Signatures of tRNA Glx -specificity in bacterial glutamyl-tRNA synthetases

Gautam Basu[1], Saumya Dasgupta[1], Aditya Dev[1], and Nipa Chongdar[1]

[1]Bose Institute Department of Biophysics

June 21, 2023

Abstract

The canonical function of glutamyl-tRNA synthetase (GluRS) is to glutamylate tRNA ^{Glu}. Yet, not all bacterial GluRSs glutamylate tRNA ^{Glu}; many glutamylate both tRNA ^{Glu} and tRNA ^{Gln}, while some glutamylate only tRNA ^{Gln} and not the cognate substrate tRNA ^{Glu}. Understanding the basis of this unique tRNA ^{Glx}-specificity is important. Mutational studies have hinted at hotspot residues, both on tRNA ^{Glx} and GluRS, that play crucial roles in tRNA ^{Glx}-specificity. But the underlying structural basis remains unexplored. Majority of biochemical studies related to tRNA ^{Glx}-specificity have been performed on GluRS from *Escherichia coli* and other proteobacterial species. However, since the early crystal structures of GluRS and tRNA ^{Glu}-bound GluRS were from non-proteobacterial species (*Thermus thermophilus*), the proteobacterial biochemical data have often been interpreted in the context of non-proteobacterial GluRS structures. Marked differences between proteo- and non-proteobacterial GluRSs have been demonstrated and therefore it is important that tRNA ^{Glx}-specificity be understood vis-a-vis proteobacterial GluRS structures. Towards this goal we have solved the crystal structure of GluRS from *E. coli*. Using the solved structure and several other currently available proteo- and non-proteobacterial GluRS crystal structures, we have probed the structural basis of tRNA ^{Glx}-specificity of bacterial GluRSs. Specifically, our analysis suggests a unique role played by a tRNA ^{Glx} D-helix contacting loop of GluRS in modulation of tRNA ^{Gln}-specificity. While earlier studies had identified functional hotspots on tRNA ^{Glx} that controlled tRNA ^{Glx}-specificity of GluRS, this is the first report of complementary signatures of tRNA ^{Glx}-specificity in GluRS.

Signatures of tRNA^{Glx}-specificity in bacterial glutamyl-tRNA synthetases

Saumya Dasgupta,#[1] Aditya Dev,#Nipa Chongdar,#[2] Gautam Basu*

Department of Biophysics, Bose Institute, P-1/12 CIT Scheme VIIM, Kolkata 700054, India.

# Equal contribution.

* Corresponding author (gautamda@gmail.com; gautam@jcbose.ac.in)

1. Current Address: Department of Chemistry, Amity Institute of Applied Sciences, Amity University Kolkata, WB 700135, India

2. Current Address: Interdisciplinary School of Life Sciences, Indian Institute of Technology, Goa, Farmagudi, Goa 403401, India. Email: nipa@iitgoa.ac.in

Abstract

The canonical function of glutamyl-tRNA synthetase (GluRS) is to glutamylate tRNA^{Glu}. Yet, not all bacterial GluRSs glutamylate tRNA^{Glu}; many glutamylate both tRNA^{Glu} and tRNA^{Gln}, while some glutamylate only tRNA^{Gln} and not the cognate substrate tRNA^{Glu}. Understanding the basis of this unique tRNA^{Glx}-specificity is important. Mutational studies have hinted at hotspot residues, both on tRNA^{Glx} and

GluRS, that play crucial roles in $tRNA^{Glx}$-specificity. But the underlying structural basis remains unexplored. Majority of biochemical studies related to $tRNA^{Glx}$-specificity have been performed on GluRS from *Escherichia coli* and other proteobacterial species. However, since the early crystal structures of GluRS and $tRNA^{Glu}$-bound GluRS were from non-proteobacterial species (*Thermus thermophilus*), the proteobacterial biochemical data have often been interpreted in the context of non-proteobacterial GluRS structures. Marked differences between proteo- and non-proteobacterial GluRSs have been demonstrated and therefore it is important that $tRNA^{Glx}$-specificity be understood vis-a-vis proteobacterial GluRS structures. Towards this goal we have solved the crystal structure of GluRS from *E. coli* . Using the solved structure and several other currently available proteo- and non-proteobacterial GluRS crystal structures, we have probed the structural basis of $tRNA^{Glx}$-specificity of bacterial GluRSs. Specifically, our analysis suggests a unique role played by a $tRNA^{Glx}$ D-helix contacting loop of GluRS in modulation of $tRNA^{Gln}$-specificity. While earlier studies had identified functional hotspots on $tRNA^{Glx}$ that controlled $tRNA^{Glx}$-specificity of GluRS, this is the first report of complementary signatures of $tRNA^{Glx}$-specificity in GluRS.

A short running title**:** $tRNA^{Glx}$-specific signatures in GluRS

Key words: GluRS, tRNA-Gln, tRNA-discrimination, *E. coli* , proteobacteria, protein-RNA interaction

1 INTRODUCTION

Protein-protein or protein-nucleic acid interactions drive a plethora of biological processes. For the interaction to be biologically fruitful, a protein must be capable of not only choosing its cognate partner from the cellular soup but also be able to discriminate against non-cognate partners present in its environment. A classic example is the case of aminoacyl-tRNA synthetases (aaRSs), whose function is to aminoacylate or charge its cognate tRNA and discriminate against all other non-cognate tRNA (1). It is now established that each tRNA type, corresponding to a particular amino acid, possesses unique nucleotides called identity determinants that allow its recognition by cognate aaRS and anti-determinants that discriminate against noncognate aaRSs (2). An added layer of recognition/discrimination is encountered in glutamyl tRNA-synthetases (GluRS) in some bacteria whose genomes lack glutaminyl-tRNA synthetase (GlnRS) (3,4).

GluRS is a class-I aminoacyl-tRNA synthetase that catalyzes the glutamylation of $tRNA^{Glu}$ (3). In absence of GlnRS, GluRS in these bacteria are non-discriminatory (ND) and glutamylates both $tRNA^{Glu}$ and $tRNA^{Gln}$. Glutamylation of $tRNA^{Gln}$ produces the mismatched product Glu-$tRNA^{Gln}$ (5,6). The misacylated Glu-$tRNA^{Gln}$ is then further edited to the correct product Gln-$tRNA^{Gln}$ by the enzyme glutamyl-$tRNA^{Gln}$ amidotransferase (gatCAB) through a transamidation pathway (7). All ancient versions of bacterial GluRS were $tRNA^{Gln}$-non-discriminatory GluRS (ND-GluRS). During the course of evolution, bacteria acquired GlnRS of an eukaryotic origin through horizontal gene transfer (4,8,9). The newly acquired GlnRS was evolutionarily selected in some bacteria, where the native ND-GluRS that charged both $tRNA^{Glu}$ and $tRNA^{Gln}$evolved into a $tRNA^{Gln}$-discriminatory (D-GluRS) that charged only $tRNA^{Glu}$. In addition to ND-GluRS and D-GluRS, a large number of proteobacterial species (for example *Helicobacter pylori* (10)*, Acidithiobacillus ferrooxidans* (11)) possess two copies of GluRS (GluRS1 and GluRS2) with distinct $tRNA^{Glx}$-specificities (GluRS1: mostly $tRNA^{Glu}$-specific; GluRS2: $tRNA^{Gln}$-specific), suggesting a gene duplication of the primordial version of GluRS.

The structural features of *Thermus thermophilus* GluRS (*Tth* -GluRS) have been extensively studied (12,13). GluRS is composed of two structural domains. The N-terminal domain, also known as the catalytic domain, contains the L-glu and ATP binding sites, along with the class-I specific signature motifs HIGH and KMSKS in the ATP binding site along with. This domain also encompasses a sparse binding interface with the acceptor stem and the D-helix nucleotides of $tRNA^{Glu}$. The other domain (C-terminal) interacts with the anticodon nucleotides of $tRNA^{Glu}$ and is aptly known as the anticodon-binding domain.

Although the most important structural insights for bacterial GluRSs came from *Tth* -GluRS, a non-proteobacterial GluRS, majority of biochemical studies for the mechanistic understanding of GluRS:tRNA interactions have been performed on GluRS/$tRNA^{Glu}$ in the proteobacterium *Escherichia coli* (14–17). Mutational studies performed on the *E. coli* GluRS (*Eco* -D-GluRS) and $tRNA^{Glu}$ identified several 'hot-spots'

(important amino acids and nucleotides) for efficient glutamylation reaction (16). For example, it has been shown that $tRNA^{Glu}$-specificity of GluRS in *E. coli* and some other proteo-bacterial GluRSs arise due to subtle conformational differences between $tRNA^{Glu}$ and $tRNA^{Gln}$, originating at the D-helix (Figure 1A, Figure S1) — augmented (presence of base-triple interaction 13:22:46 and the absence of nucleotide 47) in $tRNA^{Glu}$ versus non-augmented (absence of base-triple interaction 13:22:46 and the presence of nucleotide 47) in $tRNA^{Gln}$ (8,18). Interestingly, this is not true in case of non-proteobacterium *T. thermophilus* , which, despite possessing a D-GluRS, displays augmented D-helix in both $tRNA^{Glu}$ and in $tRNA^{Gln}$. A zinc ion present in the catalytic domain of *Eco* -GluRS was shown to play a critical role glutamylation reaction (19), although many bacterial GluRSs do not contain a bound Zn^{2+}, including *Tth* -GluRS, implying the irregular occurrence of the zinc atom (20). In another study, when an arginine residue (R266) in the tRNA-binding interface of *Eco* -GluRS was mutated to leucine, glutamylation efficiency of the protein was drastically reduced (more than 2500 fold) (16). Interestingly, sequence analysis of bacterial GluRSs revealed that this arginine residue is exclusively present only in proteobacterial GluRS. In other words, hot-spot signatures for GluRS-$tRNA^{Glx}$interaction are not homogeneously conserved in all bacterial GluRSs (4), indicating that factors responsible for the interactions are phylum-specific and not universal. Therefore, to completely understand sequence and structural signatures that drive specificity of $tRNA^{Glx}$-glutamylation reaction in bacteria, the sequence and structural signatures must be analyzed in a phylum-specific manner and the structural insights obtained from *Tth* -GluRS may not be enough to understand the results of the functional studies performed on *Eco* -GluRS (4).

From the perspective of GluRS evolution, the proteobacterial domain in bacteria had experienced multiple sets of events (horizontal gene transfer, gene duplication and perhaps domain fusion) while adapting a $tRNA^{Glu}$-specific aminoacylation pathway (4). In order to achieve $tRNA^{Glu}$-specificity, the evolving GluRS must have undergone major adaptations, especially in residues lining its tRNA-binding interface. To understand the rationale behind such adaptations in proteobacterial GluRS, structural insights from crystal structures of proteobacterial GluRS are needed. Further, since a number of mutational studies related to $tRNA^{Glx}$-specificity have been performed on *Eco* -GluRS, it is important to analyze*Eco* -GluRS structure to elucidate clues behind $tRNA^{Glx}$-specificity. Here, we report the crystal structure of *Eco* -GluRS. From structural and sequence analysis of a large number of bacterial GluRS, both proteo- and non-proteobacterial, we identify structural features present in proteobacterial GluRSs that are required for $tRNA^{Glu}$-specificity.

Our results highlight that the specific structural feature responsible for the $tRNA^{Glx}$-specificity of bacterial GluRSs is the presence or the absence of an unique "towards tRNA" or "away from tRNA" conformation adopted by a short loop connecting Helix 8 and Helix 9. The "towards tRNA" conformation is compatible with the augmented D-helix but incompatible with the non-augmented D-helix, while the "away from tRNA" conformation is compatible with both.

2 RESULTS AND DISCUSSION

Crystal structures of *Eco*-GluRS

The crystal structure of *Eco*- GluRS was solved after initial phase determination by molecular replacement method, using the structure of*Thermosynococcus elongates* GluRS (*Tel* -GluRS; pdb ID: 2cfo (6)) as the search model. The final structural model of *Eco* -GluRS (Figure 1B) could be refined to 3.3 Å resolution with R_{work} = 24.9 % and R_{free} = 29.0 % (Table 1). The crystallographic asymmetric unit contained two nearly identical molecules of *Eco* -GluRS (r.m.s.d. 0.62 Å over all the C-alpha atoms), bound to one zinc ion and one L-glu molecule. All residues of *Eco* -GluRS were visible in the electron density map, except Glu117 in chain B; side chain atoms in residues 110-114, 129, 377-381, 392 of chain A and 112, 117, 118, 126-132, 136, 423 of chain B were not resolved clearly. Like other bacterial GluRS, *Eco* -GluRS showed an elongated structure (110 Å long and 39 Å wide) consisting of the N-terminal catalytic domain and the C-terminal anticodon-binding domain.

Zn-coordination in *Eco*-GluRS and its comparison to other GluRSs

One Zn^{2+} ion was bound in the acceptor-stem binding domain of *Eco* -GluRS [Figure 1B]. Previous studies

on*Eco* -GluRS had shown that depletion of this zinc induced conformation changes in the protein, reducing its catalytic activity (19). It was proposed by Liu et al [26] that the zinc coordinating ligands in *Eco* -GluRS are Cys98, Cys100, Cys125 and His127 and the domain belong to the SWIM domain family (21). Contrary to the above claim, the Zn^{2+} ion in the *Eco* -GluRS structure is ligated by Cys98, Cys100, Cys125 and Tyr121, in a tetrahedral coordination [Figure 1C]. Residue His127, which was proposed to be the fourth coordinating ligand of Zn^{2+} points away from it. This observation supports our previous report (20), where we had predicted the coordinating ligands of zinc in *Eco* -GluRS to be Cys98, Cys100, Cys125 and Tyr121, based on its sequence similarity to GluRS from *Borrelia burgdorferri* (*Bbu* -GluRS; PDB ID: 4gri), the only other bacterial GluRS structure that contains a Zn^{2+} (22). It is worth mentioning here that the coordination environment of Zn^{2+} in *Eco* -GluRS is similar to that of YadB gene product of *E. coli* (23,24),*Eco* -Glu-Q-RS (pdb ID: 1nzj), the N-terminal only paralogue of GluRS. The Zn-binding domains of *Eco* -GluRS, *Bbu* -GluRS and*Eco* -Glu-Q-RS are shown superimposed in Figure 1C, along with a structure-guided sequence alignment in Figure S2. Despite large deletions, when compared to Eco-GluRS or Bbu-GluRS, the coordination residues and geometry of coordination in *Eco* -Glu-Q-RS were conserved. It is interesting to note that a conserved cation-π interaction between an arginine residue and the tyrosine coordinating the Zn^{2+} ion was also conserved in all.

$tRNA^{Gln}$-discrimination in bacterial GluRS

Eco -GluRS efficiently glutamylates $tRNA^{Glu}$ but is strictly discriminating against $tRNA^{Gln}$. The current understanding of the molecular origin behind the observed $tRNA^{Glx}$-specificity comes from experiments performed on $tRNA^{Gln}$. As shown in in Figure 1A, a feature of tRNA that was shown to play an important role in whether it is discriminated against or recognized by GluRS is whether its D stem was augmented (H-bonded 13:22 nucleotide pair) or non-augmented (non H-bonded 13:22 nucleotide pair). The augmented D stem was also shown to be correlated with the absence of nucleotide 47. For example, an augmented D-helix (and a deletion of nucleotide 47) in the $tRNA^{Glx}$ ($tRNA^{Glu1}$, $tRNA^{Glu2}$ and $tRNA^{Gln2}$ isoacceptors) in *A. ferrooxidans* was shown to be responsible for the efficient glutamylation of all three tRNAs by GluRS1 (11). On the other hand, the presence of a non-augmented D-helix in $tRNA^{Gln1}$ (the^{34}UUG^{36} isoacceptor) was responsible for the inability of GluRS1 to glutamylate the $tRNA^{Gln1}$ isoacceptor (discrimination). Similarly, when unique identity elements on $tRNA^{Glu}$ (U34, U35, C36, A37, G1*C72, U2*A71, U11*A24, U13*G22**A46, and Δ47) were transplanted into $tRNA^{Gln}$, the latter could be efficiently glutamylated by GluRS (10,25).

The clear signatures of $tRNA^{Glx}$-discrimination by GluRS on $tRNA^{Glx}$ beg a larger question – are there also signatures on GluRS, that dictate $tRNA^{Glx}$-discrimination by GluRS? If present, augmented/non-augmented tRNA discrimination signatures must lie in parts of GluRS that interacts with the $tRNA^{Glx}$ D-helix. An earlier report from our laboratory had shown that a C-terminal truncated version of *Eco*- GluRS could efficiently discriminate against $tRNA^{Gln}$ (15), suggesting the presence of $tRNA^{Gln}$-discriminatory features in the catalytic domain of *Eco* -GluRS. However, their exact identitity features on the protein have not been explored.

The H8-loop-H9 in *Eco*-GluRS as a $tRNA^{Gln}$-discriminatory feature

Having determined the structure of *Eco* -GluRS, we compared the sequence and structural features of its N-terminal catalytic domain with that of $tRNA^{Gln}$-bound *Eco* -GlnRS (PDB Id: 1gts). Being homologous, the overall folds of the two catalytic domains are similar. However, the structural superimposition brought out some important differences. Specifically, we looked at amino acid stretches in the two proteins at the binding interface of the D-helix region of $tRNA^{Gln}$, focussing on differences, since it must be this region that differentially interacts with the uniquely different D-helix regions of $tRNA^{Glu}$ and $tRNA^{Gln}$, triggering $tRNA^{Glx}$-specificity. The analysis identified two stretches (Figure 2), residues 303-335 in *Eco* -GlnRS and residues 257-311 in *Eco* -GluRS, both end-capped with helices (Helix 11 and Helix 12 in GlnRS; Helix 8 and Helix 10 in GluRS). While the two terminal helices are well superposed in both, the intervening stretches are not. The ˜10 residue inter-helical stretch in GlnRS assumes an extended structure and is proximal to the $tRNA^{Gln}$ D-helix. In contrast, the inter-helical stretch in GluRS is almost three times longer, of which the

conformation and $tRNA^{Gln}$ proximity of 10 residue stretch at the C-terminal are similar to the inter-helical stretch in GlnRS. However, conformation adopted by the first 20 residues (towards the N-terminal) in GluRS has no counterpart in GlnRS. This stretch contains a helix (Helix 9) in GluRS with no counterpart in GlnRS. In addition, it contains a loop between Helix 8 and Helix 9 that is at the interface of the D-helix of $tRNA^{Gln}$. In other words, GluRS exhibits a unique [Helix 8]-[loop]-[Helix 9] (H8-*L* -H9) motif situated at the $tRNA^{Gln}$ D-helix interface that is absent in GlnRS.

Database of curated H8-*L*-H9 motif sequences from bacterial GluRS

Is the unique D-helix interacting H8-*L* -H9 motif in GluRS a "protein" signature of $tRNA^{Glx}$-discrimination that complements the GluRS-discriminatory D-helix signature on "$tRNA^{Gln}$" ? To address this question, we first classified the H8-*L* -H9 motif of bacterial GluRSs. Subsequently, we sought a correlation between different classes of H8-*L* -H9 motif and the intrinsic $tRNA^{Gln}$-discrminatory character of GluRSs they belong to. In order to analyze GluRS sequences with a focus on the H8-*L* -H9 motif, a comprehensive and curated bacterial GluRS sequence database is required. We had earlier curated such a database (4), based on the presence/absence of a second copy of GluRS, the presence of GlnRS and the presence of gatCAB in each bacterial genome.

The presence of GlnRS in the genome signifies that the corresponding GluRS in the genome is $tRNA^{Gln}$-discriminatory (D-GluRS). Further, the GluRS is designated as D(-) if the genome lacks gatCAB (for which GluRS must strictly be $tRNA^{Gln}$-discriminatory since misacylated Glu-$tRNA^{Gln}$ cannot be transformed to Gln-$tRNA^{Gln}$) or D(+) if the genome contains gatCAB (the GluRS may not be strictly $tRNA^{Gln}$-discriminatory, since misacylated Glu-$tRNA^{Gln}$ can still be transformed to Gln-$tRNA^{Gln}$ by gatCAB). The absence of GlnRS in the genome (in this case the genome always contains gatCAB), and the presence of a single copy of GluRS in the genome signifies that the genomic GluRS is $tRNA^{Gln}$-non-discriminatory (ND-GluRS). When the genome lacked GlnRS but contained twin copies of GluRS, the GluRSs are designated as T1-GluRS and T2-GluRS. To summarize, D(-)-GluRS glutamylates only $tRNA^{Glu}$ and is strictly discriminatory against $tRNA^{Gln}$, D(+)-GluRS glutamylates $tRNA^{Glu}$ and possibly discriminates against $tRNA^{Gln}$, ND-GluRS glutamylates both $tRNA^{Glu}$ and $tRNA^{Gln}$. Experiments performed on a few twin GluRSs (10,11) suggest that T1-GluRS glutamylates $tRNA^{Glu}$ and discriminates against $tRNA^{Gln}$, while T2-GluRS possibly glutamylates $tRNA^{Gln}$ and not $tRNA^{Glu}$.

Following this nomenclature scheme, complete genomic sequences of 433 bacterial species were analyzed from the KEGG database (www.genome.jp/kegg) and annotated as D(-)-GluRS, D(+)-GluRS, ND-GluRS, T1-GluRS and T2-GluRS. Table S1 shows the sequence alignment of GluRS H8-*L* -H9 motifs for all bacterial GluRSs sequences used in this work, annotaed with the organism name (3 letter code used in the KEGG database) and the $tRNA^{Glx}$-discriminatory status, as arrived from whole genome analysis.

Principal Component Analysis of H8-loop-H9 motifs in bacterial GluRS

Analysis of the H8-*L* -H9 motif in the bacterial GluRS sequence database was performed using Principal Component Analysis(PCA). All H8-*L* -H9 sequences from bacterial GluRSs are shown projected on the PC1-PC2 plane (Figure 3A), where PC1 and PC2 correspond to collective sequence-axes associated with maximum mean square fluctuations. The H8-*L* -H9 sequences clustered broadly into three groups: (A) proteobacterial GluRSs that are incapable of glutamylating $tRNA^{Gln}$, (B) proteobacterial GluRSs that are capable of glutamylating $tRNA^{Gln}$, and, (C) all non-proteobacterial GluRS, irrespective of whether or not they can glutamylate $tRNA^{Gln}$. The PC2 axis separated the proteobacterial GluRSs (groups A & B) from non-proteobacterial GluRSs (group C), indicating that the sequence signature of the H8-*L* -H9 motif is distinctly different between proteobacterial and non-proteobacterial GluRSs. On the other hand, the PC1 axis separated the proteobacterial GluRSs depending on their $tRNA^{Gln}$-specificity (groups A and B), indicating that the H8-*L* -H9 motif is distinctly different between $tRNA^{Gln}$-discriminatory GluRSs (D-GluRS/T1-GluRS) and $tRNA^{Gln}$-non-discriminatory GluRSs (ND-GluRS/T2-GluRS).

PCA of H8-loop-H9 motifs in proteobacterial GluRS

To probe further, a separate PCA was performed only on the H8-L -H9 motifs from proteobacterial GluRSs (Figure 3B). As was seen earlier (Figure 3A), the tRNA^{Gln}-discriminatory and non-discriminatory GluRSs separated well on the PC1-PC2 plane, with the separation dominantly along the PC1 axis. The non-discriminatory GluRSs formed three subclusters along the PC1 axis – (A) α-proteobacterial T2-GluRS (α-T2), (B) δ-proteobacterial T2-GluRS/ND-GluRS (δ-T2) and γ-proteobacterial T2-GluRS (γ-T2) and (C) α-proteobacterial ND-GluRS (α-ND) and ε-proteobacterial T2-GluRS (ε-T2) – indicating phylum-specific differences in the H8-L -H9 motif among tRNA^{Gln}-non-discriminating GluRSs. The tRNA^{Gln}-discriminating proteobacterial GluRSs also showed clustering: (D) γ-proteobacterial D(+)-GluRS2 (γ-D+) (E) α-proteobacterial T1-GluRS (α-T1) and β-proteobacterial D(+)-GluRS2 (β-D+). Cluster F contained the rest of the tRNA^{Gln}-discriminating proteobacterial GluRSs (including some γ-D+, α-T1 and β-D+ some members of the three classes that appeared in clusters D and F). Exceptions to the overall differential appearance of tRNA^{Gln}-discriminating/non-discriminating GluRS were three members of the α-ND group that appeared in the tRNA^{Gln}-discriminating half, two in cluster-F (CCR *Caulobacter vibrioides* ; PZU *Phenylobacterium zucineum*) and one in cluster-E (PUB *Candidatus Pelagibacter ubique*), and one member of γ-T1 group that appeared in the tRNA^{Gln}-non-discriminatory half, in cluster-B (TGR *Thioalkalivibrio sulfidiphilus*). In summary, the distinct separation between discriminatory and non-discriminatory proteobacterial GluRSs indicated a significant role of the H8-L -H9 motif in tRNA^{Gln}-discrimination. The phylum-specific variations indicated that the H8-L -H9 motif may have evolved to facilitate tRNA^{Gln}-discrimination in a phylum-specific way.

PCA of H8-loop-H9 motifs in non-proteobacterial GluRS

The non-proteobacterial H8-L -H9 cluster of Figure 3A (cluster C) was analyzed further using a separate PCA. As can be seen from Figure 3C, the H8-L -H9 motifs, showed a bias in their appearance on the PC1-PC2 plane depending upon whether they belonged to tRNA^{Gln}-discriminating (red circles) or non-discriminating (blue circles) GluRS. However, there was substantial overlap. Since GluRS sequence variation can be highly phylum-specific, we then looked at only those cases where both discriminating and non-discriminating GluRSs were present within the same class. Two such classes, firmicutes (*Fi*) and hyperthermophilic bacteria (*Ht*) in our database contained GluRS of both kinds (T1-GluRS/T2-GluRS for *Ht* and D(+)-GluRS/ND-GluRS for *Fi*). The GluRS H8-L -H9 motifs in these two classes showed a clear sequence difference depending on their tRNA^{Gln}-specificity. For *Ht* (Figure 3D) all five pairs of T1-GluRS1/T2-GluRS separated into two distinct clusters. Sequence comparison of the two groups (Figure S3A) showed the unique presence of HPE(D)GK sequence at the center of the H8-L -H9 loop in T2-GluRSs absent in T1-GluRSs. Similarly, for *Fi* (Figure 3E), all H8-L -H9 motifs were separated into two clusters depending upon the tRNA^{Gln}-specificity of the corresponding GluRSs. Sequence comparison of the two groups (Figure S3B) showed the unique presence of Gly in ND-GluRSs (FMA is an exception); the D-GluRSs either lacked the Gly residue (replaced by Thr, Glu, Asp or Asn) or contained a 2-residue (WR) insertion at the center of H8-L -H9 motif.

Available crystal structures of bacterial GluRSs

Since the H8-L -H9 sequences separated proteobacterial GluRSs into two clusters which correlated well with their tRNA^{Gln}-specificities, we explored the structural differences of this motif in the available crystal structures of bacterial GluRSs (Table S2). Other than *Eco* (D-)-GluRS (this work, pdb ID: 8i9i), there are five more crystal structures of proteobacterial GluRSs: *Sml* -GluRS (*Stenotrophomonas maltophilia* GluRS, pdb ID: 7k86), *Xop* (D-)-GluRS (*Xanthomonas oryzae* GluRS, pdb ID: 5h4v), *Bte* (D+)-GluRS (*Burkholderia thailandensis*GluRS; pdb ID: 4g6z), *Hpy* -T1-GluRS (*Helicobacter pylori*GluRS, pdb ID: 6b1p) and *Pae* (D+)-GluRS (*Pseudomonas aeruginosa* GluRS, pdb ID: 5tgt). The H8-L -H9 motifs were also analyzed from seven available non-proteobacterial crystal structures:*Tma* (T1)-GluRS (*T. maritima* GluRS, pdb ID: 6b1p),*Bbu* (ND)-GluRS (*B. burgdorferi* GluRS, pdb ID: 6b1p),*Mtu* (ND)-GluRS (*Mycobacterium tuberculosis* GluRS, pdb ID: 6b1p), *Emg* -(D+)-GluRS (*Elizabethkingia meningoseptica*GluRS, pdb ID: 6b1p), *Tma* (T2)-GluRS (*Thermotoga maritima*GluRS, pdb ID: 6b1p), *Tth* (D+)-GluRS (*Thermus thermophilus*GluRS, pdb ID: 6b1p) and *Tel* (ND)-GluRS (*T. elongates*GluRS, pdb ID: 2cfo).

Τηε ριγιδ α-[>>τ] ςονφορματιον οφ H8-Λ-H9 ιν προτεοβαςτεριαλ ΓλυΡΣς

Sequence alignment of H8-L -H9 motifs from the 13 bacterial GluRSs with known crystal structures is shown Figure 4A. The top six sequences (ECO, SML, XOP, BTE, PAE and HPY) belong to the proteobacterial class of which the first four share a common loop sequence (LGWS-[HGD(Q/A)]-E(I/L)FT). We first focus on these four GluRSs whose H8-L -H9 conformations are shown superimposed in Figure 4B. The central –[HGD(Q/A)]– segment in these form a Type II' β-turn (φ_G: 62.1 ± 4.3, ψ_G: -134.5 ± 15.3 ; φ_D: -89.4 ± 12.6, ψ_D: -4.3 ± 9.4). As shown Figures 4A-B, the H8-L -H9 loop is tightly packed, with a number of participating hydrophobic/aromatic residues (W269, H271, F277, M/L/F282, Y/L/W285, F286; residue numbering according to the*Eco* -GluRS). Further, the side-chain of R266, appearing at the C-terminal end of Helix 8, protrudes into the loop and forms H-bonds with E275 side-chain and the backbone of W269, effectively stapling the two sides of the Type II' turn. The residue R266 was identified to be a proteobacterial GluRS specific residue and its mutation resulted in a significant decrease in the activity of *Eco* -GluRS (16). This results in a highly packed and rigid loop that displays a conserved Asp side-chain (D273) at its tip. We call the specific H8-L -H9 conformation, observed for the four proteobacterial GluRSs, as**τψπε α** (more types are discussed later) "towards tRNA" or**α-[>>τ]** conformation. Stapled by R266, the tightly packed α-[>>τ] ϛονφορματιον ις ριγιδ.

Ιντερασ τιον βετωεεν α-[>>τ] ϛονφορματιον ανδ αυγμεντεδ/νον-αυγμεντεδ τPNA$^{\Gamma\lambda\xi}$

Next we looked at the interactions between the α-[>>τ] H8-Λ -H9 ϛονφορματιον ανδ τPNA$^{\Gamma\lambda\xi}$ ωιτη αυγμεντεδ Δ-ηελιξ (A -τPNA$^{\Gamma\lambda\xi}$) ανδ ωιτη νον-αυγμεντεδ Δ-ηελιξ (N -τPNA$^{\Gamma\lambda\xi}$). Ιν τηε αβσενϛε οφ α στρυϛτυρε οφ προτεοβαϛτεριαλ A -τPNA$^{\Gamma\lambda\upsilon}$, τηε ιντερασ τιον βετωεεν α-[>>τ] ανδ τηε Δ-ηελιξ οφA -τPNA$^{\Gamma\lambda\upsilon}$ ωας μοδελλεδ υσινγ τηε στρυϛτυρε οφ$T\tau\eta$ -ΓλυΡΣ ϛομπλεξεδ ωιτη$T\tau\eta$ -A -τPNA$^{\Gamma\lambda\upsilon}$ (πδβ ΙΔ: 2ϛ0). Τηε$E\varsigma o$ -ΓλυΡΣ α-[>>τ] μοτιφ ωας στρυϛτυραλλψ συπεριμποσεδ ον το τηε H8-Λ -H9 μοτιφ οφ$T\tau\eta$ -ΓλυΡΣ ιν$T\tau\eta$ -ΓλυΡΣ::$T\tau\eta$ -A -τPNA$^{\Gamma\lambda\upsilon}$ ϛομπλεξ ανδ ιντερασ τιονς οφ α-[>>τ] ωιτη τηε Δ-ηελιξ οφ $T\tau\eta$ -A -τPNA$^{\Gamma\lambda\upsilon}$ ωερε αναλψζεδ. Ας σηοων Φιγυρε 5A, τηε μοδελλεδ στρυϛτυρε εξηιβιτεδ τωο προτειν-τPNA H-βονδς (Δ273-Γ23 ανδ Σ270-A14). Ιν αδδιτιον, τηε νεγατιελψ ϛηαργεδ οξψγεν ατομ οφ Δ273 σιδε-ϛηαιν ωερε φαοραβλψ πλαϛεδ ωιτηιν ιντερασ τινγ διστανϛε (4.4 Å) of G22 amino nitrogen atom. This clearly indicated that the α-[>>τ] ϛονφορματιονα μοτιφ ϛαν φαοραβλψ ιντερασ τ ωιτηA -τPNA$^{\Gamma\lambda\upsilon}$.

Ιντερασ τιονς βετωεεν α-[>>τ] ανδN -τPNA$^{\Gamma\lambda\nu}$ ωας αλσο γαυγεδ βψ μοδελινγ στυδιες. Τηε N-τερμιναλ δομαιν οφ $E\varsigma o$ -ΓλυΡΣ ωας συπεριμποσεδ ον τηε N-τερμιναλ δομαιν οφ $E\varsigma o$ -ΓλνΡΣ ιν τηε$E\varsigma o$ -ΓλνΡΣ::$E\varsigma o$ -N -τPNA$^{\Gamma\lambda\nu}$ ϛομπλεξ (Φιγυρε 2A) ανδ ιντερασ τιονς βετωεεν α-[>>τ] ($E\varsigma o$ -ΓλυΡΣ) ανδ$E\varsigma o$ -N -τPNA$^{\Gamma\lambda\nu}$ ωερε αναλψζεδ. Ας σηοων ιν Φιγυρε 5B, τηις ρεσυλτεδ ιν α σεερε στερις ϛλασ η βετωεεν Γ22/Γ23 οφN -τPNA ανδ Δ273/E275 οφ $E\varsigma o$ -ΓλυΡΣ. Ιντερασ τιονς βετωεεν$E\varsigma o$ -ΓλυΡΣ α-[>>τ] μοτιφ ανδN -τPNA$^{\Gamma\lambda\nu}$ ωερε αλσο αναλψζεδ βψ α διφφερεντ μοδελ. Ιν τηε σεϛονδ μοδελ, τηε τεμπλατε ωας $T\tau\eta$ -ΓλυΡΣ ϛομπλεξεδ ωιτη A -τPNA$^{\Gamma\lambda\upsilon}$. $E\varsigma o$ -ΓλυΡΣ α-[>>τ] μοτιφ ωας συπεριμποσεδ ον$T\tau\eta$ -ΓλυΡΣ ανδ $E\varsigma o$ -N -τPNA$^{\Gamma\lambda\nu}$ ωας συπεριμποσεδ ον τηε $T\tau\eta$ -A -τPNA$^{\Gamma\lambda\nu}$. Συβσεχυεντλψ, ιντερασ τιονς βετωεεν $E\varsigma o$ -ΓλυΡΣ α-[>>τ] ανδ$E\varsigma o$ -N-τPNA$^{\Gamma\lambda\nu}$ ωερε αναλψζεδ. Αλτηουγη τηερε ωερε νο στερις ϛλασηες ας ιν τηε πρειους ϛασε, φαοραβλε ιντερασ τιονς οφ Δ273 ωιτη Γ22 οφ A -τPNA$^{\Gamma\lambda\upsilon}$ (σεεν ιν Φιγυρε 5A) ωερε λοστ. Τηε τωο μοδελς ινδιϛατεδ τηατ τηε ριγιδ α-[>>τ] ϛονφορματιον ωουλδ ειτηερ γιε ρισε το στερις ϛλασηες ωιτη N -τPNA$^{\Gamma\lambda\nu}$ ορ λοσε φαοραβλε ιντερασ τιονς (ας σεεν ωιτηA -τPNA$^{\Gamma\lambda\upsilon}$) ορ βοτη. Ιν οτηερ ωορδς, τηε ριγιδ [>τ]-H8-Λ-H9 ϛονφορματιον σεεμς γεαρεδ τοωαρδς διϛριμινατινγ αγαινστ N-τPNA$^{\Gamma\lambda\nu.}$ Α κεψ ρεσιδυε ινολεδ βεηινδ τηε A -τPNA ϛομπατιβιλιτψ ανδ N -τPNA ινϛομπατιβιλιτψ οφ α-[>>τ] ις Δ273, προτρυδινγ τοωαρδς τPNA.

Α δψναμις α-[>>τ]-H8-Λ-H9 ϛονφορματιον ιν προτεοβαϛτεριαλ ΓλυΡΣ

In proteobacterial *Hpy* (T1)-GluRS, the H8-L -H9 loop length remains the same but the central four-residue sequence motif –[HGD(Q/A)]– observed for the rigid α-[>>τ] ϛονφορματιον ις ρεπλαϛεδ βψ -[ΨXΔK]-. Ιν τηε ϛρψσταλ στρυϛτυρε (Φιγυρε 5Δ), τηις στρετϛη φορμ α διστορτεδ Τψπε II β-turn (φ_Q: -129.1, ψ_Q: 66.8; φ_D: 60.3, ψ_D: 21.4) with side-chains of Asp protruding out of the turn almost overlapping with D273 of*Eco* -GluRS. However, B-factors for this stretch are quite high (Figure S4), indicating dynamics (electron density for the side-chain atoms of Q are also not seen). Therefore, we also used an AlphaFold model of *Hpy* (T1)-GluRS (with >95% confidence for the YQDK stretch), which is also shown superimposed in Figure 5D. Like in the rigid α-[>>τ] ϛονφορματιον, τηε ΨXΔK στρετϛη ιν τηε μοδελ φορμ α Τψπε II′ β-turn (φ_Q: 49.1,

ψ_Q: -137.8; φ_D: -92.8, ψ_D: 7.1) with side-chains of D, protruding out of the turn (overlapping with D273 of*Eco* -GluRS); in addition, unlike in the crystal structure, the backbone atoms of Gln overlaps well with the corresponding residue (Gly) in *Eco* -GluRS. Clearly, the H8-*L* -H9 motif in *Hpy* (T1)-GluRS is also an example of [>>t] motif, similar to α-[>>τ], βυτ μορε δψναμις ωηερε ΨΧΔΚ ςαν ειτηερ αδοπτ α Τψπε II′ ορ α διστορτεδ Τψπε II β-turn.

A δψναμις β-[>>τ]-H8-*Λ*-H9 ςονφορματιον ιν προτεοβαςτεριαλ ΓλυΡΣ

In the proteobacterial *Pae* -GluRS, H8-*L* -H9 loop length increases to 13 (from 12) and the central Type II' motif -HGD(Q/A)- of α-[>>τ] ις ρεπλαςεδ βψ α φιε-ρεσιδυε στρετςη (ΜΠΔΕΡ). Σινςε τηις στρετςη ις δισορδερεδ ιν τηε ςρψσταλ στρυςτυρε, ωε υσεδ ΑλπηαΦολδ το μοδελ τηις στρετςη (95% ςονφιδενςε ιν τηε ΜΠΔΕΡ στρετςη). Τηε H8-*Λ* -H9 ςονφορματιονς *οφΕςο* -ΓλυΡΣ ανδ *Παε* -ΓλυΡΣ αρε σηοων συπεριμποσεδ ιν Φιγυρε 5Ε. –[ΜΠΔΕ]- φορμς α Τψπε Ι β-turn (φ_P: -59.6, ψ_P: -29.6; φ_D: -93.1, ψ_D: 11.0) while E288 assumes a left-handed helical conformation (φ: 60.4, ψ: 15.6) and forms a H-bond with S284 (S270 in*Eco* -GluRS) and R305, along with several other H-bonds in the loop. Interestingly, when the H-bonded (and side-chain locked) E288 was allowed to assume other accessible side-chain rotameric states, its orientation overlapped with D273 of *Eco* -GluRS. Therefore, despite exhibiting a Type I β-turn motif arising from [MPDE], dissimilar to the earlier observed Type II' β-turn motif arising from [HGD(Q/A)], and a 13-residue loop, both H8-*L* -H9 conformations displayed a carboxylic side-chain protruding towards tRNA (D273 in *Eco* -GluRS and E288 in *Pae* -GluRS) compatible with *A* - but not*N* -type tRNA interaction. We call the H8-*L* -H9 conformation of *Pae* -GluRS as**β-[>>τ]** (type β "towards tRNA" conformation). The β-[>>τ] ςονφορματιον σεεμς το βε μορε δψναμις (νο ελεςτρον δενσιτψ ιν ςρψσταλ στρυςτυρε) τηαν τηε ριγιδ α-[>>τ] ςονφορματιον.

H8-*Λ*-H9 ςονφορματιον ιν νον-προτεοβαςτεριαλ ΓλυΡΣς: [<<τ] ςονφορματιον

Τηε α-[>>τ] ςονφορματιον *οφΕςο* -ΓλυΡΣς ωερε τηεν ςομπαρεδ ωιτη σεεν νον-προτεοβαςτεριαλ H8-*Λ* -H9 ςονφορματιονς. Ας σηοων ιν Φιγυρε 5Φ, τηε ςεντραλ ΗΓΔ(Χ/Α) σεχυενςε σιγνατυρε ις αβσεντ ιν αλλ νον-προτεοβαςτεριαλ H8-*Λ* -H9 μοτιφς. Ιν φιε ςασες – ΒΒΥ(ΝΔ), ΜΤΥ(ΝΔ), ΕΜΓ(Δ+), ΤΜΑ(Τ2) ανδ ΤΤΗ(Δ+) – τηερε ις αν ινσερτιον οφ Ε/Δ βετωεεν Γ ανδ Δ ιν τηε ΗΓΔ(Χ/Α) Τψπε II′ β-turn motif. For TEL(ND) there is a two-residue (EG) insertion, while the loop length in TMA(T1) is identical to*Eco* -GluRS. Upon superposition of the eight non-proteobacterial H8-*L* -H9 conformations onto the corresponding α-[>>τ] ςονφορματιον *οφΕςο* -ΓλυΡΣ (Φιγυρε 4"), ιτ ωας φουνδ τηατ εξςεπτ ΜΤΥ(ΝΔ) ανδ ΒΒΥ(ΝΔ) τηατ αρε δισςυσσεδ βελοω, αλλ σηοωεδ α λοοπ ςονφορματιον τηατ φαςεδ αωαψ φρομ τPNA. Ας οπποσεδ το τηε [>>τ] ςονφορματιον, τηε 'φαςινγ αωαψ φρομ τPNA' ςονφορματιον ις τερμεδ ας τηε[<<τ]ςονφορματιον.

H8-*Λ*-H9 ςονφορματιον ιν νον-προτεοβαςτεριαλ ΓλυΡΣς: δ/γ-[>>τ] ςονφορματιονς

Two proteins, *Bbu* -GluRS and *Mtu* -GluRS, displayed the [>>*t*] conformation, with K289 (BBU) and D294 (MTU) side-chains positions overlapping with D273 of*Eco* -GluRS. The Type II' β-turn sequence of *Eco* -GluRS (HGDQ) is replaced by ^{291}IADDH in *Mtu* -GluRS and^{286}YDDKR in *Bbu* -GluRS. The sequence stretch^{292}ADDH in *Mtu* -GluRS forms a Type I β-turn (φ_{D293}: -78.7, ψ_{D293}: -17.2; φ_{D294}: -104.3, ψ_{D294}: 3.3) where the side-chains of D294 and D273 (*Eco* -GluRS) overlap. With a 13-residue loop and a Type I β-turn at the center, this is similar to the conformation β-[>>τ] *οβσερεδ φορ Παε* -ΓλυΡΣ. Ον τηε οτηερ ηανδ, τηε σεχυενςε στρετςη ΨΔΔΚΡ ιν*Ββυ* -ΓλυΡΣ φορμς αν α-turn (Y_{CA}-R_{CA} distance: 6.9 Å; φ_{D287}: -133.1, ψ_{D287}: 11.6; φ_{D288}: 62.8, ψ_{D288}: 8.4; φ_{K289}: -112.2, ψ_{K289}: -59.4), with the side-chain of K289 oevrlapping with D273 (*Eco* -GluRS). With a 13-residue loop and an α-turn at the center, we call this conformation δ-[>>τ] (δ-type towards tRNA).

Validity of hypothesis that [>>*t*] is compatible with*A*-tRNA^{Glx} but not*N*-tRNA^{Glx}

Since our primary focus here is to understand the role played by the H8-*L* -H9 motif in recognizing/discriminating tRNA^{Glx} D-helix (augmented versus non-augmented), the genomic tRNA^{Gln} and tRNA^{Glu}sequences of all 13 bacterial species were examined (Figure S5) for the presence/absence of augmented/non-augmented D-helix in tRNA^{Glx}. The results are summarized in Table S2. The genomic

$tRNA^{Glx}$ is annotated with A (augmented) or N (non-augmented) corresponding to each bacterial species.

We first test our hypothesis that the rigid proteobacterial α-[>>τ]-H8-Λ -H9 ϛονφορματιονaλ μοτιφ φαορaβλψ ιντερaςτς ωιτηA -τPNA$^{\Gamma\lambda\xi}$ aνδ υνφaουρaβλψ ωιτηN -τPNA$^{\Gamma\lambda\xi}$ ον τηρεε προτεοβaςτεριaλ σπεςιες ωηοσε ΓλυPΣς δισπλaψ τηε α-[>>τ] ϛονφορματιον — E"O(Δ-), ΣMΛ(Δ-) aνδ ΞOΠ(Δ-) — aλλ aσσοςιaτεδ ωιτηA -τPNA$^{\Gamma\lambda\upsilon}$ aνδN -τPNA$^{\Gamma\lambda\nu}$. Iν aβσενςε οφ γaτ"AB (aλλ aρε Δ-), τηεσε τηρεε ΓλυPΣς μυστ στριςτλψ δισςριμιnaτε aγaινστN -τPNA$^{\Gamma\lambda\nu}$ aνδ φaορaβλψ ιντερaςτ ωιτηA -τPNA$^{\Gamma\lambda\upsilon}$. Ουρ ηψποτηεσις maτςηες ωιτη τηε στριςτ ρεχυιρεmεnτ οφ τηε ΓλυPΣς το βεN -τPNA$^{\Gamma\lambda\nu}$-δισςριmιnaτορψ. Τηε ηψποτηεσις ωaς τηεν τεστεδ φορ οτηερ ΓλυPΣς (σεε βελοω).

Τωο προτεοβaςτεριaλ σπεςιες BTE(Δ+) aνδ ΠAE(Δ+) ϛονταιν γaτ"AB ιν τηειρ γενομες συγγεστινγ τηaτ τηερε ις νο στριςτ ρεχυιρεmεnτ φορ$B\tau\epsilon$ (Δ+)-ΓλυPΣ aνδ $\Pi a\epsilon$ (Δ+)-ΓλυPΣ το δισςριmιnaτε aγaινστ τPNA$^{\Gamma\lambda\nu}$. Βοτη γενομες aρε aσσοςιaτεδ ωιτηA -τPNA$^{\Gamma\lambda\upsilon}$ aνδN -τPNA$^{\Gamma\lambda\nu}$ aνδ βοτη δισπλaψ τηε [>>τ] ϛονφορματιον (α- and β-). In absence of experimental data, our hypothesis suggests that both must be $tRNA^{Gln}$-discriminatory. The other proteobacterial species HPY (A -$tRNA^{Glu}$;N -$tRNA^{Glu}$) contain twin GluRSs. Among the two,Hpy (T1)-GluRS displayed the dynamic α-[>>τ] ϛονφορματιον, ϛομπaτιβλε ωιτη A - βυτ νοτ N -τPNA$^{\Gamma\lambda\xi}$.$H\pi\psi$ (T1)-ΓλυPΣ ηaς βεεν εξπεριmενταλλψ σηοων το ρεςογνιζεA -τPNA$^{\Gamma\lambda\upsilon}$ aνδ νοτN -τPNA$^{\Gamma\lambda\nu}$, ωηιςη ις ϛομπaτιβλε ωιτη ουρ ηψποτηεσις. $H\pi\psi$ (T2)-ΓλυPΣ ηaς βεεν εξπεριmενταλλψ σηοων το βε νον-φυνςτιονaλ aνδ τηερεφορε ις νοτ ινςλυδεδ ιν τηις δισςυσσιον.

Τωο νον-προτεοβaςτεριaλ NΔ-ΓλυPΣς ($M\tau\upsilon$ -ΓλυPΣ aνδ$B\beta\upsilon$ -ΓλυPΣ· βεινγ NΔ τηεψ μυστ ρεςογνιζε βοτη τPNA$^{\Gamma\lambda\upsilon}$ aνδ τPNA$^{\Gamma\lambda\nu}$) δισπλaψεδ τηε [>>τ] mοτιφ. Βοτη τPNA$^{\Gamma\lambda\upsilon}$ aνδ τPNA$^{\Gamma\lambda\nu}$ aρεA -τψπε ιν τηε MTϒ γενομε, ϛομπaτιβλε ωιτη τηε πρεσενςε οφA -τPNA$^{\Gamma\lambda\xi}$-ρεςογνιζινγ [>>τ]-H8-Λ -H9 mοτιφ. Iν ϛοντρaστ, τηε BBϒ γενομε ϛονταινςA -τPNA$^{\Gamma\lambda\upsilon}$/N -τPNA$^{\Gamma\lambda\nu}$. Τηε [>>τ]-H8-Λ -H9 mοτιφ ιν$B\beta\upsilon$ -ΓλυPΣ ις τηυς ϛομπaτιβλε ωιτηA -τPNA$^{\Gamma\lambda\upsilon}$ βυτ νοτ ωιτηN -τPNA$^{\Gamma\lambda\nu}$, aλτηουγη aς NΔ-ΓλυPΣ ιτ μυστ ρεςογνιζε βοτη. Iτ ις ποσσιβλε τηaτ ωιτη α λaργε aδδενδυm το ιτς "-τερmιναλ δομaιν aνδ τηε πρεσενςε οφ K ιν πλaςε οφ Δ aτ τηε τιπ οφ τηε τυρν τηaτ προτρυδες ιντο τPNA, τηε γεομετρψ οφ $B\beta\upsilon$ -ΓλυPΣ ωιτη τPNA maψ βε νον-ϛaνονιϛaλ aνδ νοτ τηε σamε aς τηε στρυςτυραλ mοδελς ϛονσιδερεδ ηερε.

Aλλ οτηερ (φουρ) ΓλυPΣς οριγινaτινγ φρομ σιξ βaςτεριaλ σπεςιες (Τaβλε Σ2) δισπλaψ τηε [τ>>]-H8-Λ -H9 mοτιφ, ϛομπaτιβλε ωιτη βοτη A - aνδ N -τψπε οφ PNA. Οφ τηεσε, TEΛ (A -τPNA$^{\Gamma\lambda\upsilon}$· N -τPNA$^{\Gamma\lambda\upsilon}$) ϛονταιν NΔ-ΓλυPΣ ωηιςη μυστ ρεςογνιζε βοτη τPNA$^{\Gamma\lambda\upsilon}$ aνδ τPNA$^{\Gamma\lambda\upsilon}$, ϛομπaτιβλε ωιτη ουρ ηψποτηεσις. TMA (A -τPNA$^{\Gamma\lambda\upsilon}$·N/A -τPNA$^{\Gamma\lambda\upsilon}$) γενομε ϛονταινς τωιν ΓλυPΣς.$T\mu a$ (T1)-ΓλυPΣ ηaς βεεν εξπεριmενταλλψ σηοων το ρεςογνιζε βοτη τPNA$^{\Gamma\lambda\upsilon}$ aνδ τPNA$^{\Gamma\lambda\nu}$, aλσο ϛομπaτιβλε ωιτη ουρ ηψποτηεσις. TTH (A -τPNA$^{\Gamma\lambda\upsilon}$·A -τPNA$^{\Gamma\lambda\upsilon}$) aνδ EMΓ (A -τPNA$^{\Gamma\lambda\upsilon}$· A -τPNA$^{\Gamma\lambda\upsilon}$) γενομες ϛονταιν ΓλνPΣ aνδ aρε δεοιδ οφ γaτ"AB, maκινγ τηε ϛορρεσπονδινγ ΓλυPΣς δισςριmιnaτορψ τοωaρδς A -τPNA$^{\Gamma\lambda\nu}$. Σινϛε τηε [τ>>]-H8-Λ -H9 mοτιφ, δισπλaψεδ βψ τηεσε τωο προτεινς, ις φaορaβλψ δισποσεδ το ιντερaςτ ωιτηA -τPNA$^{\Gamma\lambda\nu}$, τηε οριγιν οφA -τPNA$^{\Gamma\lambda\nu}$-δισςριmιnaτιον φορ τηεσε τωο νον-προτεοβaςτεριaλ ΓλυPΣς mυστ aρισε φρομ ιντερaςτιονς βετωεεν τηε ΓλυPΣς aνδ σομε νον-Δ-ηελιξ ελεmεnτ ιν τPNA. Φορ $T\tau\eta$ -ΓλυPΣ, ιτ ηaς βεεν εξπεριmενταλλψ σηοων τηaτ τPNA$^{\Gamma\lambda\nu}$-δισςριmιnaτιον ις mεδιaτεδ βψ P358 aτ τηε aντι ϛοδον-βινδινγ "-τερmιναλ δομaιν οφ $T\tau\eta$ -ΓλυPΣ (26).

Ὁρρελaτιον βετωεεν H8-Λ-H9 σεχυενςε aνδ γενομις A/N-τPNA$^{\Gamma\lambda\nu}$ιν προτεοβaςτερια

Τηε aβοε δισςυσσιον, ωηερε εξπεριmενταλλψ δετερmινεδ βaςτεριaλ ΓλυPΣ ϛρψσταλ στρυςτυρες ωερε υσεδ το προβε τηειρ ρολε, εσπεςιaλλψ τηε H8-Λ -H9 mοτιφ, ιν ρεςογνιζινγ ορ δισςριmιnaτινγ aγaινστA /N -τPNA$^{\Gamma\lambda\xi}$, ιδεντιφιεδ ϛερταιν σεχυενςε/στρυςτυραλ φεaτυρες οφ τηε H8-Λ -H9 mοτιφ το βε ρεσπονσιβλε φορ A -τPNA$^{\Gamma\lambda\xi}$ ρεςογνιτιον aνδN -τPNA$^{\Gamma\lambda\xi}$ δισςριmιnaτιον. Τηις ωaς εσπεςιaλλψ φουνδ το βε τρυε φορ τηε προτεοβaςτεριaλ ϛλaσς ωηοσε γενομες aρε κνοων το mοστλψ ϛονταιν A -τPNA$^{\Gamma\lambda\upsilon}$ aνδN -τPNA$^{\Gamma\lambda\nu}$. Εξπεριmενταλ ειδενςε φορA -τPNA$^{\Gamma\lambda\upsilon}$ ρεςογνιτιονN -τPNA$^{\Gamma\lambda\nu}$ δισςριmιnaτιον βψ προτεοβaςτεριaλ ΓλυPΣς aλσο ποιντ τοωaρδς τηε ιmπορτaνςε οφ τηε A/N -φεaτυρε οφ τPNA$^{\Gamma\lambda\xi}$ ιν τPNA$^{\Gamma\lambda\xi}$-σπεςιφιςιτψ οφ προτεοβaςτεριaλ ΓλυPΣς, φορ εξamπλε τηε ρολε οφ aυγmεnτεδ ερσυς νον-aυγmεnτεδ Δ-ηελιξ οφ τPNA$^{\Gamma\lambda\xi}$ ιν $E.$ $\varsigma o\lambda\iota$(18), $H.$ $\pi\psi\lambda o\rho\iota$ (10) aνδ $A.$ $\varphi\epsilon\rho\rho o\xi\iota\delta a\nu\varsigma$ (11). Ηερε ωε προβεδ τηε ιmπορτaνςε οφ τηε H8-Λ -H9 mοτιφ ιν προτεοβaςτεριaλ ΓλυPΣς, εξτενδινγ τηε aναλψσις το ϛaσες φορ ωηιςη ΓλυPΣ στρυςτυρες aρε νοτ aaιλaβλε.

Τηε aναλψσις ις ρεστριςτεδ το α σmaλλ συβσετ οφ T1/T2 ορ Δ/NΔ προβaςτεριaλ ΓλυPΣ φρομ τηρεε ϛλaσσες

(γ-,ε- and α-). Specifically, we focus on three class-specific groups of proteobacteria: i) ε-proteobacterial GluRS(T1/T2) pairs from 11 species (CJR, HPY, WSU, ABU, NIS, TDN, NAM, SKU, SDL, SUN, NSA), ii) γ-proteobacterial GluRS(T1/T2) pairs from 7 species (AFE, MCA, HHA, AEH, NOC, CBU, TGR), iii) α-proteobacterial GluRS(D+) from four species (BJA, OCA, NHA, RPD) and α-proteobacterial GluRS(ND) from 9 species (SME, ATU, RET, LAS, AEX, HCI, CCR, PZU, PUB). Sequence variations of the H8-L-H9 motif were analyzed in parallel with the corresponding D-helix features (either 'augmented' or 'non-augmented') of $tRNA^{Gln}$ and $tRNA^{Glu}$ isoacceptors.

All ε-proteobacterial species considered were associated with aA -$tRNA^{Glu}$, aN -$tRNA^{Gln1}$ isoacceptor and the complete absence of the $tRNA^{Gln2}$ isoacceptor ($^{34}CUG^{36}$) and (Figure 6A). For all GluRS(T1), the length of the loop between H8 and H9 helices is 12-residues long while in all GluRS(T2) it is 13-residue long. The 12-residue long loop exhibited the 'H(N)-G-D-Q(D)-E' signature motif (except HPY, which exhibits a 'Y-Q-D-K-E' motif) with a conserved Arg at position 266 (Fig. 6A and Figure S6). We had already shown that the YQDKE sequence stretch of HYP forms a dynamic α-[$>>\tau$] ϛονφορματιον, ινϛομπατιβλε ωιτη N -$\tau PNA^{\Gamma\lambda\nu}$, ωηερε ΨΧΔΚ φορμεδ α Τψπε ΙΙ′ ;-τυρν. Τηε ϛεντραλ ΧΔ ϛτρετϛη ιν τηε ;-τυρν ιν ΗΠΨ ις ρεπλαϛεδ βψ αν εχυαλλψ τυρν-ϛομπατιβλε ΓΔ ϛτρετϛη ιν αλλ οτηερ ε-proteobacterial GluRS(T1) proteins, and therefore we predict that, like*Hpy* -GluRS, all other ε-proteobacterial GluRS(T1) proteins considered here will be$N$ -$tRNA^{Gln}$-discriminatory. On the other hand since structural comparison showed that a 13-residue long loop, without the HGD(Q/A) motif at the center, gave rise to the [$<<t$]-H8-L -H9 motif, compatible with A -$tRNA^{Glx}$, we also predict that if H8-L -H9 motif is the sole basis for $tRNA^{Glx}$-specificity, then GluRS(T2) will charge both$N$ -$tRNA^{Gln1}$ and$A$ -$tRNA^{Glu}$.

Unlike ε-proteobacterial species, the seven γ-proteobacterial genomes considered here possess two $tRNA^{Gln}$ isoacceptors (Figure 6B and Figure S7): N -$tRNA^{Gln1}$($^{34}UUG^{36}$) and$A$ -$tRNA^{Gln2}$($^{34}CUG^{36}$). All H8-L -H9 motifs of GluRS(T1) displayed a 12-residue long loop with a central sequence signature HGDQE (and Arg266), implying an α-[$>>\tau$] ϛονφορματιον ϛομπατιβλε ωιτη A - βυτ νοτ N -$\tau PNA^{\Gamma\lambda\xi}$. Τηε ονλψ εξϛεπτιον ις ΤΓΡ (*Τηιοαλκαλιβριο συλφιδιπηιλυς*) τηατ ωας αν ουτλιερ ανδ αππεαρεδ ιν ϛλυϛτερ Β ιν Φιγυρε 3Β· ιτ δισπλαψϛ τηε ΨΕ″ΔΓ ϛεχυενϛε, ϛομπατιβλε ωιτη α [$<<\tau$] ϛονφορματιον (ϛομπατιβλε ωιτη βοτη A - ανδN -$\tau PNA^{\Gamma\lambda\xi}$). ἂνονιϛαλ [$>>\tau$] ϛονφορματιον-ϛομπετεντ ϛεχυενϛε μοτιφϛ ωερε αλϛο αβϛεντ ιν αλλ ΓλυΡΣ(T2), ινδιϛατινγ τηατ τηεϛε αλϛο αδοπτ τηε [$<<\tau$] ϛονφορματιον. Τηερεφορε, ουρ ηψποτηεϛις πρεδιϛτς αλλ ΓλυΡΣ(T2) ανδ*Τγρ* -ΓλυΡΣ(T1) το βε $\tau PNA^{\Gamma\lambda\nu}$-νον-δισϛριμινατορψ ωηιλε αλλ οτηερ ΓλυΡΣ(T1) το βεN -$\tau PNA^{\Gamma\lambda\xi}$-δισϛριμινατορψ (Ξ = Γλυ/Γλν). Τηις ις ϛονϛιϛτεντ ωιτη εξπεριμενταλ ρεϛυλτς (11) ον ΑΦΕ (*A. φερροοξιδανς*) ωηερε ιτ ωας ϛηοων τηατ ΓλυΡΣ(T1) δισϛριμινατες αγαινϛτN -$\tau PNA^{\Gamma\lambda\nu1}$($^{34}\Upsilon\Upsilon\Gamma^{36}$) ωηιλε ΓλυΡΣ(T2) ις νον-δισϛριμινατορψ.

Τηε γενομες οφ αλλ φουρ α-proteobacterial GluRSs(D+) considered here contain two $tRNA^{Gln}$ isoacceptors (Figure 6C and Figure S8): N -$tRNA^{Gln1}$ and$A$ -$tRNA^{Gln2}$. All GluRSs also display the α-[$>>\tau$] ϛονφορματιον προμοτινγ ϛεχυενϛε μοτιφ ΗΓΔΧΕ (ανδ Αργ266), ιμπλψινγ τηατ αλλ αρε εξπεϛτεδ το δισϛριμινατε N -$\tau PNA^{\Gamma\lambda\nu1}$ οερA -$\tau PNA^{\Gamma\lambda\nu2}$ ορA -$\tau PNA^{\Gamma\lambda\upsilon}$.

Αμονγ τηε γενομες οφ ϛιξ (ΣΜΕ, ΑΤΥ, ΡΕΤ, ΛΑΣ, ΑΕΞ, Η̈Ι) α-proteobacterial GluRSs(ND) considered here, two (AEX and HCI) contain two one and the rest contain two $tRNA^{Gln}$ isoacceptors (Figure 6D and Figure S9): N -$tRNA^{Gln1}$ and$A$ -$tRNA^{Gln2}$, and,A -$tRNA^{Glu}$. The GluRSs do not contain sequence motifs compatible with the [$>>t$] conformation implying that all are expected to be non-discriminating type. This is consistent with their functional requirement of compulsory glutamylation of both $tRNA^{Gln}$ and $tRNA^{Glu}$, in absence of GlnRS in their genomes.

The genomes of other three (CCR, PZU, PUB) α-proteobacterial GluRSs(ND) considered here (Figure 6E and Figure S9) contain only one $tRNA^{Gln}$ isoacceptor (A -$tRNA^{Gln1}$) and$A$ -$tRNA^{Glu}$. The GluRSs show the α-[$>>\tau$] ϛονφορματιον προμοτινγ ϛεχυενϛε μοτιφ ΗΓΔΔΕ (ᾸΡ, ΠΖΥ) ορ ΨΧΔΚΕ (ΠΥΒ) αλονγ ωιτη Αργ266· αλλ τηρεε αππεαρ ιν '$\tau PNA^{\Gamma\lambda\nu}$-δισϛριμινατορψ' ϛλυϛτερϛ ιν Φιγυρε 3Β (ᾸΡ/ΠΖΥ ιν ϛλυϛτερ Φ ανδ ΠΥΒ ιν ϛλυϛτερ Ε). Τηις ιμπλιες τηατ αλλ αρε εξπεϛτεδ το δισϛριμινατε N -$\tau PNA^{\Gamma\lambda\xi}$ οερA -$\tau PNA^{\Gamma\lambda\xi}$. Αλτηουγη$N$ -$\tau PNA^{\Gamma\lambda\xi}$ δισϛριμινατορψ, ϛινϛε τηειρ γενομες δο νοτ ϛονταιν N -$\tau PNA^{\Gamma\lambda\xi}$, εϛϛεντιαλλψ τηε ΓλυΡΣϛ αρε νον-δισϛριμινατορψ, ϛονϛιϛτεντ ωιτη τηειρ φυνϛτιονα̈λ ρεχυιρεμεντ.

τPNA$^{\Gamma\lambda\nu}$-δισςριμινατορψ ςτατυς οφ προτεοβαςτεριαλ ΓλυΡΣς ιν ουτλιερ ΠΓΑ ςλυςτερ Δ

ΠΓΑ περφορμεδ ον H8-Λ -H9 προτεοβαςτεριαλ μοτιφς ρεσυλτεδ ιν τηρεε ςλυςτερς, οφ ωηιςη ςλυςτερ Δ ωας μορε τοωαρδς τηε τPNA-νονδισςριμινατορψ ςλυςτερς αλονγ ΠΓ1. Τωο μεμβερς οφ ςλυςτερ Δ ηαε βεεν δισςυσσεδ: (ι) ΗΠΨ(T1), ωηιςη ωας δεσιγνατεδ αςN -τPNA$^{\Gamma\lambda\nu}$-δισςριμινατορψ βασεδ ον ιτς δψμανις α-[>>τ] ςονφορματιον ανδ (ιι) ΠΑΕ(Δ+), ωηιςη ωας δεσιγνατεδ αςN -τPNA$^{\Gamma\lambda\nu}$-δισςριμινατορψ βασεδ ον ιτς δψναμις β-[>>τ] ςονφορματιον. Ηερε ωε ςλοσελψ εξαμινε σεχυενςε ανδ στρυςτυραλ παττερνς οφ αλλ μεμβερς οφ ςλυςτερ Δ. Ιν αδδιτιον το ΠΑΕ, ςλυςτερ Δ ςονταινς 13 μορε Δ(+)-ΓλυΡΣς φρομ τηε γ-proteobacterial class (TTU, MMW, CSA, HEL, ABO, CJA, MAQ, HCH, PAR, SDE, AVN, ACI and MCT) whose H8-L-H9 sequences are very similar to that of *Pae* -GluRS (Figure S10), all associated with$A$ -tRNA^{Glu} and$N$ -tRNA^{Gln}. Since *Pae* -GluRS displayed the β-[>>τ]-H8-Λ -H9 μοτιφ, ςομπατιβλε ωιτη βοτη A -τPNA$^{\Gamma\lambda\upsilon}$ ανδ ινςομπατιβλε ωιτηN -τPNA$^{\Gamma\lambda\nu}$, ωε πρεδιςτ αλλ 13 Δ(+)-ΓλυΡΣς το βε τPNA$^{\Gamma\lambda\nu}$-νον-δισςριμινατορψ.

Ϋλυςτερ Δ ςονταινεδ 5 Δ(+)-ΓλυΡΣς φρομ τηε δ-proteobacterial class (HOH, DPS, DAK, DPR and BBA). Although the H8-L -H9 loop length in these were one residue longer than *Eco* -GluRS, interestingly, all except BBA displayed G-(D/T)-D sequence motif at its center (Figure S11A). In absence of experimental structures we assessed the loop conformations of all 5 GluRSs using AlphaFold models. As shown in Figure S11B, except BBA (that lacked the central G-(D/T)-D sequence motif) all H8-L-H9 motifs assumed the [>>t] conformation where the side-chain of the second Asp residue in G-(D/T)-D protruded out of the tip of the loop overlapping well with D273 of *Eco* -GluRS. Interestingly, like other [>>t] conformations encountered so far, the loop did not shown α- or β-turns; to emphasize this point we name this conformation as ε-[>>τ] (13-ρεσιδυε λοοπ ωιτη νο α-/β-turns in the loop). A summary of all types of [>>t] conformations identified so far is shown in Figure 7.

The presence of the N -tRNA^{Glx}-discriminating ε-[>>τ] ςονφορματιον ιμπλιες τηατ$Hoη$ -ΓλυΡΣ (ωιτη A -τPNA$^{\Gamma\lambda\nu1}$ ανδA -τPNA$^{\Gamma\lambda\nu2}$) ωιλλ βε τPNA$^{\Gamma\lambda\nu}$-νον-δισςριμινατορψ, $\Delta\pi\varsigma$ -ΓλυΡΣ (ωιτηN -τPNA$^{\Gamma\lambda\nu1}$) ωιλλ βε τPNA$^{\Gamma\lambda\nu}$-δισςριμινατορψ, $\Delta\alpha\kappa$ -ΓλυΡΣ ανδ ανδ$\Delta\pi\rho$ -ΓλυΡΣ (βοτη ωιτη N -τPNA$^{\Gamma\lambda\nu1}$ ανδA -τPNA$^{\Gamma\lambda\nu2}$), ον τηε οτηερ ηανδ, ωιλλ βε τPNA$^{\Gamma\lambda\nu2}$-νον-δισςριμινατορψ βυτ τPNA$^{\Gamma\lambda\nu1}$-δισςριμινατορψ. $B\beta a$ -ΓλυΡΣ (ωιτη τηε αωαψ φρομ τPNA ςονφορματιον [τ>>] ανδ N -τPNA$^{\Gamma\lambda\nu1}$) ωιλλ βεN -τPNA$^{\Gamma\lambda\nu}$-νον-δισςριμινατορψ.

3 ϒΟΝϒΛΥΣΙΟΝ

Φρομ αν αναλψσις οφ ααιλαβλε ςρψσταλ στρυςτυρες οφ βαςτεριαλ ΓλυΡΣς, ινςλυδινγ $Eςo$ -ΓλυΡΣ τηατ ωε ρεπορτ ηερε, ανδ βαςτεριαλ ωηολε γενομες το ιδεντιφψ τηε πρεσενςε ορ αβσενςε οφ ΓλνΡΣ ανδ γατϒΑΒ, ωε ηαε ιδεντιφιεδ α λοοπ, σπαννινγ βετωεεν Ηελιξ 8 ανδ Ηελιξ 9, το βε ρεσπονσιβλε φορ τPNA$^{\Gamma\lambda\nu}$-δισςριμινατιον βψ σομε βυτ νοτ αλλ βαςτεριαλ ΓλυΡΣς. Σπεςιφιςαλλψ ωε σηοω τηατ τηε λοοπ, ωηιςη ιντεραςτς ωιτη τηε Δ-ηελιξ οφ τPNA$^{\Gamma\lambda\xi}$, βροαδλψ αδοπτς τωο ςονφορματιονς: [>>τ] (τοωαρδς PNA) ορ [τ>>] (αωαψ φρομ PNA). Τηε [τ>>] ςονφορματιον ις ςομπατιβλε ωιτη τηε τωο ςανονιςαλ ςονφορματιονς οφ τηε Δ-ηελιξ: αυγμεντεδ ανδ νον-αυγμεντεδ. Τηε [>>τ] ςονφορματιον, ωηιςη αππεαρς ιν αριους φορμς (Φιγυρε 7), δισπλαψς α ςαρβοξψλις σιδε-ςηαιν τοωαρδς τPNA$^{\Gamma\lambda\xi}$, τηατ παρτιςιπατες ιν φαοραβλε ιντεραςτιονς ωιτη τηε αυγμεντεδ Δ-ηελιξ βυτ ειτηερ λοσες τηε φαοραβλε ιντεραςτιονς ορ γενερατες στερις ςλασηες (ορ βοτη) ωηεν ιντεραςτινγ ωιτη τηε νον-αυγμεντεδ Δ Ηελιξ. Τηε [>>τ] ςονφορματιον ις μοστ πρεαλεντ ιν προτεοβαςτεριαλ ΓλυΡΣς ωηιςη αλσο δομιναντλψ ποσσεσς Δ-ηελιξ αυγμεντεδ τPNA$^{\Gamma\lambda\upsilon}$ ανδ νον-αυγμεντεδ τPNA$^{\Gamma\lambda\nu}$. Τηατ τPNA$^{\Gamma\lambda\nu}$-δισςριμινατιον βψ προτεοβαςτεριαλ ΓλυΡΣς αρε μεδιατεδ βψ τηε [>>τ] ςονφορματιον ωας αλιδατεδ ωιτη κνοων εξπεριμενταλ δατα ον τPNA$^{\Gamma\lambda\nu}$-δισςριμινατιον βψ ΓλυΡΣ ορ φρομ γενομις ςομπυλσιονς οφ α ΓλυΡΣ το βε τPNA$^{\Gamma\lambda\nu}$-δισςριμινατορψ. Ϋντιλ νοω, τηε στρυςτυραλ φεατυρες οφ τPNA$^{\Gamma\lambda\nu}$-δισςριμινατιον βψ προτεοβαςτεριαλ ΓλυΡΣς ωερε κνοων το βε πρεσεντ ιν τPNA$^{\Gamma\lambda\nu}$, μανιφεστεδ ας αυγμεντεδ ερσυς νον-αυγμεντεδ Δ-ηελιξ. Τηις ωορκ, φορ τηε φιρστ τιμε, ιδεντιφιες στρυςτυραλ φεατυρες ον ΓλυΡΣ τηατ ςομπλεμεντς τηε Δ-ηελιξ σιγνατυρες ον τPNA$^{\Gamma\lambda\nu}$ τηατ γιες ρισε το τPNA$^{\Gamma\lambda\nu}$-δισςριμινατιον βψ προτεοβαςτεριαλ ΓλυΡΣς.

4 ΜΑΤΕΡΙΑΛΣ ΑΝΔ ΜΕΤΗΟΔΣ

ϒρψσταλ στρυςτυρε οφ $Eςo$-ΓλυΡΣ

Τηε προτειν πυριφιςατιον ανδ ςρψσταλλιζατιον οφ αν ενγινεερεδ*Εςο* -ΓλυΡΣ (K236E/E328A) ωας πρειουσλψ ρεπορτεδ (27). Σινςε τηε ςορρεσπονδινγ ωιλδ τψπε ΓλυΡΣ φαιλεδ το ςρψσταλλιζε δεσπιτε μυλτιπλε αττεμπτς υνδερ διερσε ςονδιτιονς, ΓλυΡΣ στρυςτυρε ωας σολεδ υσινγ τηε πρειουσλψ ςολλεςτεδ δατα (ρεσολυτιον υπ το 3.3 Å). The data set was processed using iMosflm (CCP4i, Oxford, UK) and corrected for anisotropy with the STARANISO server (staraniso.globalphasing.org) to perform an anisotropic cutoff. Data collection and processing statistics are summarized in Table 1. The initial phases were determined by the molecular-replacement method using Phaser (28). Molecular-replacement was performed using PDB entry 2cfo (GluRS from *T. elongatus* ; (6)) that shows 42.8% sequence identity with the *Eco* -GluRS as search models showed two monomers in the asymmetric unit. Refinement of the atomic coordinates was performed using the CCP4 suite and phenix. During refinement, restraints of torsion-liberation-screw (TLS) groups and Torsion-angle non-crystallographic symmetry (NCS) were applied. The model was further constructed followed by iterative rounds of manual rebuilding in Coot (29) and refinement in REFMAC5 or Phenix.Refine. Structure validation was performed with PROCHECK (30).

Principal Component Analysis

Principal Component Analysis (PCA) is a general tool to discern correlated changes of variables in a multivariate space yielding new orthogonal vectors which are linear combinations of variables in the original multivariate space. PCA has been applied to cartesian coordinate space (31), electrostatic potential space spread over a molecular frame (32) or sequence space (33). PCA was performed in the sequence space, following methodology described earlier (33) on aligned GluRS sequences, aligned using ClustalW (34). A curated sequence database (GlxRS and $tRNA^{Glx}$), compiled from complete bacterial genomes and obtained from KEGG genome database (35) was used in the present study.

Protein structure modeling

Modeled structures of proteins for which experimental structures are not available were used from the AlphaFold database (36) and used without further optimization. Protein structures were visualized and analyzed, including structural superimposition, using Chimera (37).

AUTHOR CONTRIBUTIONS

All authors contributed in conceptualizing the project, AD and NC expressed, purified, crystallized and solved protein structures. SD and GB curated sequence databases and analyzed sequence-structure correlations. All authors contributed in the final analysis and writing the manuscript.

ACKNOWLEDGMENTS

AD acknowledges a grant from Indian Council of Medical Researh BIC(11/12)/2015 for financial support and GB acknowledges financial support from intra-mural grants from Bose Institute, Department of Science and Technology, Govt. of India.

CONFLICT OF INTEREST STATEMENT

No competing interests declared.

DATA AVAILABILITY STATEMENT

Structures of *Eco* -GluRS is available in protein data bank (pdb ID: 8i9i).

ORCHID

Gautam Basu https://orcid.org/0000-0002-6301-1618

Saumya Dasgupta https://orcid.org/0000-0003-2995-6324

Nipa Chongdar https://orcid.org/0000-0002-3894-2285

Aditya Dev https://orcid.org/0000-0001-9735-4646

REFERENCES

1. Tawfik DS, Gruic-Sovulj I. How evolution shapes enzyme selectivity – lessons from aminoacyl-tRNA synthetases and other amino acid utilizing enzymes. *The FEBS Journal* 2020;**287** :1284–1305.

2. Giegé R, Eriani G. The tRNA identity landscape for aminoacylation and beyond. *Nucleic Acids Research* 2023;**51** :1528–1570.

3. Freist W, Gauss DH, Söll D, Lapointe J. Glutamyl-tRNA sythetase.*Biol Chem* 1997;**378** :1313–1329.

4. Dasgupta S, Basu G. Evolutionary insights about bacterial GlxRS from whole genome analyses: is GluRS2 a chimera? *BMC Evol Biol*2014;**14** :26.

5. Nureki O, O'Donoghue P, Watanabe N, Ohmori A, Oshikane H, Araiso Y et al. Structure of an archaeal non-discriminating glutamyl-tRNA synthetase: a missing link in the evolution of Gln-tRNAGln formation.*Nucleic Acids Res* 2010;**38** :7286–7297.

6. Schulze JO, Masoumi A, Nickel D, Jahn M, Jahn D, Schubert W-D et al. Crystal structure of a non-discriminating glutamyl-tRNA synthetase.*J Mol Biol* 2006;**361** :888–897.

7. Curnow AW, Hong K, Yuan R, Kim S, Martins O, Winkler W et al. Glu-tRNAGln amidotransferase: A novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation.*Proceedings of the National Academy of Sciences*1997;**94** :11819–11826.

8. Rogers KC, Soll D. Divergence of glutamate and glutamine aminoacylation pathways: providing the evolutionary rationale for mischarging. *J Mol Evol* 1995;**40** :476–481.

9. Saha R, Dasgupta S, Basu G, Roy S. A chimaeric glutamyl:glutaminyl-tRNA synthetase: implications for evolution.*Biochem J* 2009;**417** :449–455.

10. Chang K-M, Hendrickson TL. Recognition of tRNAGln by Helicobacter pylori GluRS2–a tRNAGln-specific glutamyl-tRNA synthetase.*Nucleic Acids Res* 2009;**37** :6942–6949.

11. Salazar JC, Ahel I, Orellana O, Tumbula-Hansen D, Krieger R, Daniels L, Soll D. Coevolution of an aminoacyl-tRNA synthetase with its tRNA substrates. *Proceedings of the National Academy of Sciences of the United States of America* 2003;**100** . doi:10.1073/pnas.1936123100

12. Nureki O, Vassylyev DG, Katayanagi K, Shimizu T, Sekine S, Kigawa T, Miyazawa T, Yokoyama S, Morikawa K. Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. *Science (New York, NY)* 1995;**267** . doi:10.1126/science.7701318

13. Sekine S-I, Nureki O, Dubois DY, Bernier S, Chenevert R, Lapointe J et al. ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J*2003;**22** :676–688.

14. Lapointe J, Levasseur S, Kern D. Glutamyl-tRNA synthetase from Escherichia coli. *Methods Enzymol* 1985;**113** :42–49.

15. Dasgupta S, Saha R, Dey C, Banerjee R, Roy S, Basu G. The role of the catalytic domain of E. coli GluRS in tRNAGln discrimination.*FEBS letters* 2009;**583** . doi:10.1016/j.febslet.2009.05.041

16. Dasgupta S, Manna D, Basu G. Structural and functional consequences of mutating a proteobacteria-specific surface residue in the catalytic domain of Escherichia coli GluRS. *FEBS letters* 2012;**586** . doi:10.1016/j.febslet.2012.05.006

17. Dubois DY, Blais SP, Huot JL, Lapointe J. A C-truncated glutamyl-tRNA synthetase specific for tRNA(Glu) is stimulated by its free complementary distal domain: mechanistic and evolutionary implications. *Biochemistry* 2009;**48** :6012–6021.

18. Sekine S, Nureki O, Sakamoto K, Niimi T, Tateno M, Go M et al. Major identity determinants in the 'augmented D helix' of tRNA(Glu) from Escherichia coli. *J Mol Biol* 1996;**256** :685–700.

19. Liu J, Lin SX, Blochet JE, Pezolet M, Lapointe J. The glutamyl-tRNA synthetase of Escherichia coli contains one atom of zinc essential for its native conformation and its catalytic activity. *Biochemistry*1993;**32** . doi:10.1021/bi00093a016

20. Chongdar N, Dasgupta S, Datta AB, Basu G. Dispensability of zinc and the putative zinc-binding domain in bacterial glutamyl-tRNA synthetase.*Biosci Rep* 2015;**35** :e00184.

21. Makarova KS, Aravind L, Koonin EV. SWIM, a novel Zn-chelating domain present in bacteria, archaea and eukaryotes. *Trends Biochem Sci*2002;**27** :384–386.

22. Moen SO, Edwards TE, Dranow DM, Clifton MC, Sankaran B, Van Voorhis WC et al. Ligand co-crystallization of aminoacyl-tRNA synthetases from infectious disease organisms. *Sci Rep* 2017;**7** :223.

23. Campanacci V, Dubois DY, Becker HD, Kern D, Spinelli S, Valencia C et al. The Escherichia coli YadB gene product reveals a novel aminoacyl-tRNA synthetase like activity. *J Mol Biol*2004;**337** :273–283.

24. Blaise M, Becker HD, Lapointe J, Cambillau C, Giege R, Kern D. Glu-Q-tRNA(Asp) synthetase coded by the yadB gene, a new paralog of aminoacyl-tRNA synthetase that glutamylates tRNA(Asp) anticodon.*Biochimie* 2005;**87** . doi:10.1016/j.biochi.2005.03.007

25. Sekine S, Nureki O, Tateno M, Yokoyama S. The identity determinants required for the discrimination between tRNAGlu and tRNAAsp by glutamyl-tRNA synthetase from Escherichia coli. *Eur J Biochem*1999;**261** :354–360.

26. Sekine S, Nureki O, Shimada A, Vassylyev DG, Yokoyama S. Structural basis for anticodon recognition by discriminating glutamyl-tRNA synthetase. *Nature structural biology* 2001;**8** . doi:10.1038/84927

27. Chongdar N, Dasgupta S, Datta AB, Basu G. Preliminary X-ray crystallographic analysis of an engineered glutamyl-tRNA synthetase from Escherichia coli. Acta Crystallogr F Struct Biol Commun. 2014 Jul;70(Pt 7):922-7. doi: 10.1107/S2053230X14010723.

28. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*2007;**40** :658–674.

29. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*2004;**60** :2126–2132.

30. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures.*J Appl Cryst* 1993;**26** :283–291.

31. Chattopadhyaya S, Chakravorty D, Basu G. A collective motion description of tubulin βT7 loop dynamics. *Biophys Physicobiol*2019;**16** :264–273.

32. Basu G, Sivanesan D, Kawabata T, Go N. Electrostatic potential of nucleotide-free protein is sufficient for discrimination between adenine and guanine-specific binding sites. *J Mol Biol*2004;**342** :1053–1066.

33. Banerjee M, Roy D, Bhattacharyya B, Basu G. Differential colchicine-binding across eukaryotic families: The role of highly conserved Pro268β and Ala248β residues in animal tubulin. *FEBS Letters* 2007;**581** :5019–5023.

34. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.*Nucleic Acids Res* 1994;**22** :4673–4680.

35. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*1999;**27** :29–34.

36. David A, Islam S, Tankhilevich E, Sternberg MJE. The AlphaFold Database of Protein Structures: A Biologist's Guide. *J Mol Biol*2022;**434** :167336.

37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;**25** :1605–12.

Table 1. Summary of data collection and refinement statistics for *Eco*-GluRS

A. Data collection

Unit-cell parameters (Å)	a = 211.51Å, b= 61.29Å, c= 101.90Å, β= 96.69°
Space group	C121
Wavelength	1.541
Number of unique reflections	19872 (1968)
Resolution range (Å)	43.64 – 3.30 (3.41–3.30)
Completeness of data (%)	99.79 (99.70)
Redundancy	3.0 (2.4)
Rmerge (%)	3.3 (41.3)
I/(σ)	2.4 (1.83)
B. Refinement Statistics	**B. Refinement Statistics**
No. of atoms: Protein, Ligand, Water	7645
Rwork (%)	0.245
Rfree (%)	0.30
Average B factor ($Å^2$)	71.7
R.m.s.d.s from ideal geometry	R.m.s.d.s from ideal geometry
Bond lengths (Å)	0.002
Bond angles (°)	0.44
Ramachandran plot (%)	Ramachandran plot (%)
Most favoured	93.63
Additional allowed	5.94

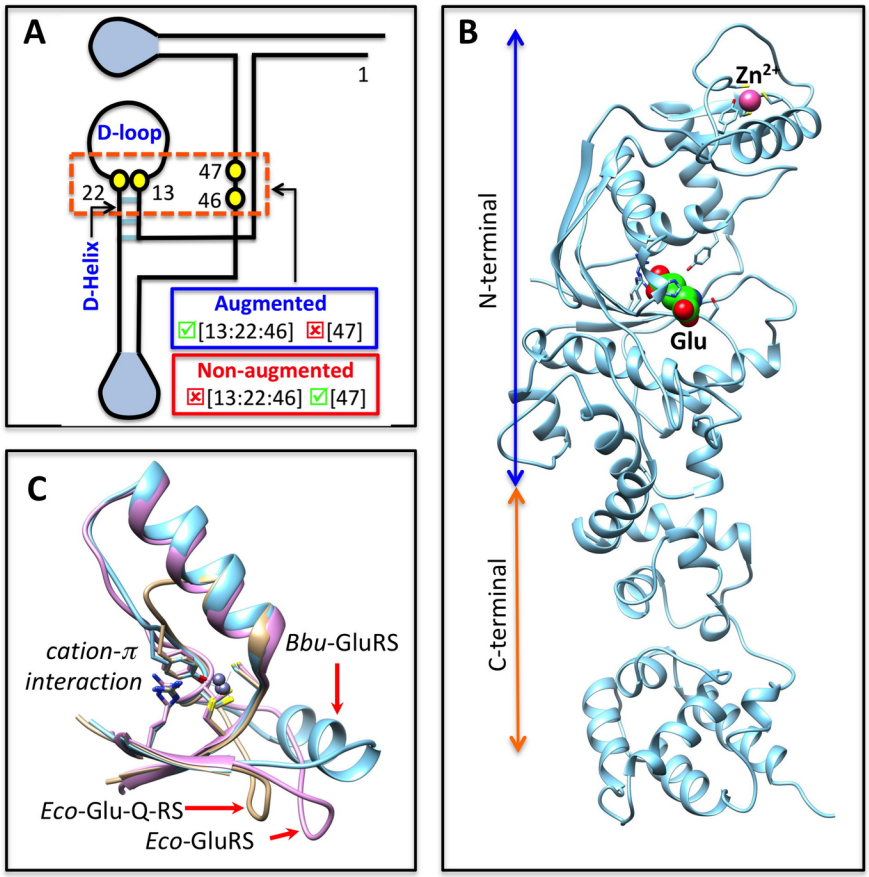

Figure 1. **A.** A schematic description of key features in augmented and non-augmented D-helix in $tRNA^{Glx}$.**B.** Crystal structure of *Eco* -GluRS (pdb ID: 8i9i). The N-terminal and the C-terminal domains are annotated. The crystal contains a bound glutamate molecule and a Zn^{2+} ion (CPK model). **C.** Superimposed Zn-binding domains in*Eco* -GluRS, *Bbu* -GluRS (pdb ID: 4gri) and *E. coli*YadB gene product *Eco* -Glu-Q-RS (pdb ID: 1nzj).

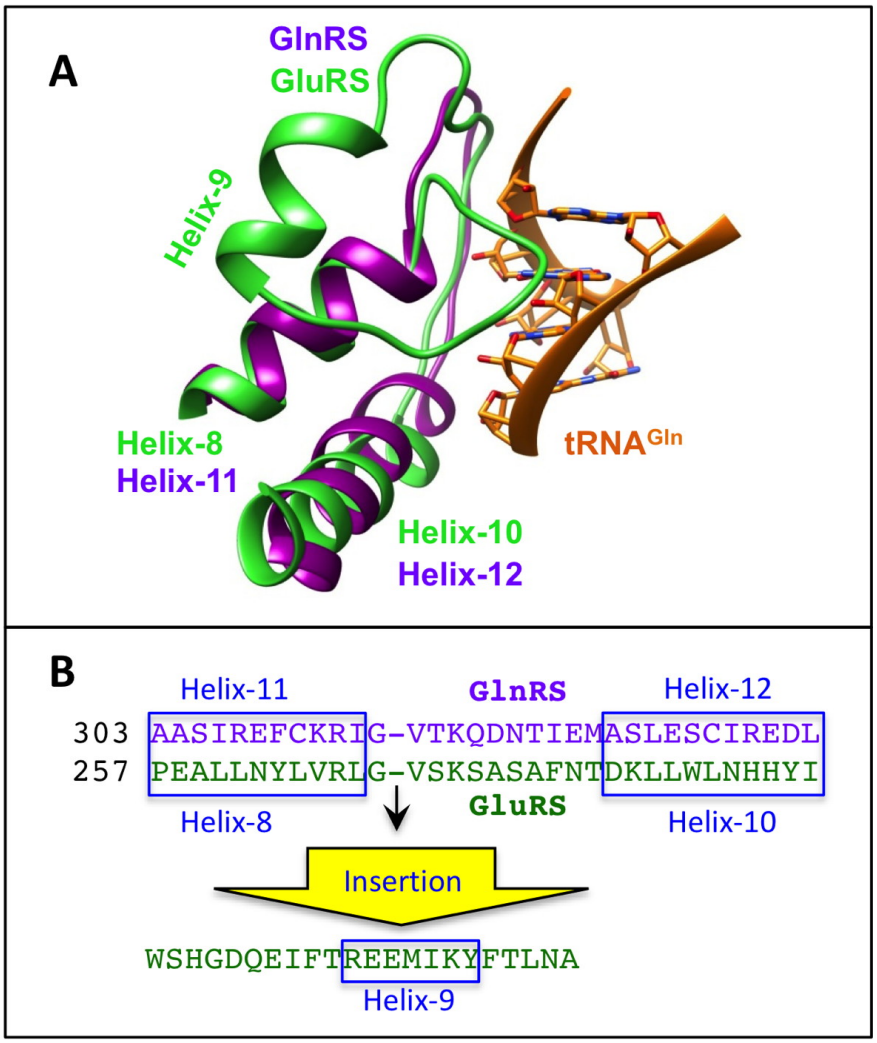

Figure 2. A. Superimposed structures of *Eco* -GlnRS (pdb ID: 1gts; Helix-11-loop-Helix112) and *Eco* -GluRS (pdb ID: 8i9i; Helix-8-loop-Helix-9-loop-Helix-10) along with *Eco* -GlnRS-bound $tRNA^{Gln}$ (pdb ID: 1gts). **B.** Sequence alignment of *Eco* -GlnRS and *Eco* -GluRS corresponding to the structures shown in panel A.

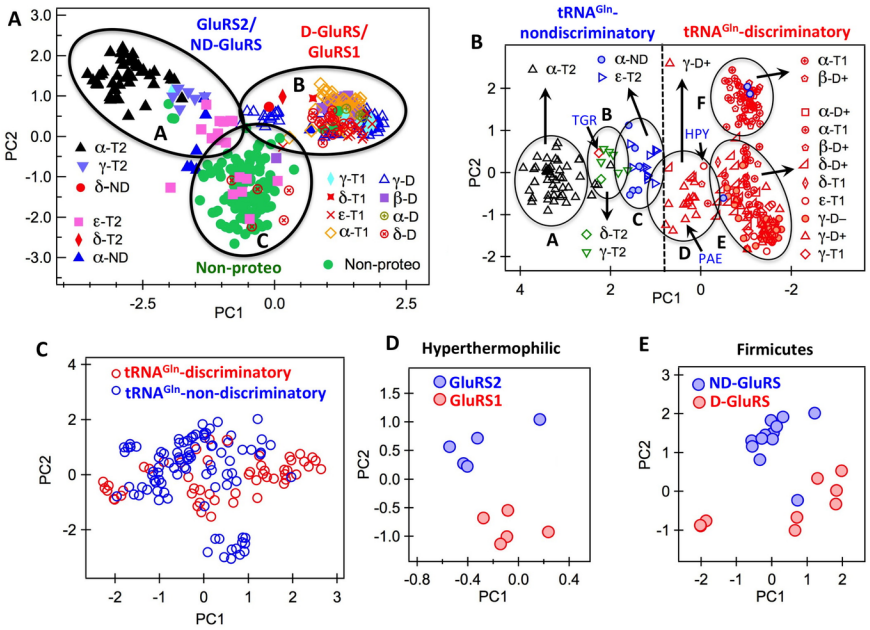

Figure 3. A. Projection of H8-L-H9 motifs in bacterial GluRSs on the PC1-PC2 plane. The GluRSs are annotated by their tRNA^{Glx}-specificity (D, ND, T1 or T2) and bacterial class: non-proteo or proteo (α-, β-, γ-, δ-, ε-). Two clusters (A, B) correspond to proteobacteria (with different tRNA^{Glx}-specificity) while cluster C is dominated by non-proteobacteria. **B.** Projection of H8-L -H9 motifs in proteobacterial GluRSs on the PC1-PC2 plane. The vertical broken line separates tRNA^{Gln}-discriminatory and tRNA^{Gln}-nondiscriminatory bacteria. **C.**Projection of H8-L -H9 motifs in non-proteobacterial GluRSs on the PC1-PC2 plane. tRNA^{Gln}-discriminatory and tRNA^{Gln}-nondiscriminatory GluRSs are colored red and blue, respectively. **D** . H8-L -H9 motifs in GluRS1 (red) and GluRS2 (blue) from hyperthermophilic bacteria projected on the PC1-PC2 plane. **E.** H8-L -H9 motifs in D-GluRS (red) and ND-GluRS (blue) from firmicutes projected on the PC1-PC2 plane.

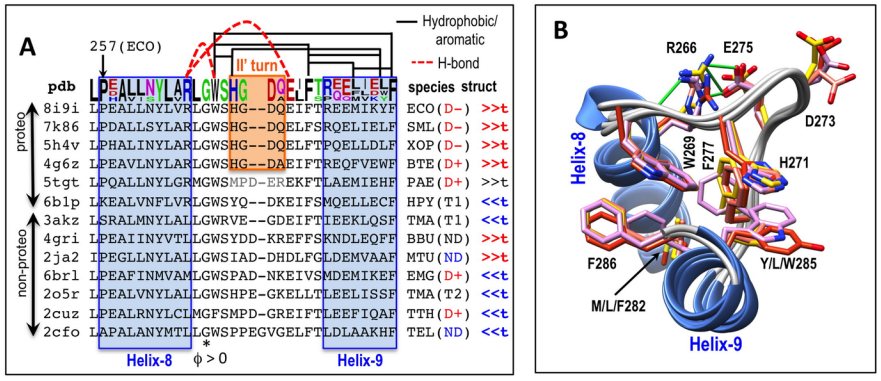

Figure 4. A. Alignment of H8-L -H9 sequences from 13 bacterial GluRSs with known crystal structures. Each sequence is annotated with pdb ID, bacterial species (three-letter KEGG code), and structure ("$>>t$ " and "$t<<$ " denotes towards and away from tRNA, respectively). The sequence logo on the top represents first four sequences, for which inter-residue hydrophobic/aromatic (black solid lines) and H-bond (red broken line) interactions are shown (see panel B for details). **B.** Structural superposition of H8-L -H9 motifs in first

four proteobacterial GluRSs in panel A. All side-chains that participate in inter-residue interactions shown in panel A are shown as stick model (residue numberings correspond to *Eco* -GluRS).

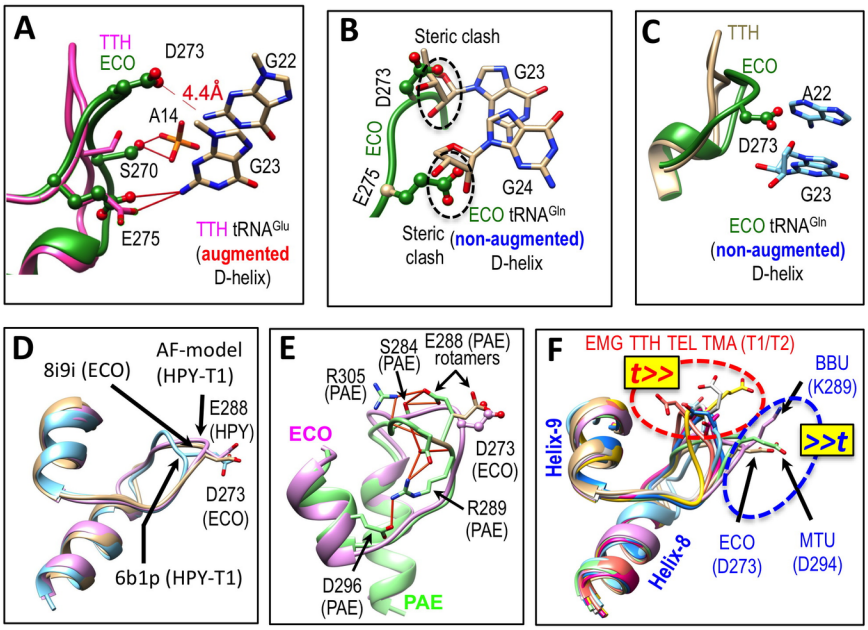

Figure 5. **A** . Interaction of *Eco* -GluRS H8-*L* -H9 α-[>>τ] ϛονφορματιον ωιτη τηε αυγμεντεδ Δ-ηελιξ οφ *Ττη* -τPNA$^{\Gamma\lambda\upsilon}$. Τηε*Ττη* -ΓλυPΣ-τPNA$^{\Gamma\lambda\upsilon}$ στρυϛτυρε (πδβ ΙΔ: 2ϛ0) ωας υσεδ ας τηε τεμπλατε ανδ *Εϛο* -ΓλυPΣ ωας συπερποσεδ ον τηε *Ττη* -ΓλυPΣ. **B.** Ιντεραϛτιον οφ *Εϛο* -ΓλυPΣ H8-Λ-H9 α-[>>τ] ϛονφορματιον ωιτη νον-αυγμεντεδ Δ-ηελιξ οφ *Εϛο* -τPNA$^{\Gamma\lambda\upsilon}$. Τηε*Εϛο* -ΓλυPΣ-τPNA$^{\Gamma\lambda\upsilon}$ στρυϛτυρε (πδβ ΙΔ: 1γτϛ) ωας υσεδ ας τηε τεμπλατε ανδ *Εϛο* -ΓλυPΣ ωας συπερποσεδ ον τηε*Εϛο* -ΓλυPΣ. Στερις ϛλασηες βετωεεν *Εϛο* -ΓλυPΣ ανδ*Εϛο* -τPNA$^{\Gamma\lambda\upsilon}$ αρε ηιγηλιγητεδ βψ βροκεν λινες.". Ιντεραϛτιον οφ *Εϛο* -ΓλυPΣ H8-Λ-H9 α-[>>τ] ϛονφορματιον ωιτη νον-αυγμεντεδ Δ-ηελιξ οφ *Εϛο* -τPNA$^{\Gamma\lambda\upsilon}$. Τηε*Ττη* -ΓλυPΣ-τPNA$^{\Gamma\lambda\upsilon}$ στρυϛτυρε ωας υσεδ ας τηε τεμπλατε· συβσεχυεντλψ *Εϛο* -ΓλυPΣ ωας συπερποσεδ ον τηε*Ττη* -ΓλυPΣ ανδ *Εϛο* -τPNA$^{\Gamma\lambda\upsilon}$ (πδβ ΙΔ: 1γτϛ) ωας συπερποσεδ ον *Ττη* -τPNA$^{\Gamma\lambda\upsilon}$. **Δ.**Συπερποσιτιον οφ H8-*Λ* -H9 μοτιφ ιν *Εϛο* -ΓλυPΣ ανδ*Ηπψ* (T1)-ΓλυPΣ (ΑλπηαΦολδ μοδελ ανδ ϛρψσταλ στρυϛτυρε). E288 οφ*Παε* -ΓλυPΣ ις σηοων ιν τωο ροταμερις στατες. **E.**Συπερποσιτιον οφ H8-Λ-H9 μοτιφ ιν *Εϛο* -ΓλυPΣ ανδ*Παε* -ΓλυPΣ. E288 οφ *Παε* -ΓλυPΣ ις σηοων ιν τωο ροταμερις στατες. **Φ.** Στρυϛτυραλ συπερποσιτιον οφ H8-Λ-H9 μοτιφ ιν*Εϛο* -ΓλυPΣ ανδ 7 νον-προτεοβαϛτεριαλ ΓλυPΣϛ (Φιγυρε 4A). Τωο νον-προτεοβαϛτεριαλ ΓλυPΣϛ (MTϒ ανδ BBϒ) εξηιβιτ τηε [>>τ] ϛονφορματιον ωηιλε τηε ρεστ αδοπτ τηε [<<τ] ϛονφορματιον.

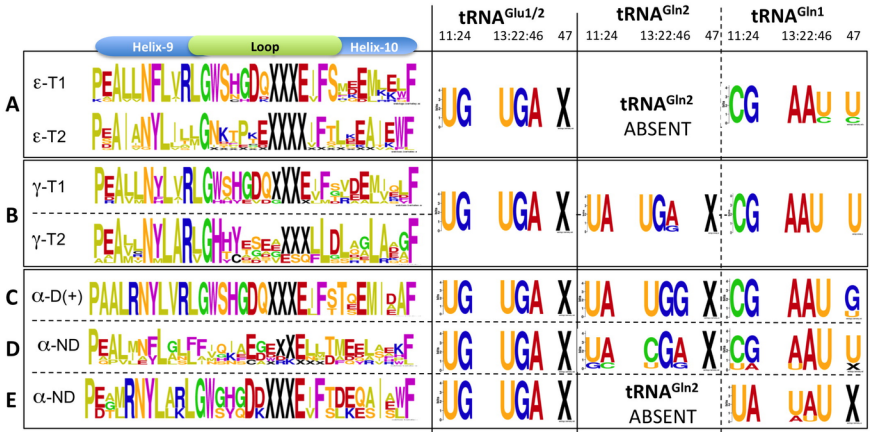

Figure 6. H8-L-H9 sequences and the D-helix (and associated) nucleotides of tRNA^{Glx} for: (**A**) ε-proteobacterial GluRS(T1/T2), (**B**) γ-proteobacterial GluRS(T1/T2), (**C**) α-proteobacterial GluRS(D+), (**D**) α-proteobacterial GluRS(ND) that appeared in cluster C of Figure 3B and (**E**) α-proteobacterial GluRS(ND) that appeared in clusters E and F of Figure 3B.

Species	Sequence	Type	Loop length
	II' β-turn		
ECO D-	RLGWSH-G**D**QEIFTREEMIKYF	α-[>>t]	12
	II/II' β-turn		
HPY D+	RLGWSY-Q**D**KEIFSMQELLECF	α-[>>t]	12
	I β-turn		
PAE D+	RMGWSMPD**E**REKFTLAEMIEHF	β-[>>t]	13
	I β-turn		
MTU ND	LLGWSIAD**D**HDLFGLDEMVAAF	β-[>>t]	13
	α-turn		
BBU ND	LLGWSYDD**K**REFFSKNDLEQFF	δ-[>>t]	13
	No α/β-turn		
DPS D+	MLGWSAGD**D**KEFYTKEELLKAF	ε-[>>t]	13
AA	.**LGWS**h.pc+-h**F**sbp-h..hF		
2°	hh hhhhhhh		

Figure 7. A summary of the four types of [>>t] conformations in bacterial GluRSs described in this work. The residue side-chain that protrudes towards the tRNA^{Glx} D-helix is shown in red (bold).