) ECOD: new developments in the
488 evolutionary classification of domains. Nucleic Acids Res 45:D296-d302.
489 37. Naganuma M, Sekine S, Chong YE, Guo M, Yang XL, Gamper H, Hou YM, Schimmel P,
490 Yokoyama S (2014) The selective tRNA aminoacylation mechanism based on a single
491 G•U pair. Nature 510:507-511.
492 38. Kumar A, Åqvist J, Satpati P (2019) Principles of tRNA(Ala) Selection by Alanyl-tRNA
493 Synthetase Based on the Critical G3·U70 Base Pair. ACS Omega 4:15539-15548.
494 39. Penev PI, Alvarez-Carreño C, Smith E, Petrov AS, Williams LD (2021) TwinCons: Conservation
495 score for uncovering deep sequence similarity and divergence. PLoS Comput Biol
496 17:e1009541.
497 40. McClain WH, Foss K, Jenkins R, Schneider J (1991) Four sites in the acceptor helix and one
498 site in the variable pocket of tRNA (Ala) determine the molecule's acceptor identity.
499 Proc Natl Acad Sci USA 88:9272-9276.
500 41. Mohan S, Hsiao C, VanDeusen H, Gallagher R, Krohn E, Kalahar B, Wartell RM, Williams LD
501 (2009) Mechanism of RNA double helix-propagation at atomic resolution. J Phys Chem B
502 113:2614-2623.

503 42. Vishwanath P, Favaretto P, Hartman H, Mohr SC, Smith TF (2004) Ribosomal protein-
504 sequence block structure suggests complex prokaryotic evolution with implications for
505 the origin of eukaryotes. Mol Phylogenet Evol 33:615-625.
506 43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search
507 tool. J Mol Biol 215:403-410.
508 44. Frickey T, Lupas A (2004) CLANS: a Java application for visualizing protein families based on
509 pairwise similarity. Bioinformatics 20:3702-3704.
510 45. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J (2019) HH-suite3
511 for fast remote homology detection and deep protein annotation. BMC Bioinformatics
512 20:473.
513 46. Swairjo MA, Otero FJ, Yang X-L, Lovato MA, Skene RJ, McRee DE, de Pouplana LR, Schimmel
514 P (2004) Alanyl-tRNA synthetase crystal structure and design for acceptor-stem
515 recognition. Mol Cell 13:829-841.
516 47. Antika TR, Chrestella DJ, Tseng YK, Yeh YH, Hsiao CD, Wang CC (2023) A naturally occurring
517 mini-alanyl-tRNA synthetase. Commun Biol 6:314.
518 48. Frenkel-Pinter M, Petrov AS, Matange K, Travisano M, Glass JB, Williams LD (2022)
519 Adaptation and Exaptation: From Small Molecules to Feathers. J Mol Evol 90:166-175.
520 49. Schimmel P, Yang XL (2004) Two classes give lessons about CCA. Nat Struct Mol Biol 11:807-
521 808.
522 50. Tomita K, Fukai S, Ishitani R, Ueda T, Takeuchi N, Vassylyev DG, Nureki O (2004) Structural
523 basis for template-independent RNA polymerization. Nature 430:700-704.
524 51. Westhof E, Dumas P, Moras D (1988) Restrained refinement of two crystalline forms of
525 yeast aspartic acid and phenylalanine transfer RNA crystals. Acta Crystallogr A 44 (Pt
526 2):112-123.
527 52. Parker ET, Cleaves HJ, Dworkin JP, Glavin DP, Callahan M, Aubrey A, Lazcano A, Bada JL
528 (2011) Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark
529 discharge experiment. Proc Natl Acad Sci USA 108:5526-5531.
530 53. Botta O, Glavin DP, Kminek G, Bada JL (2002) Relative amino acid concentrations as a
531 signature for parent body processes of carbonaceous chondrites. Orig Life Evol Biosph
532 32:143-163.
533 54. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated
534 alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-1973.
535 55. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms
536 and methods to estimate maximum-likelihood phylogenies: assessing the performance
537 of PhyML 3.0. Syst Biol 59:307-321.
538 56. Lefort V, Longueville JE, Gascuel O (2017) SMS: Smart Model Selection in PhyML. Mol Biol
539 Evol 34:2422-2424.
540 57. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O (2011) Survey of branch support
541 methods demonstrates accuracy, power, and robustness of fast likelihood-based
542 approximation schemes. Syst Biol 60:685-699.

543