

# Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases<sup>\*S</sup>♦

Received for publication, March 31, 2016, and in revised form, May 9, 2016. Published, JBC Papers in Press, May 23, 2016, DOI 10.1074/jbc.M116.730382

Marco Igor Valencia-Sánchez<sup>†1</sup>, Annia Rodríguez-Hernández<sup>‡S2</sup>, Ruben Ferreira<sup>¶1</sup>, Hugo Aníbal Santamaría-Suárez<sup>‡</sup>, Marcelino Arciniega<sup>‡</sup>, Anne-Catherine Dock-Bregeon<sup>||</sup>, Dino Moras<sup>\*\*</sup>, Brice Beinstener<sup>\*\*</sup>,  Haydyn Mertens<sup>‡‡</sup>,  Dmitri Svergun<sup>‡‡</sup>, Luis G. Briebe<sup>§</sup>, Morten Grøtli<sup>¶</sup>, and  Alfredo Torres-Larios<sup>‡3</sup>

From the <sup>†</sup>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, the <sup>§</sup>Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, Guanajuato 04510, México, the <sup>¶</sup>Department of Chemistry and Molecular Biology, University of Gothenburg, SE-412 96 Gothenburg, Sweden, the <sup>||</sup>Laboratoire de Biologie Intégrative des Milieux Marins, CNRS UMR8227, 29680 Roscoff, France, the <sup>\*\*</sup>Centre for Integrative Biology, Department of Integrated Structural Biology, Institute of Genetics and of Molecular and Cellular Biology, CNRS UMR 7104, 1 Rue Laurent Fries, Illkirch, France, and the <sup>‡‡</sup>European Molecular Biology Laboratory, Hamburg Outstation, c/o DESY, Notkestrasse 85, Hamburg 22603, Germany

Glycyl tRNA synthetase (GlyRS) provides a unique case among class II aminoacyl tRNA synthetases, with two clearly widespread types of enzymes: a dimeric ( $\alpha_2$ ) species present in some bacteria, archaea, and eukaryotes; and a heterotetrameric form ( $\alpha_2\beta_2$ ) present in most bacteria. Although the differences between both types of GlyRS at the anticodon binding domain level are evident, the extent and implications of the variations in the catalytic domain have not been described, and it is unclear whether the mechanism of amino acid recognition is also dissimilar. Here, we show that the  $\alpha$ -subunit of the  $\alpha_2\beta_2$  GlyRS from the bacterium *Aquifex aeolicus* is able to perform the first step of the aminoacylation reaction, which involves the activation of the amino acid with ATP. The crystal structure of the  $\alpha$ -subunit in the complex with an analog of glycyl adenylate at 2.8 Å resolution presents a conformational arrangement that properly positions the cognate amino acid. This work shows that glycine is recognized by a subset of different residues in the two types of GlyRS. A structural and sequence analysis of class II catalytic domains shows that bacterial GlyRS is closely related to alanyl tRNA synthetase, which led us to define a new subclassification of these ancient enzymes and to propose an evolution-

ary path of  $\alpha_2\beta_2$  GlyRS, convergent with  $\alpha_2$  GlyRS and divergent from AlaRS, thus providing a possible explanation for the puzzling existence of two proteins sharing the same fold and function but not a common ancestor.

Aminoacyl tRNA synthetases are ancient enzymes that attach cognate amino acids to their corresponding tRNAs (1–3). This task is performed in two steps as follows: amino acid activation with ATP, followed by attachment of the resulting aminoacyl adenylate to the cognate tRNA (4–6). According to their sequence and structural features, there are two main classes of non-related aaRS,<sup>4</sup> which evolved independently from two different domains corresponding to two modes of ATP binding (7–10). These classes are further divided into subclasses according to protein sequence, structural features of the catalytic domain, the presence of accessory domains, and similarity of amino acids (11–13). Most aaRSs are descended from a single ancestor (monophyletic); however, there are two clearly recognized exceptions, lysyl tRNA synthetase and GlyRS (2, 12, 14–16). GlyRS, a class II aaRS, can be found as an  $\alpha_2$  homodimer in eukarya, archaea, and some bacteria (14, 17–20) and as an  $\alpha_2\beta_2$  heterotetramer in most bacteria and chloroplasts (21–23). Notably, the limited extent of the sequence conservation between the two forms does not allow inference to a common ancestor between them. Nonetheless, the catalytic domains contained in the  $\alpha$ -subunits of both GlyRSs share the same fold (11, 14, 16, 22–24). The eukaryotic GlyRS belongs to the subclass IIa, specific for hydrophobic and small polar amino acids, with similarities with HisRS, ThrRS, ProRS, and SerRS.

\* The authors declare that they have no conflicts of interest with the contents of this article.

M. I. V. S. and A. T. L. dedicate this work to Prof. Enrique Merino Pérez, IBT-Universidad Nacional Autónoma de México, for his constant support and guidance.

✂ Author's Choice—Final version free via Creative Commons CC-BY license.

♦ This article was selected as a Paper of the Week.

 This article contains supplemental Figs. S1 and S2.

The atomic coordinates and structure factors (code 5F5W) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

<sup>1</sup> Recipient of a grant from CONACyT to perform graduate studies. Part of this work in was performed to fulfill the requirements of a doctoral dissertation in the Posgrado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México.

<sup>2</sup> Supported by a CONACyT postdoctoral fellowship.

<sup>3</sup> Supported by Consejo Nacional de Ciencia y Tecnología (CONACyT) Grants CB2015-253323 and PDCPN2014-47543 and Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, Universidad Nacional Autónoma de México (PAPIIT-DGAPA) Grant IN202416. To whom correspondence should be addressed: Circuito Exterior s/n, Ciudad Universitaria, Apartado Postal 70-243, México D.F. 04510, México. Fax: 52-55-5622-5630; E-mail: torres@ifc.unam.mx.

<sup>4</sup> The abbreviations used are: aaRS, aminoacyl tRNA synthetase; Gly-SA, G5A, or GSAd, 5'-O-[N-(L-glycyl)sulfamoyl] adenosine; SAXS, small angle x-ray scattering; TCS, transitive consistency score; SEC-MALS, size exclusion chromatography coupled to multiangle light scattering; PCA, principal component analysis; r.m.s.d., root mean square deviation; PDB, Protein Data Bank; GlyRS, glycyl tRNA synthetase; HisRS, histidyl tRNA synthetase; ThrRS, threonyl tRNA synthetase; ProRS, prolyl tRNA synthetase; SerRS, seryl tRNA synthetase; PheRS, phenylalanyl tRNA synthetase; AlaRS, alanyl tRNA synthetase; SepRS, phosphoseryl tRNA synthetase; PylRS, pyrrolysyl tRNA synthetase.

Bacterial GlyRS, however, belongs to the subclass IIc, together with PheRS, AlaRS, SepRS, and PylRS, the most heterogeneous group of the three subclasses (25–27). However, there are some studies that place AlaRS or bacterial GlyRS within subclass IIa and/or do not consider the existence of two types of GlyRS belonging to different subclasses (28–32).

The ~35-kDa  $\alpha$ -subunit of bacterial GlyRS contains the aminoacylation site (22, 24, 33, 34). However, no activity has been demonstrated for this domain alone, suggesting that the  $\beta$ -subunit is required for catalysis (21, 33, 35–37). The ~65-kDa  $\beta$ -subunit has the most important tRNA recognition elements but is not homologous to the anticodon binding domain of eukaryotic GlyRS (23, 37). Two crystal structures are available for the bacterial GlyRS  $\alpha$ -subunit in the apo-conformation (PDB entries 3rf1 and 1j5w), leaving unclear whether the recognition details of the small substrates, glycine and ATP, are shared by the two types of enzymes.

Here, we show that the bacterial GlyRS  $\alpha$ -subunit, which has all of the molecular determinants needed for the first step of the reaction, is indeed able to perform this catalysis. The crystal structure of this subunit in complex with an analog of glycyl adenylate (5'-O-[N-(L-glycyl)sulfamoyl] adenosine, also known as Gly-SA, G5A, or GSAd) allows the conformational changes correlated with glycine recognition to be defined. A comparison of the  $\alpha$ -subunit of bacterial GlyRS with the activation domain of archaeal and eukaryotic GlyRS establishes that the two classes of GlyRS employ different chemical strategies to recognize glycine as a substrate. Moreover, a structural and sequence analysis performed on the activation domain of class II aaRS defines a new subclass II<sub>d</sub>, comprising AlaRS and bacterial  $\alpha_2\beta_2$  GlyRS. This consequently modifies subclass IIc to contain ( $\alpha\beta$ )<sub>2</sub> PheRS, SepRS, and possibly PylRS. This classification standing on common structural motifs and active site residues in class II<sub>d</sub> enzymes allows us to propose a divergence of bacterial GlyRS from AlaRS, which resolves the puzzling existence of two proteins sharing the same fold and function but not a common ancestor.

## Results

**$\alpha$ -Subunit of Bacterial GlyRS Is Able to Perform the Amino Acid Activation**—We were able to obtain highly purified  $\alpha$ -subunit GlyRS of the hyperthermophilic bacterium *Aquifex aeolicus* ( $\alpha$ -AaGlyRS) by means of a heat treatment and an astringent His tag affinity chromatography step. A final purification step using size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) indicated a homogeneous dimeric population of 69.1 kDa (theoretical mass = 67.4 kDa), in agreement with previous reports (21, 36). Small angle x-ray scattering (SAXS) further confirmed the dimeric nature of the ensemble. Because of the evolutionary conservation of all amino acids involved in glycine activation (see below), we speculated that the  $\alpha$ -subunit alone would be able to catalyze the first step of the reaction, the attachment of glycine to ATP. With the use of an alternative method based on thin layer chromatography to monitor the activity, we found that  $\alpha$ -AaGlyRS was indeed able to perform the first step of aminoacylation (Fig. 1). In contradiction with previous reports, the  $\alpha$ -subunit showed weak activity at pH values ranging from 6.0 to 8.0 and

glycine concentrations from 80  $\mu$ M to 10 mM. Under the best possible reaction conditions, the observed  $K_m$  for glycine was  $0.11 \pm 0.016$  mM, similar to a previously reported value for the full-length *Escherichia coli* enzyme (21). Nevertheless, the  $k_{cat}$  reaction was extremely slow, only  $2.3 \times 10^{-4}$  s<sup>-1</sup>. Previous studies have performed an ATP-PP<sub>i</sub> exchange assay, which follows the reverse reaction of incorporation of <sup>32</sup>P from PP<sub>i</sub> into ATP (21, 33, 35–37), whereas in our study the Gly-AMP synthesis was measured directly. These findings provide biological relevance to our reported crystal structure (see below) and support the functional existence of an  $\alpha_2\beta_2$  quaternary structure, rather than an ( $\alpha\beta$ )<sub>2</sub> organization.

**Binding to a Transition State Analog Promotes Conformational Changes in the  $\alpha$ -Subunit of Bacterial GlyRS**—To understand amino acid and nucleotide recognition in bacterial GlyRS, we solved the crystal structure of  $\alpha$ -AaGlyRS in complex with GSAd at 2.81 Å resolution (Table 1). The electron density map unambiguously showed all features of the bound GSAd and its molecular surroundings in all five molecules in the asymmetric unit (Fig. 2B).

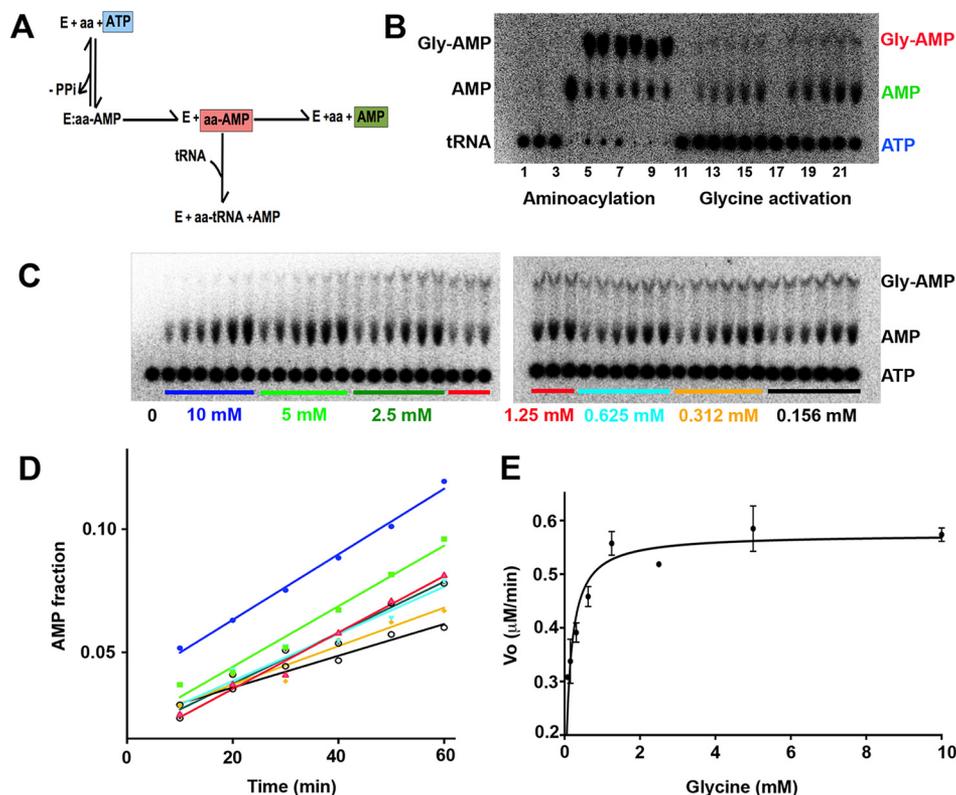
As shown previously (PDB code 1j5w (24)), the  $\alpha$ -subunit forms a homodimer (Fig. 2A), with 58% of its 2492 Å<sup>2</sup> interface formed by a 97-residue helical region located at the C-terminal part of the subunit and situated on top of the signature antiparallel  $\beta$ -strand of class II synthetases. Further details on the description of the general architecture of this subunit have been given previously (24).

Comparison of  $\alpha$ -AaGlyRS with previously solved structures of this subunit in the apo-form reveals conformational changes in the region formed by residues 112–123 (Fig. 2, C and D). This segment is the topological equivalent of the so-called amino acid loop found in HisRS, ProRS, ThrRS, SerRS, and eukaryotic GlyRS (38). Here, the movement of this region causes an ~5 Å displacement of Trp-115 when compared with the apo structure, allowing it to form a cation- $\pi$  interaction with the glycine moiety of GSAd (Fig. 2D, see below) and to form a hydrogen bond with the O1S atom of GSAd (equivalent to the O1P atom of the cognate Gly-AMP (Fig. 3D)). As a result, the size of the active site pocket is reduced from an area and volume of 706 Å<sup>2</sup> and 960 Å<sup>3</sup> to 451 Å<sup>2</sup> and 636 Å<sup>3</sup>.

In correlation with this movement, thermal shift assays showed a displacement of the  $T_m$  values of  $\alpha$ -AaGlyRS in the presence of GSAd, from 83.2 °C (no ligand) to 92.3 °C (8 mM GSAd). We could not detect significant changes in the presence of glycine, ATP, or ATP + glycine. The reported crystal structure of the  $\alpha$ -subunit of bacterial GlyRS shows a productive, biologically meaningful complex that allows the recognition mechanisms in the two types of GlyRS to be compared.

**Mechanism of Glycine Recognition Differs in the Two Types of Known GlyRS**—Although nucleotide recognition shows common features in both types of GlyRSs (Fig. 3), critical differences are found at the amino acid recognition level, which is centered on the amino group. In eukaryotic GlyRS, the amino group is recognized by three conserved and negatively charged glutamic acid residues (Glu-522, Glu-296, and Glu-245 in PDB code 2zt8) (Fig. 4, A and C) (39). In contrast, in the bacterial type GlyRS, the amino group is recognized through five conserved different residues (Trp-115, Gln-76, Gln-78, Thr-33, and Glu-

## Insights into the Polyphyletic Origins of GlyRSs



**FIGURE 1.  $\alpha$ -Subunit of the tetrameric  $\alpha_2\beta_2$  GlyRS from *A. aeolicus* ( $\alpha$ -AaGlyRS) is able to activate the amino acid.** *A*, schematic diagram of the first step of aminoacylation. *B*, control experiments. Comparison of full aminoacylation reaction versus amino acid activation. Lanes 1–10, aminoacylation reaction performed as described previously (57–59). Lane 1, no enzyme added, and no P1 nuclease added. Lane 2, *Anaerolinea thermophila*, a protein that bears in the same sequence both subunits ( $\alpha\beta$ -GlyRS) added, and no P1 nuclease added. Lane 3, *A. aeolicus* ( $\alpha + \beta$ -GlyRS) added, and no P1 nuclease added. Lane 4, no enzyme added, and P1 nuclease added. Lanes 5–7, 5, 10, and 15 min of aminoacylation reaction using  $\alpha + \beta$ -GlyRS (with P1 nuclease added). Lanes 8–10, 5, 10, and 15 min of aminoacylation reaction using  $\alpha\beta$ -GlyRS (with P1 nuclease added). Lanes 11–22, glycine activation reaction. Lane 11, zero time point using  $\alpha$ -AaGlyRS. Lanes 12–16, 10, 20, 30, 40, and 50 min of the glycine activation reaction using  $\alpha$ -AaGlyRS. Lane 17, zero time point using  $\alpha\beta$ -GlyRS. Lanes 18–22, 10, 20, 30, 40, and 50 min of the glycine activation reaction using the  $\alpha\beta$ -GlyRS. *C*, amino acid activation.  $\alpha$ -AaGlyRS at 40  $\mu$ M, in the presence of decreasing glycine concentrations, 0.5 mM ATP, 50 mM Tris, pH 8.0, 50 mM KCl, 10 mM MgCl<sub>2</sub>. Time points were taken every 10 min for 60 min for each concentration, and the formation of AMP was monitored for each point. *D*, initial velocities (kinetics of AMP formation from the experiment in *C*). Steady state time courses using different glycine concentrations. *E*, Michaelis-Menten plot. Initial velocities were plotted against substrate concentration; error bars indicate the standard deviation for each point.

**TABLE 1**  
Data collection and refinement statistics

One crystal was used.

| $\alpha$ -AaGlyRS-GSAd  |                                     |
|---|-------------------------------------|
| <b>Data collection</b>  |                                     |
| Space group   | P22 <sub>1</sub> 2 <sub>1</sub>     |
| Cell dimensions   |                                     |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                              | 101.8, 130.0, 145.5                 |
| $\alpha$ , $\beta$ , $\gamma$ (°)                               | 90.0, 90.0, 90.0                    |
| Resolution (Å)  | 83.43–2.81 (2.91–2.81) <sup>a</sup> |
| <i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>            | 0.084 (0.584)                       |
| <i>I</i> / $\sigma$ <i>I</i>                                    | 9.7 (2.2)                           |
| Completeness (%)  | 100 (100)                           |
| Redundancy  | 4.4 (4.5)                           |
| Mean ( <i>I</i> ) half-set correlation <i>CC</i> <sub>1/2</sub> | 0.990 (0.770)                       |
| Wilson <i>B</i> -factor (Å <sup>2</sup> )                       | 58.4                                |
| <b>Refinement</b>   |                                     |
| Resolution (Å)  | 80.19–2.81 (2.86–2.81)              |
| No. of reflections  | 47,741 (2607)                       |
| <i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>             | 0.239/0.252 (0.362/0.363)           |
| No. of atoms  |                                     |
| Protein   | 11,740                              |
| Ligand/ion  | 135                                 |
| Water   | 2                                   |
| <b><i>B</i>-Factors</b>   |                                     |
| Protein (by chain)  | 45.1, 53.6, 56.9, 69.6, 73.1        |
| Ligand/ion  | 40.8, 47.5, 52.6, 68.0, 74.6        |
| Water   | 45.8                                |
| <b>r.m.s.d.</b>   |                                     |
| Bond lengths (Å)  | 0.011                               |
| Bond angles (°)   | 1.462                               |

<sup>a</sup> Highest resolution shell is shown in parentheses.

156) (Figs. 2*D* and 4, *B*, *D*, and *E*). Interestingly, Gln-76 interacts directly with the glycine carbonyl group and also with the amino moiety through hydrogen bonding, with Gln-78, of a solvent molecule (Fig. 4*D*). The amino group of GSAd also interacts with the side chains of Thr-33 and Glu-156 (Fig. 4, *B* and *D*).

The differences with eukaryotic GlyRS are also extended to the region of the amino acid loop. In eukaryotic GlyRS, the hydroxyl group of a non-conserved Tyr-386 contacts the carbonyl group of GSAd (Fig. 4*E*). In contrast, in bacterial GlyRS, an absolutely conserved Trp-115 makes a clear cation- $\pi$  interaction with the amino moiety of GSAd and also contacts the O1S atom of GSAd (Figs. 2*D* and 4*E*).

In addition, in eukaryotic GlyRS, the binding of the noncognate alanine is prevented in part by a non-conserved Ser (or Ala) residue (524) and the highly negatively charged recognition cavity (Fig. 4, *C* and *F*). In contrast, in bacterial GlyRS, this steric hindrance is made by a pair of highly conserved Thr residues (158 and 140) together with a nearby solvent molecule (Fig. 4*F*).

The structural sequence alignment combined with multiple sequence alignments indicate seven highly conserved residues that are shared by the two types of GlyRS (Fig. 3*B*). Five of these

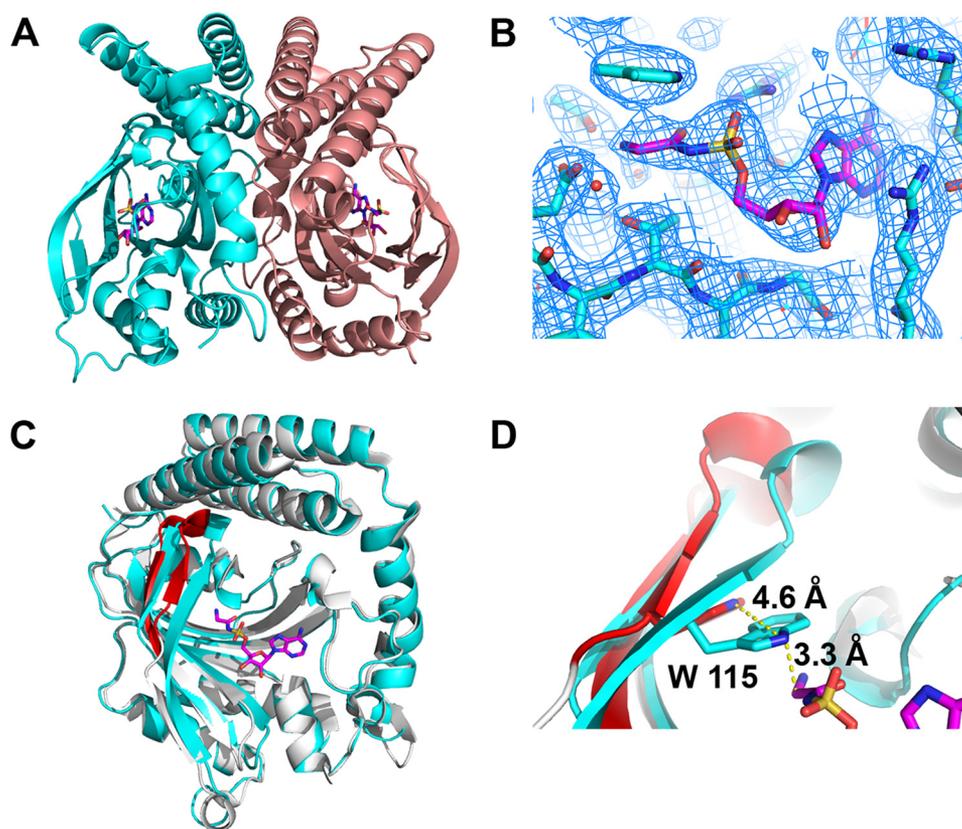


FIGURE 2. **Binding of a glycyl adenylate analog promotes a conformational change in  $\alpha$ -AaGlyRS.** *A*, overall dimeric structure of  $\alpha$ -AaGlyRS with GSAd shown in magenta. There are five monomers, each with a ligand bound, in the asymmetric unit. *B*, simulated annealed,  $2F_o - F_c$  omit electron density map contoured at  $1\sigma$  around the GSAd-binding site. Two solvent molecules were located near the glycine recognition site that could be well fitted and observed on double difference and omit maps for monomer B and could be seen as peaks bigger than  $3.5\sigma$  on difference,  $F_o - F_c$  electron density maps on the four other monomers. No other solvent molecules were added to the model. *C*, superposition of  $\alpha$ -AaGlyRS-GSAd with the apo structure of *Campylobacter jejuni* (PDB code 3rgl). The sequences of the subunits have an identity of 60% and a similarity of 77%. The overall r.m.s.d. is 0.959 Å. The binding of GSAd causes conformational changes in the region of residues 112–123, shown in red in *C* and *D*. *D*, close-up of the superposition showing the conformational change upon binding of GSAd, which involves a 4.6 Å movement of Trp-115 to make a cation- $\pi$  interaction with GSAd. The superposition with another  $\alpha$ -subunit structure from *T. maritima* (PDB code 1j5w) does not show this displacement, suggesting that this conformational change is due to GSAd binding.

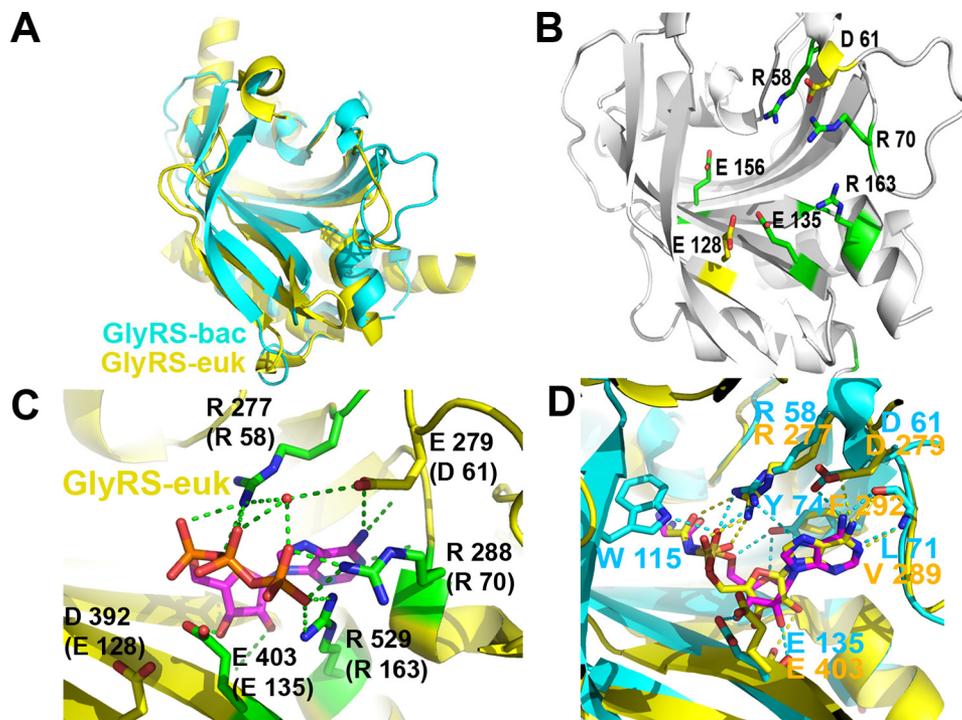
are involved in ATP recognition, and as such, these residues are highly conserved and shared by many aaRSs. There is only one residue that is shared by the two types of GlyRS for the recognition of glycine (Glu-156 in bacterial GlyRS and Glu-522 in eukaryotic GlyRS). Not another single residue is shared in any other region of the catalytic domain. Comparison of the electrostatic potentials within the glycine binding pockets shows a highly negatively charged region in eukaryotic GlyRS and a less polar environment in bacterial GlyRS. The comparison of the two types of GlyRSs and further analysis as shown below indicate that their differences are so profound, up to the level of amino acid recognition, that it is extremely unlikely for them to share a common ancestor, even when their catalytic domains have the same overall fold.

**Bacterial GlyRS Presents Key Differences Compared with Other Class II Synthetases**—A hallmark of class II synthetases is the presence of key residues located in each of their three characteristic motifs (Fig. 5A) (7). In bacterial GlyRS, the signature sequence of motif 2, formed by the first two strands of the antiparallel  $\beta$ -sheet that forms the floor of the active site, presents several changes compared with the consensus sequence (Fig. 5B). In particular, there are three key differences. 1) Unique to bacterial GlyRS, an arginine residue (Arg57) is located where a hydrophobic residue is otherwise always pres-

ent (Fig. 5, A–C). Additionally, an Arg residue makes a salt bridge with two highly conserved Asp residues (Fig. 5C). In the rest of class II synthetases, a Phe, Tyr, Val, or His residue helps to stabilize a hydrophobic region located at the dimeric interface (Fig. 5C). 2) Similar to AlaRS, there is an insertion (Pro-59 in bacterial GlyRS and two to six residues in AlaRS) between an absolutely conserved Arg and an acidic residue (Fig. 5, A and B). 3) In many aaRSs, a fully conserved acidic residue recognizes the amino moiety of the cognate amino acid (Fig. 5, A, B and D). In bacterial GlyRS, AlaRS, HisRS, SepRS, and PylRS, another residue occupies this position (Fig. 5B). In bacterial GlyRS, a glutamine residue is substituted in place of this acidic residue (Fig. 5, A, B and D). In  $\alpha$ -AaGlyRS, Gln-78 makes a hydrogen bond with a solvent molecule that interacts with the amino moiety of the cognate glycine (Fig. 5D). These sequence-structure observations confirm that bacterial GlyRS has non-canonical features that are somewhat shared with other atypical aaRSs, like AlaRS, and not with eukaryal GlyRS.

**Bacterial GlyRS and AlaRS Form a Structural Subclass of aaRS**—To gain further insight into the evolutionary origin of bacterial GlyRS, we performed pairwise structural alignments of the catalytic core (proposed to reflect the primordial synthetase (40)), among the available crystal structures of all types of class II aaRS using two different algorithms (Dali (41) and

## Insights into the Polyphyletic Origins of GlyRSs



**FIGURE 3. Features shared by the two types of GlyRSs.** *A*, superposition of the catalytic core, according to Dali, of bacterial GlyRS (residues 1–169) and eukaryotic GlyRS (residues 97–536, with the exception of extensions comprising residues 129–137, 145–242, and 301–349, PDB code 2zt8). The Z-score is 14.9 for the 152 residues aligned, with an r.m.s.d. of 2.7 and 17% identity. *B*, residues that are highly (yellow) or absolutely conserved (green) and also shared by bacterial GlyRS and eukaryotic GlyRS according to the structural and sequence alignment of all available sequences are shown on the structure of bacterial GlyRS. The following panels show the role and/or location of some of these residues in more detail. *C*, five of the seven residues depicted in *B* are highly conserved class II residues that recognize ATP (PDB code 2zt7). *D*, in comparison with eukaryotic GlyRS,  $\alpha$ -AaGlyRS Tyr-74 makes additional interactions through its OH group, which contacts the O4' and O5' atoms of GSAd and the NE1 atom of Trp-115 that contacts the O1S atom of GSAd.

STAMP (42)). We also performed a multiple sequence alignment using the structural information through T-Coffee Expresso (43). Finally, we performed a principal component analysis (PCA) on the C $\alpha$  atoms of 80 core residues common to the structure of class II aaRSs, according to the algorithm implemented in Bio3D (44). The information derived from each of the four approaches is summarized in Figs. 6–10, Tables 2 and 3, and supplemental Figs. S1 and S2. We were able to clearly identify four subclasses of class II aaRSs. These subclasses could be defined either using the whole catalytic domain just devoid of big insertions (Fig. 6, Table 2, and supplemental Figs. S1 and S2) or a small C $\alpha$  core of 80 atoms (Figs. 7–10). The analysis of the C $\alpha$  core of 80 atoms showed that the main differences between the groups are found in two helical regions and one  $\beta$ -strand (PC1 and PC2, residues 16–22, 46–58, and 61–64 of the C $\alpha$  core, Figs. 7 and 9A). Notably, the PCA analysis (Fig. 9) shows that only these two components, which together account for almost 50% of the structural variance of the core between the class II structures (Fig. 9B), allow us to define a clear separation between the subgroups, most importantly, the group formed by bacterial GlyRS and AlaRS and the rest of class II aaRSs (Figs. 9C and 10). The four subgroups are maintained even if bacterial GlyRS or AlaRS is removed from the PCA. At the whole catalytic domain level, the superposition of representatives of the four subclasses shows a general fold agreement (Fig. 6, average Dali Z-score of 13.4). A closer approach allows us to visualize the differences among the four established divisions (groups with Dali Z-scores equal to or higher than 18.5, which turned out to be equal to the branches

of the STAMP analysis based on r.m.s.d.) and the agreement between them (Fig. 6, Table 2, and supplemental Fig. S1). Each subclass is mostly defined by the relative angles of two helices located at the back of the active site (subclass IIa *versus* IIb), by the length of these helices (subclass IIa and IIb *versus* IIc and IId), or by the length and relative orientation of three strands of the active site  $\beta$ -sheet (subclass IIc *versus* IId). Notably, each subclass is in full agreement with previously proposed divisions. However, subclass IIc is now divided in two distinct subclasses: IIc, including PheRS, SepRS, and possibly PylRS, which groups with subclass IIc according to the structural classification (Figs. 8–10 and supplemental Fig. S1) but with subclass IIb according to the sequence analysis (supplemental Fig. S2). The other subclass, IId, includes AlaRS and bacterial GlyRS, which match not only in helical length but also the length and orientation of three strands of the active site  $\beta$ -sheet (Fig. 11).

**Bacterial GlyRS and AlaRS Share Key Residues Important for Both Protein Structure and Substrate Recognition**—The structural sequence alignment coupled with multiple sequence alignments (Fig. 11, A and B) indicate that there are 15 residues either absolutely or highly conserved by bacterial GlyRS and AlaRS. Five of these are involved in ATP recognition (Fig. 11, B and C). Outside of this, the amino acid recognition region surprisingly revealed that the equivalent of Trp-115 in bacterial GlyRS is also found in AlaRS (Trp-192, PDB code 2zzg) (Fig. 11, B, C and E) (45). This residue is absolutely conserved in AlaRSs from all domains of life and bacterial GlyRS and only in these synthetases. Notably, there is a conformational change in the region of Trp-192 in AlaRS in response to ligand binding (45),

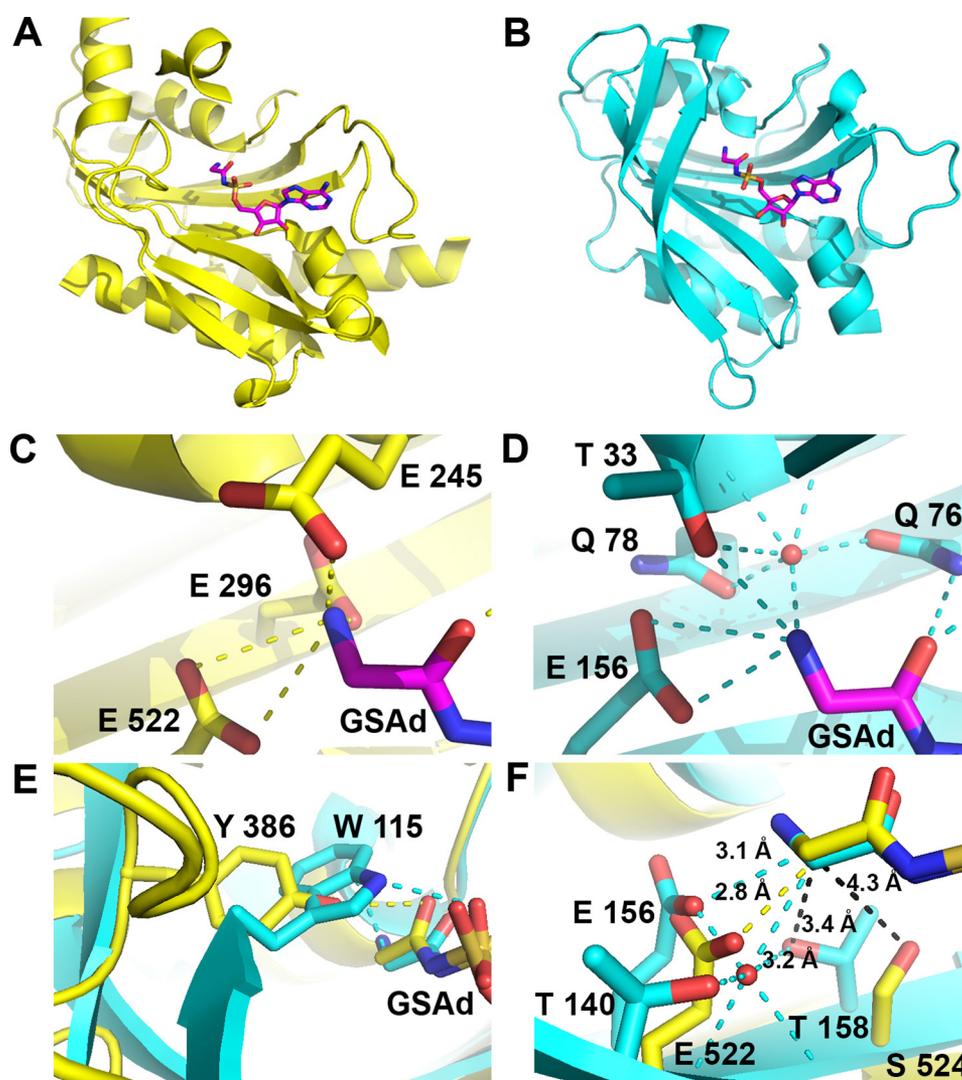


FIGURE 4. **Distinct modes of amino acid recognition in polyphyletic GlyRSs.** *A*, ribbon diagram of the catalytic core of eukaryotic GlyRS (PDB code 2zt8). *B*, ribbon diagram of the catalytic core of bacterial GlyRS. *C*, stick diagram of residues involved in the amino group of glycine recognition in eukaryotic GlyRS. *D*, stick diagram of residues involved in the amino group of glycine recognition in bacterial GlyRS. Thr-33 and Gln-76 have no topological equivalent in eukaryotic GlyRS. All residues involved in the amino group recognition are absolutely conserved in each type of GlyRS. *E*, non-conserved Tyr-386 of the amino acid loop in eukaryotic GlyRS (yellow) compared with the fully conserved Trp-115 of bacterial GlyRS (cyan). *F*, discrimination of non-cognate amino acids by steric hindrance. Non-conserved Ser-524 in eukaryotic GlyRS (yellow) compared with highly conserved Thr-158 and 140 in bacterial GlyRS (cyan).

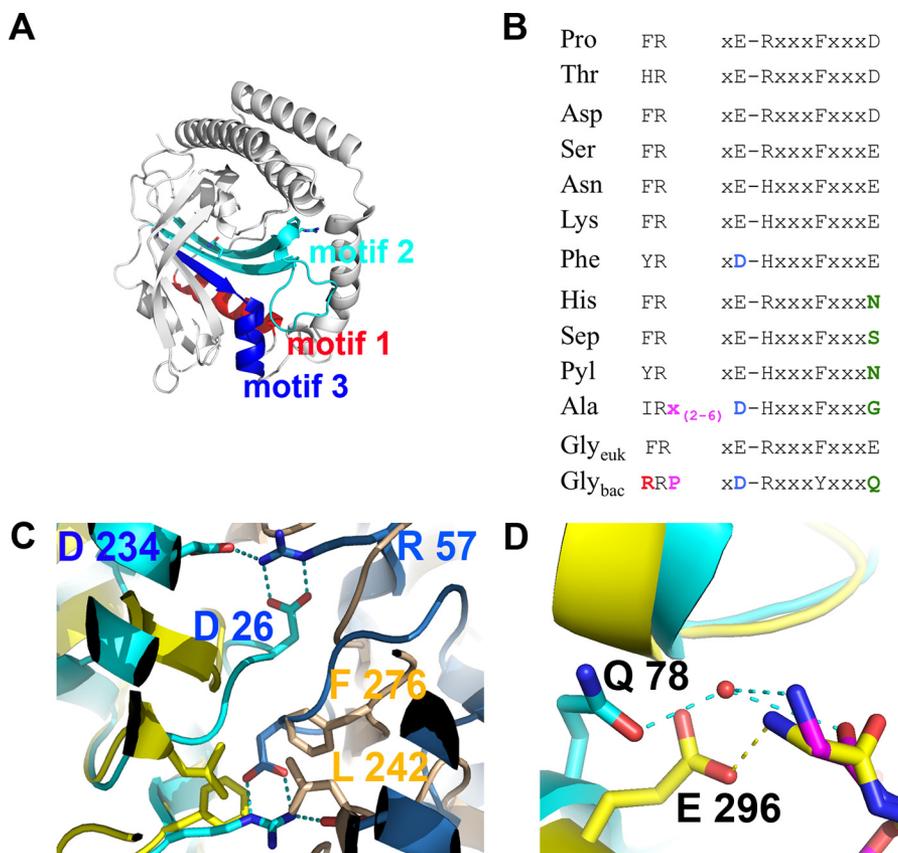
similar to the one seen in bacterial GlyRS. There is also a highly conserved Ala-31 (Ala-99 in AlaRS), which is 4 Å from Trp-115 (Fig. 11, *B* and *E*), besides an equivalent of Glu-156 (Asp-248 in AlaRS) that recognizes the amino group of the cognate amino acid (Fig. 11, *B*, *C* and *D*). In addition, the location of a water molecule that interacts with the amino moiety of the cognate amino acid (Fig. 11*D*) complements the remarkable similarities in the active site cavity. Interestingly, a partial solvent-mediated recognition of the amino group is also found, for example, in ThrRS (46).

Furthermore, there are four residues, outside the active site cavity, that are highly conserved in AlaRS and bacterial GlyRS (Fig. 11, *B* and *F*). These amino acids are Phe-13, Gly-124, Tyr-141, and Gln-143 in bacterial GlyRS, which correspond to Phe-69, Gly-199, Phe-216, and Gln-218 in AlaRS (Fig. 11, *B* and *F*). As mentioned previously, there are no such residues, indicative of a much higher sequence-structure relatedness, in the comparison of both types of GlyRS.

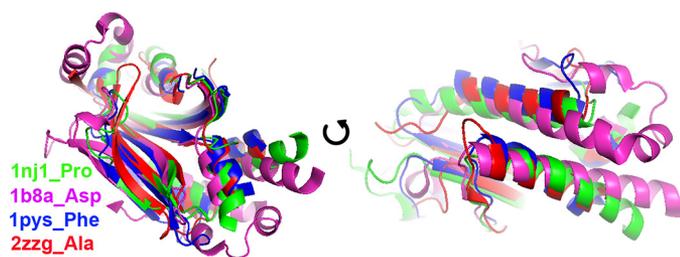
Taken together, the structural and sequence similarities between the catalytic domains of AlaRS and bacterial GlyRS demonstrate a greater degree of conservation than that observed between the two types of GlyRSs. These similarities and differences allow for a new subclass of aaRSs to be defined that is clearly different from any other subgroup.

## Discussion

In this work, we explored three long-standing questions regarding bacterial GlyRS. The first was based on previous experimental evidence suggesting that the full  $\alpha_2\beta_2$  tetramer was needed for any reaction step. The second concerned the comparison of substrate recognition and the active sites of the two types of GlyRSs at the molecular level. Finally, the third question deals with the subclassification of class IIc aaRSs and the possible origin of bacterial GlyRS.



**FIGURE 5. Bacterial GlyRS shows key differences with class II aaRSs.** *A*, three class II motifs are color-coded and labeled on the bacterial  $\alpha$ -GlyRS monomer. Motif 1, residues 3–17; motif 2, residues 48–82; and motif 3, residues 155–169. *B*, signature sequences of motif 2. The consensus sequences were obtained from the analysis of the seed alignments from the Pfam families PF00587 (subclass Ia), PF00152 (subclass Ib), PF01409 (PheRS), PF01411 (AlaRS), and PF02091 (bacterial GlyRS), as well as from our sequence analysis based on structure using Multiseq (see “Experimental Procedures”). The four main differences of bacterial GlyRS with most of class II synthetases are marked in color. Arg-57 and Gln-78 are further analyzed in *C* and *D*, respectively. *C*, differences at the dimerization interface (Arg-57). In bacterial GlyRS (cyan and blue), Arg-57 mediates a salt bridge with Asp-26 and Asp-234. In eukaryotic GlyRS (yellow and light orange), a hydrophobic patch made by Phe-276 and Leu-242 is found in the equivalent region. In some sequences, an Ile residue substitutes Arg-57. This change is correlated with a substitution of Asp-26 by a Thr or Asn residue. *D*, differences in the cognate amino acid binding pocket. Gln-78 in bacterial GlyRS (cyan) helps to bind a solvent molecule that recognizes in part the amino group of the cognate amino acid. In the equivalent position in all class II aaRS, exemplified with eukaryotic GlyRS (yellow), Glu-296 helps to recognize the amino group of glycine.



**FIGURE 6. Initial selection of a catalytic core of class II aaRSs.** A collection of 111 crystal structures from class II aaRSs was chosen according to their diversity in specificity, species, and ligands bound in the active site. The catalytic cores of these structures was selected by including the residues located between motifs 1 and 3 and excluding additional insertions or domains not including common structural motifs for most aaRSs. An overall multiple structural alignment was made using the STAMP algorithm (60) as implemented in MultiSeq in VMD (61, 62). This panel shows the superposition of structures of the catalytic core described and used to obtain the Dali Z-score matrix shown in Table 2, as well as the dendrograms shown in supplemental Fig. S1 (based on structure according to STAMP) and supplemental Fig. S2 (based on sequence according to T-Coffee and built with PhyML). It represents a structure from each obtained subgroup.

*$\alpha$ -Subunit of Bacterial GlyRS Is Dimeric and Active for the First Step of Aminoacylation*—The  $\alpha$ -subunit of bacterial GlyRS contains all the determinants for aminoacyl-adenylate synthesis, and here we show that the activation of glycine can indeed

take place with this chain alone. Given the high sequence conservation of this subunit in bacteria (Pfam family PF2091), it is likely that the same observation would apply to all available enzymes.

The rigorous purification of the  $\alpha$ -subunit, which included a heating step to 75 °C for 30 min ( $T_m$  of the  $\alpha$ -subunit = 83.2 °C, according to thermal shift assays), nickel-nitrilotriacetic acid affinity purification in the presence of relatively high amounts of salts (1 M KCl and 0.5 M urea), and SEC-MALS purification, allowed us to obtain a homogeneous dimeric population of 69.1-kDa protein, free of even trace amounts of higher molecular weight complexes. Previous works have reported that the  $\alpha$ -subunit of bacterial GlyRS contains the aminoacyl adenylate synthesis site, but, unlike this study, no activity had been seen. These findings also support the functional existence of an  $\alpha_2\beta_2$  quaternary structure for bacterial GlyRS, rather than an  $(\alpha\beta)_2$  organization, as seen in most PheRSs (47).

*Two Types of GlyRS Recognize Glycine in Distinct Ways*—The crystal structure of a bacterial GlyRS  $\alpha$ -subunit in a complex with GSAad allows comparison between both types of GlyRS. To function in a meaningful biological context, these synthetases must exclude any amino acid other than glycine, including the

**TABLE 2**
**DALI Z-score of pairwise comparisons of representative structures of class II aaRSs**

Z-scores higher than 18.5 are indicated in bold and grouped by colors that indicate class II subclasses: a, (green); b, (magenta); c, (blue); d, (red).

|    | PDB         | 1           | 2           | 3           | 4           | 5           | 6           | 7           | 8           | 9           | 10          | 11          | 12          | 13          | 14          |
|----|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1  | 1ati<br>Gly |             | <b>18.8</b> | 17.0        | <b>19.0</b> | <b>19.6</b> | 18.2        | 13.4        | 14.1        | 14.0        | 18.1        | 14.2        | 16.1        | 14.5        | 14.7        |
| 2  | 1qe0<br>His | <b>18.7</b> |             | 16.6        | 16.7        | <b>18.5</b> | 16.7        | 12.6        | 12.5        | 12.7        | 18.3        | 14.7        | 17.1        | 12.6        | 13.7        |
| 3  | 1evk<br>Thr | 17.0        | 16.6        |             | <b>23.2</b> | <b>19.2</b> | 17.4        | 10.0        | 13.0        | 10.6        | 13.7        | 12.8        | 14.5        | 11.9        | 12.0        |
| 4  | 1ser<br>Ser | <b>19.0</b> | 16.7        | <b>23.2</b> |             | <b>21.0</b> | <b>19.1</b> | 10.5        | 14.2        | 11.7        | 15.5        | 14.5        | 14.9        | 13.6        | 12.6        |
| 5  | 1nj1<br>Pro | <b>19.6</b> | <b>18.5</b> | <b>19.2</b> | <b>21.0</b> |             | <b>22.4</b> | 12.7        | 12.6        | 13.0        | 17.5        | 14.2        | 16.1        | 13.4        | 13.1        |
| 6  | 2cj9<br>Ser | 18.2        | 16.7        | 17.4        | <b>19.1</b> | <b>22.4</b> |             | 10.6        | 11.3        | 11.6        | 15.3        | 12.9        | 13.7        | 12.8        | 13.1        |
| 7  | 1e1o<br>Lys | 13.4        | 12.6        | 10.1        | 10.5        | 12.7        | 10.6        |             | <b>19.2</b> | <b>23.3</b> | 15.2        | 13.4        | 14.3        | 11.5        | 13.5        |
| 8  | 1b8a<br>Asp | 14.1        | 12.5        | 13.0        | 14.2        | 12.6        | 11.3        | <b>19.2</b> |             | <b>23.4</b> | 15.0        | 11.8        | 13.4        | 12.9        | 14.2        |
| 9  | 1x54<br>Asn | 13.8        | 12.7        | 10.6        | 11.7        | 13.0        | 11.6        | <b>23.3</b> | <b>23.4</b> |             | 15.0        | 11.9        | 13.3        | 11.6        | 12.9        |
| 10 | 2zim<br>Pyl | 18.1        | 18.3        | 13.8        | 15.4        | 17.5        | 15.3        | 15.2        | 15.0        | 15.0        |             | <b>18.5</b> | <b>19.4</b> | 16.1        | 17.3        |
| 11 | 1pys<br>Phe | 11.9        | 14.7        | 12.8        | 12.9        | 14.2        | 12.9        | 13.4        | 11.8        | 11.9        | <b>18.5</b> |             | <b>20.9</b> | 13.2        | 14.6        |
| 12 | 2du3<br>Sep | 16.1        | 17.1        | 14.5        | 14.9        | 16.1        | 13.7        | 14.3        | 13.4        | 13.3        | <b>19.4</b> | <b>20.9</b> |             | 14.0        | 14.8        |
| 13 | 2zzg<br>Ala | 14.5        | 12.6        | 11.9        | 13.6        | 13.4        | 12.8        | 11.5        | 12.9        | 11.6        | 16.1        | 12.0        | 14.0        |             | <b>19.5</b> |
| 14 | 5f5w<br>Gly | 14.7        | 13.7        | 11.8        | 12.7        | 13.1        | 13.1        | 13.5        | 14.2        | 12.8        | 17.3        | 12.2        | 14.8        | <b>19.5</b> |             |

near-cognate alanine. This recognition problem was solved in different ways by the two types of GlyRSs, which further highlight their distinct origin.

Both types of GlyRSs share 1) the same basic architecture of class II synthetases, a six-stranded antiparallel  $\beta$ -sheet. However, the active site domains are not similar enough to share the same subclass, 2) the same ATP-binding site, which is a common feature of all class II synthetases, built using residues of the characteristic motifs 2 and 3. However, the motif 2 signature is different in the two types of GlyRSs, 3) a glutamic acid residue (Glu-156) that helps to recognize the amino moiety of the cognate amino acid glycine.

Beyond these general similarities, there are key differences at the molecular level between the two types of GlyRSs. First, other than the class II antiparallel  $\beta$ -sheet, there are no other structural elements or residues shared by the two types of GlyRS in this domain. Second, the character of the eukaryotic GlyRS dimerization interface is mainly hydrophobic like most class II synthetases, whereas that of the bacterial GlyRS is stabilized by a number of salt bridges. Third, the most impressive difference between the two classes of GlyRSs is their highly distinct strategies for amino acid recognition. Whereas bacterial GlyRS interacts with the amino group of glycine through five absolutely conserved residues, eukaryotic GlyRS uses three absolutely conserved glutamic acids, creating an electronegative pocket, to perform this task. To further illustrate the different character of this region, an absolutely conserved residue in

eukaryotic GlyRS (Arg-410) that interacts with two highly conserved residues, Glu-245 and Glu-522, is substituted by a highly conserved Phe (142) in bacterial GlyRS.

The dynamics of the reaction, exemplified by the amino acid loop and its topological analog in bacterial GlyRS, also differs in both types of GlyRS. In eukaryotic GlyRS, the hydroxyl group of a non-conserved Tyr residue contacts the carboxyl moiety of the cognate glycine. In bacterial GlyRS, an absolutely conserved Trp (115) residue closes the amino acid binding cavity, forming a cation- $\pi$  interaction with the amino group of the cognate glycine. Finally, the discrimination of the near-cognate alanine is made in bacterial GlyRS by two nearly absolutely conserved Thr residues (140 and 158) that also bind a solvent molecule, although in eukaryotic GlyRS this task is done mainly through a non-conserved Ser or Ala residue (524), equivalent of Thr-158 (Fig. 4B). The equivalent of Thr-140 is a non-conserved position (Ala, His, or Ser-408) in eukaryotic GlyRS.

Such disparate differences have only been partially observed until now in the case of SerRSs, where a subset of methanogenic archaea possesses a SerRS with different amino acid recognition elements than those found in the majority of SerRSs (48). However, in this case, both types of SerRS appear to belong to subclass IIa of aaRSs (Figs. 8, 9C, and 10 and PDB entries 2dq0 and 2cj9). Overall, the features of the homologous catalytic core of the two types of GlyRSs are so different that it is extremely unlikely for them to share a common origin.

## Insights into the Polyphyletic Origins of GlyRSs

**TABLE 3**

Proposed subgroups of class II aaRSs

| Subclass | Type of amino acid             | Amino acid       | Quaternary structure | aa site <sup>1</sup> | m2 (#outliers) <sup>2</sup> | Particular Features <sup>3</sup>                  | Post transfer editing <sup>4</sup>         |     |
|----------|--------------------------------|------------------|----------------------|----------------------|-----------------------------|---|--|-----|
| IIa      | small hydrophobic, small polar | Gly              | $\alpha_2$           | 3'                   |                             | ACBD <sup>5</sup> shared with Pro, Thr and HisRSs |  |     |
|          |                                | Pro              | $\alpha_2$           | 3'                   |                             |   | Ala, Cys                                   |     |
|          |                                | Thr              | $\alpha_2$           | 3'                   |                             |   | Bacterial editing domain shared with AlaRS | Ser |
|          |                                | His              | $\alpha_2$           | 3'                   | 1                           |   |  |     |
|          |                                | Ser              | $\alpha_2$           | 3'                   |                             |   |  |     |
| IIb      | charged and large              | Asp              | $\alpha_2$           | 3'                   |                             | ACBD shared with Asn and LysRSs                   |  |     |
|          |                                | Asn              | $\alpha_2$           | 3'                   |                             |   |  |     |
|          |                                | Lys              | $\alpha_2$           | 3'                   |                             |   |  |     |
|          |                                | Pyl <sup>6</sup> | $\alpha_2$           | 3'                   | 1                           |   |  |     |
| IIc      | large                          | Phe              | $(\alpha\beta)_2$    | 2'                   | 1                           |   | Tyr  |     |
|          |                                | Sep              | $\alpha_4$           | 2'                   | 1                           |   |  |     |
| IId      | hydrophobic and small          | Ala              | $\alpha_2$           | 3'                   | 3                           | Helicoidal domain on the C-ter <sup>7</sup>       | Gly, Ser                                   |     |
|          |                                | Gly              | $\alpha_2\beta_2$    | 3'                   | 4                           | Helicoidal domain on the C-ter <sup>7</sup>       |  |     |

<sup>1</sup> Aminoacylation site (2' or 3' OH of the terminal ribose of the cognate tRNA).

<sup>2</sup> Number of residues in motif two that do not agree with the canonical sequence.

<sup>3</sup> Although this proposed classification is based solely on the structural and sequence features of the catalytic core of class IIaaRSs, there are some characteristics that are shared among its members.

<sup>4</sup> Editing carried out after the first step of the reaction against standard amino acids.

<sup>5</sup> ABCD is anticodon binding domain.

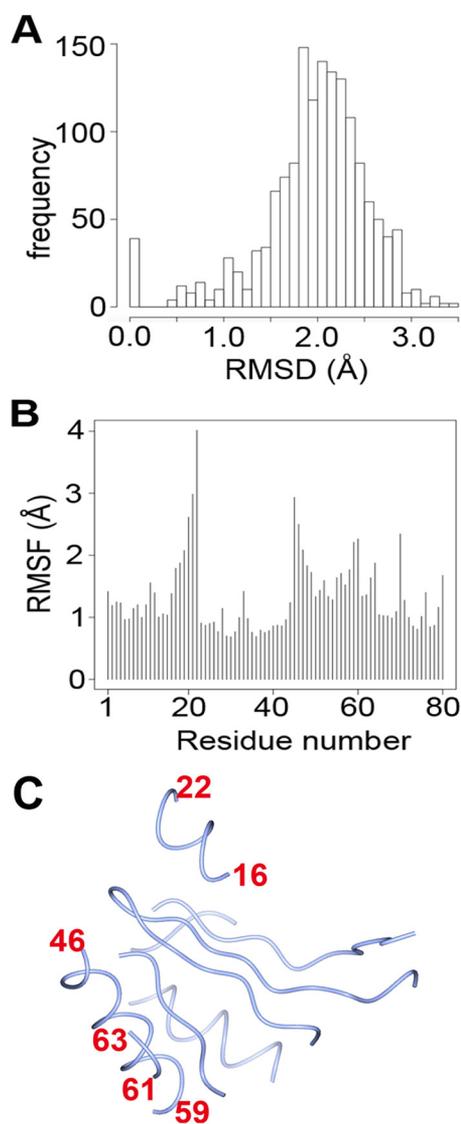
<sup>6</sup> PylRS can be either subclass IIc according to the structural analysis or subclass IIb according to the sequence analysis. It is placed in the frontier in this table because of the aminoacylation site properties of PheRS and SepRS, which are different from PylRS.

<sup>7</sup> These domains have limited similarity between each other (Dali Z-score 2.6, r.m.s.d. 4.5 Å, % identity = 12).

*There Are Four Structural Subclasses of Class II aaRSs*—We used the standard accepted structure and sequence comparison methods to detect structural relationships, including the following: an all-against-all similarity matrix using Dali Z-scores (41, 49, 50); a dendrogram built according to the overall C $\alpha$  r.m.s.d. values derived from the STAMP superposition algorithm (42) as implemented in MultiSeq (51, 52); a PCA and r.m.s.d. dendrogram based on a conserved core of 80 C $\alpha$  atoms using bio3D (44); and a dendrogram built with PhyML (53) according to a weighted multiple sequence alignment made using structural information, as implemented in T-Coffee Expresso (43) and validated using Transitive Consistency Score (TCS) (54). Using these tools and the crystal structures from several species and in complex with different ligands, we were able to define four different class II divisions that are in general agreement with previous proposals. We made this classification proposal based mainly on the structural features of a monomeric catalytic core. Remarkably, even if this core is reduced to

a minimal structure of 80 C $\alpha$  atoms, it is still possible to define the four subclasses. It is noteworthy that the new adjustment of class IIc and IId takes into account the physicochemical character of the cognate amino acids, the aminoacylation site, and correct quaternary structures among the four proposed subclasses (Table 3). PheRS and SepRS are clear functional outsiders from the subclass IId (AlaRS and bacterial GlyRS); the corresponding amino acids are much bigger than Ala and Gly; the aminoacylation sites are different (3'-OH for AlaRS and GlyRS and 2'-OH for Phe and SepRS (27)), and the quaternary structures are also different, most notably between bacterial GlyRS ( $\alpha_2\beta_2$ ) and PheRS ( $(\alpha\beta)_2$ ) (Table 3). In this way, subclasses IIc and IId encompass much more homogeneous and coherent members, as opposed to the previous subclass IIc, which included a wide variety of aaRSs (25–27).

These analyses define a clear subgroup for AlaRS and bacterial GlyRS. According to the sequence analysis, however, it is less clear how to classify PylRS, which has been suggested to



**FIGURE 7. Structural features of the conserved core of 80 C $\alpha$  atoms of class II aaRSs.** 39 structures from the alignment described in Fig. 6 were further trimmed, and the largest positional differences were excluded to determine the invariant core of class II aaRSs using the Bio3D package in R (68). An r.m.s.d. and a PCA, shown in Figs. 8 and 9, were performed using the 80 residues core depicted in C. A, histogram of pairwise r.m.s.d. values B, root mean square fluctuations (RMSF) plot. C, location on the 80-residue core of the main differences shown in B.

belong to subclass IIc (55) (consistent with our structure-based classification) or IIb (56) (as in our sequence analysis). There also appears to be a discrepancy in the case of HisRS at the sequence level. However, HisRS can be confidently placed in class IIa due to the presence of an anticodon recognition domain located on the C-terminal region, which is another signature of this subclass.

Recent phylogenetic trees have not commented on the proposed subdivision (27, 55). However, a relationship between AlaRS and bacterial GlyRS has even been suggested based on the insertions in the catalytic core and the C-terminal helical domain (57, 58). To address this, in this work we presented a complete analysis of all class II synthetases, defined a new subclass, and explored the similarities between AlaRS and bacterial GlyRS.

**Bacterial GlyRS and AlaRS Are Highly Related**—In contrast to its comparison with eukaryal GlyRS, bacterial GlyRS shares several key features with AlaRS. First, these two enzymes display similar catalytic cores whose architectures are distinct from other subclasses. Second, they present a similarly modified motif 2 signature: an insertion between an Arg and an acidic residue and a variation in this position of Glu for Asp, where the overwhelming consensus is a Glu residue in most class II aaRSs. Third, in the active site, they share the same strategy for the recognition of the amino group. Fourth, they share an absolutely conserved Trp involved in the recognition of the cognate amino acids and a highly conserved Ala in the vicinity of this site. An interesting case of a structural/functional convergence is seen in this particular feature in methanogenic archaea SerRS, where a different motif approaches a Trp (396) in a different conformation but with the same apparent role (PDB code 2cj9). Fifth, outside of the active site, AlaRS and bacterial GlyRS share four other highly conserved residues.

**Proposed Models for the Origin of the Two Types of GlyRS**—GlyRSs exhibit an unusual scenario for two enzymes with the same activity and the same overall fold of the catalytic domain. In most cases, these similarities would indicate that they share a common origin. Accordingly, we may consider different scenarios for the structural source of the catalytic core of GlyRSs as follows. 1) A single common ancestor, where an initial  $\alpha_2\beta_2$  GlyRS gave rise to another GlyRS  $\alpha_2$  or vice versa. We consider this hypothesis unlikely, according to the analysis presented in this work. 2) Two different ancestors with an emergence of a dimeric eukaryotic type  $\alpha_2$  from an ancient ancestor for all class II synthetases and another emergence of  $\alpha_2\beta_2$  GlyRS from AlaRS or a pre-AlaRS.

We believe there are several points that support this second hypothesis. First, there is high structural and sequence similarity between the catalytic cores of AlaRS and bacterial GlyRS. Second, the amino acids Ala and Gly are related in nature, only differing by one methyl group and thus may be recognized in a similar manner. In fact, it is possible to delineate changes in bacterial GlyRS in the positions of Thr-158, Thr-140, and/or Glu-256 or in the equivalent and absolutely conserved Gly-250, Val-215, and Asp-248 positions of AlaRS to partially interconvert the amino acid specificity (45, 59). Third, AlaRS presents editing activity against Gly and Ser to cope with the misincorporation of these amino acids (1/300 for Ser and 1/170 for Gly) in the active site (60). It is possible that a pre-AlaRS, able to aminoacylate both Ala and Gly, may have given rise to GlyRS. Taken together, we favor that the second scenario would more easily solve the paradox of the presence of two proteins with the same function and sharing the same primordial fold but not a single common ancestor. The two types of GlyRS represent a beautiful case of isofunctional paralogs between species (61).

## Experimental Procedures

### Protein Purification

The cDNA coding for  $\alpha$ -AaGlyRS was synthesized by GenScript, optimized according to *E. coli* codon frequency, and subcloned into the pET-28 vector (Novagen). Recombinant  $\alpha$ -

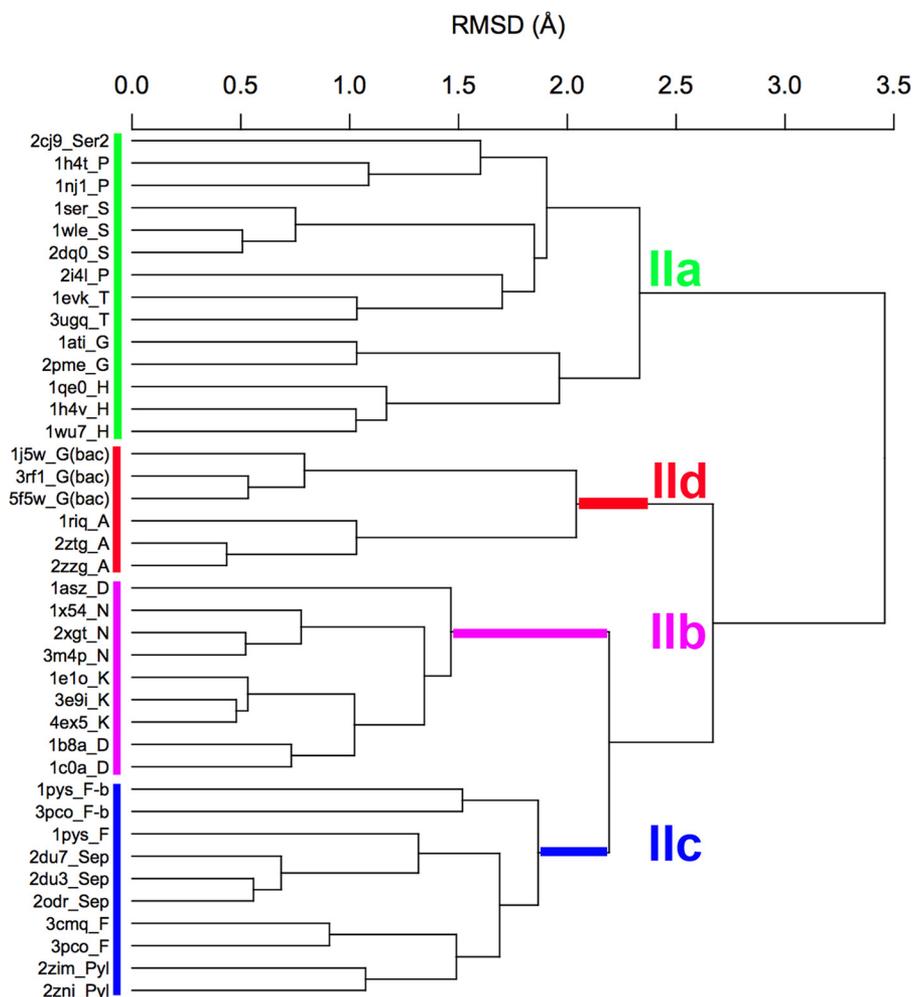


FIGURE 8. **Bacterial GlyRS and AlaRS form a subclass of aaRSs.** r.m.s.d. cluster dendrogram of class II aaRSs. The dendrogram was calculated according to the C $\alpha$  core of 80 residues (see Fig. 7). This tree is in full agreement with previous subclass definitions but with an additional subgroup formed by bacterial GlyRS and AlaRS.

AaGlyRS was expressed with an N-terminal His<sub>6</sub> affinity purification tag in *E. coli* BL21(DE3) cells. Cells were grown at 37 °C until an  $A_{600}$  of 1.0 was reached. At that time, cells were induced for 10 h with a final concentration of 1 mM isopropyl  $\beta$ -D-thiogalactopyranoside at 37 °C and were harvested by centrifugation ( $5515 \times g$  for 20 min). The resulting cell pellet from 2 liters of culture was frozen at  $-80$  °C until needed. For  $\alpha$ -AaGlyRS purification, the cell pellet was thawed and then resuspended in 20 ml of buffer A (50 mM sodium phosphate, pH 6.3, 100 mM potassium chloride, 1 mM MgCl<sub>2</sub>). Cells were lysed by sonication on ice and then centrifuged at  $20,410 \times g$  for 30 min. The resulting supernatant was heated at 75 °C for 30 min and then centrifuged again at  $20,410 \times g$  for 30 min. The supernatant was supplemented with the following additives: 1 M KCl, 0.5 M urea, 50 mM imidazole, and 10% glycerol (final concentrations) and applied to a 5-ml Ni<sup>2+</sup>-Sepharose FF column (GE Healthcare) connected to an Äkta FPLC system (GE Healthcare). The column was then washed with buffer A plus additives to remove nonspecifically bound proteins. The target protein was eluted with a 100-ml gradient of buffer A plus 500 mM imidazole. Fractions containing  $\alpha$ -AaGlyRS were pooled and concentrated (Amicon Ultra-filter Millipore, 30 kDa) to 15 mg/ml. The protein was >99% pure as judged by denaturing gel electrophore-

sis. The protein was loaded onto a size exclusion chromatography Superdex S-75 10/300 analytical column (GE Healthcare) connected to DAWN Heleos-II and Optilab RI detectors (Wyatt Technologies). The column was run at a flow rate of 0.5 ml/min. One peak was eluted that included 99.9% of the injected mass and corresponded to the dimeric form of  $\alpha$ -AaGlyRS, according to the light scattering measurements. Prior to crystallization, the protein was desalted into 20 mM HEPES, pH 7.2, 50 mM NaCl, and 1 mM DTT.

#### Crystallization and Structure Determination

Crystals of  $\alpha$ -AaGlyRS were obtained by sitting-drop vapor diffusion. Drops containing 1  $\mu$ l of  $\alpha$ -AaGlyRS at 15 mg/ml yielded crystals at 18 °C when mixed with 1  $\mu$ l of a reservoir solution containing 30% polyethylene glycol monomethyl ether 2000 and 100 mM potassium thiocyanate. Crystals were soaked for 2 days in a solution containing a final concentration of 28 mM GSAd by mixing 1.3  $\mu$ l of the drop containing crystals, 2.5  $\mu$ l of reservoir solution, and 1.5  $\mu$ l of GSAd at 100 mM. The synthesis of GSAd was performed as described previously (62). Crystals were cryoprotected in a solution prepared with the mother liquor supplemented with 15% glycerol and then flash-frozen in liquid nitrogen. Diffraction data were collected at 100

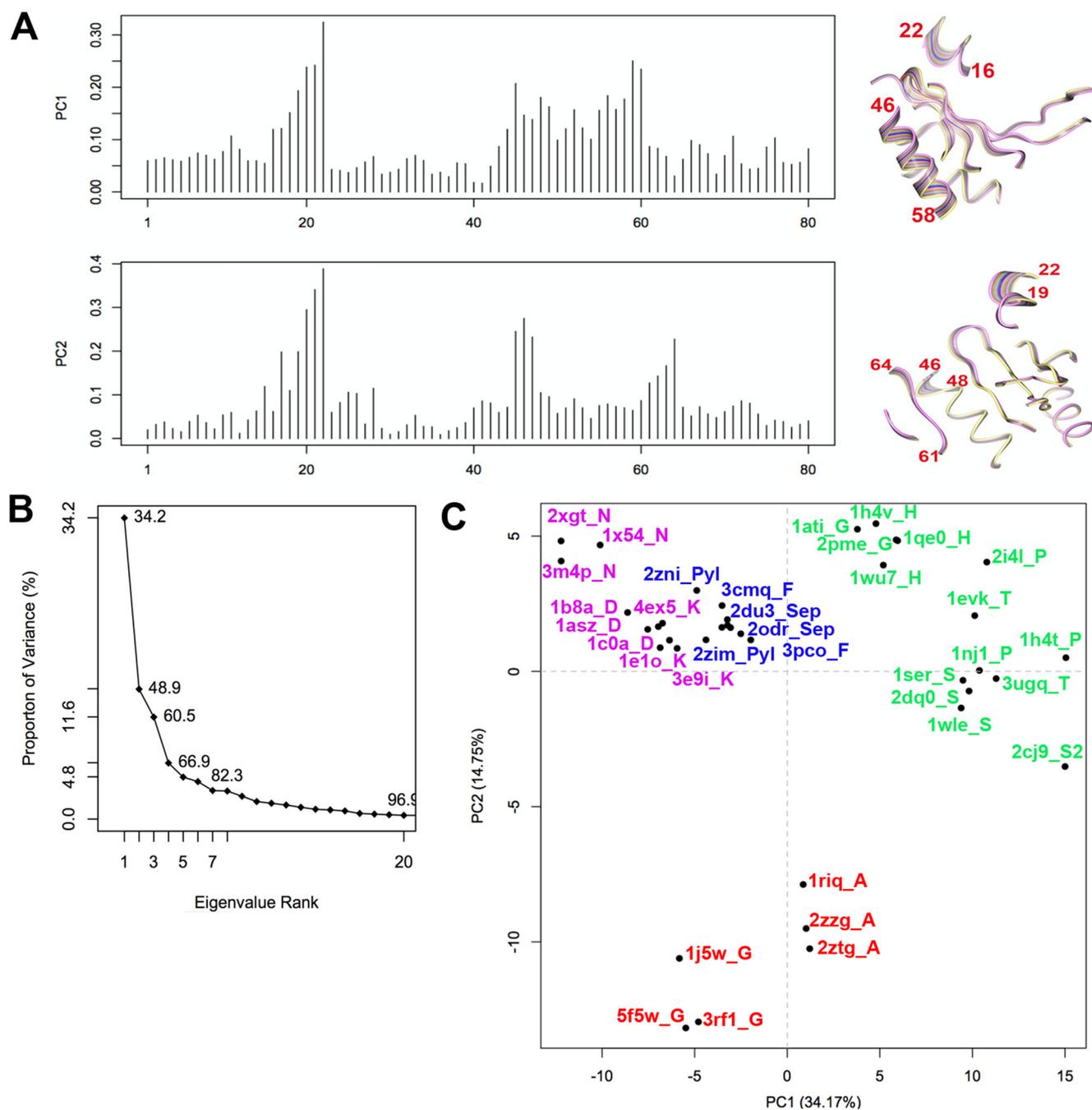


FIGURE 9. **Bacterial GlyRS and AlaRS form a subclass of aaRSs.** PCA of class II aaRSs. *A*, contribution of each of the 80 residues C $\alpha$  core, as defined in the analysis described in Fig. 7, to the first two principal components found. *B*, contribution of each component to the total variance of the distribution of structures. *C*, plot along PC1 and PC2, which describe the change that covers 49% of the structural variance in class II aaRSs.

K using a wavelength of 0.9785 Å at the Life Sciences Collaborative Access Team (LS-CAT) 21-ID-G beamline at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL). Data were indexed and processed with iMOS-FLM (63) and reduced with Aimless (64). The structure was solved by molecular replacement using Balbes (65), with a final search model based on the structure of the  $\alpha$ -subunit GlyRS from *Thermotoga maritima* (PDB code 1j5w). There are five monomers in the asymmetric unit, with a crystal solvent content of 57%. The initial  $R_{\text{free}}$  was 0.35. The ligand GSAd was fitted by means of the LigandFit program of the PHENIX suite (66). Refinement was alternated with manual building/refine-

ment in COOT (67), PHENIX, and the PDB\_REDO server (68). Non-crystallographic symmetry restraints were used for the refinement. The model presents no Ramachandran outliers, with 1355 residues (97%) on the favored and 36 residues (3%) on the allowed regions. *B*-factor refinement was limited to one value per amino acid. Monomers A and B present the best adjustment to the electron density according to the wwPDB validation report (0 residues with RSR-Z scores >2), and all figures were prepared using these monomers as templates. Two solvent molecules were located near the glycine recognition site. They could be well fitted and seen on double difference and omit maps for monomer B, and they can be seen as peaks bigger

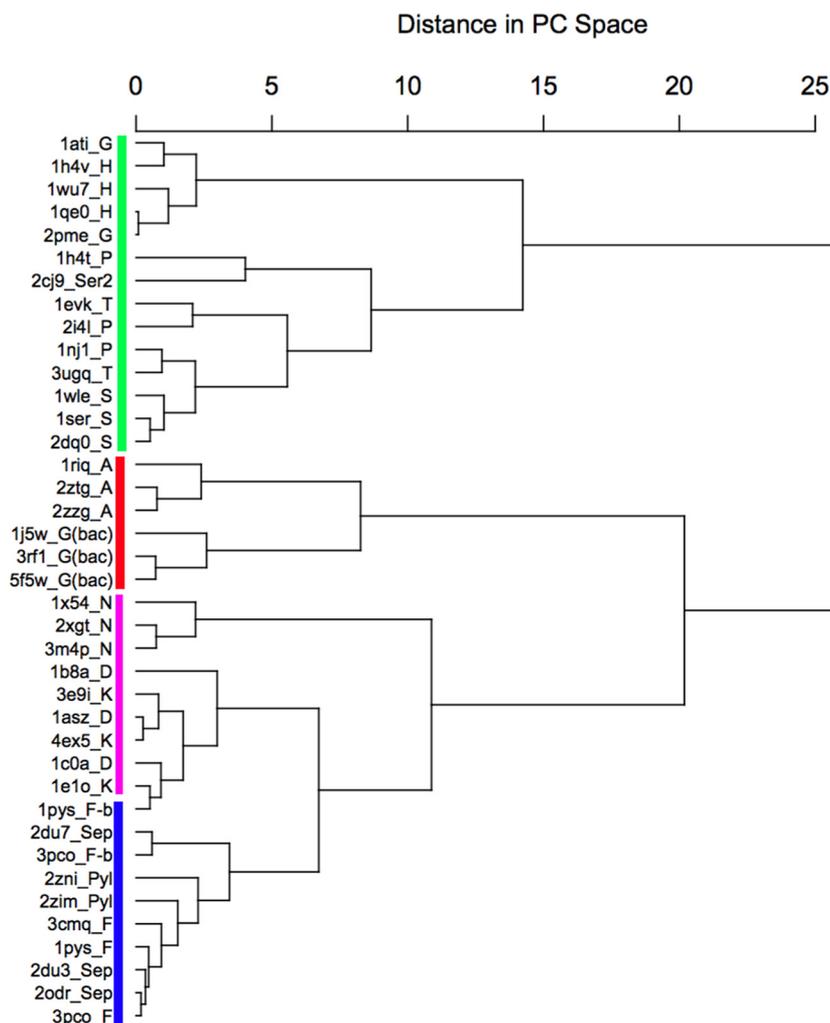


FIGURE 10. PCA cluster dendrogram of classII aaRSs. Clustering dendrogram along PC1 and PC2 as described in Fig. 9.

than  $3.5\sigma$  on difference,  $F_o - F_c$  electron density maps on the four other monomers. No other solvent molecules were added to the model.  $\sigma A$ -weighted  $2F_o - F_c$ -simulated annealing omit maps were used to validate further the quality of the model. Data collection and refinement statistics are summarized in Table 1. Analysis of the interface was performed with the PISA server (69). Because of the absence of ideal geometric parameters from MSDchem, the geometry of the GSAd ligand was adjusted to that observed in PDB code 3hy0 (structure of catalytic fragment of *E.coli* AlaRS in the complex with GlySA). Validation of the GSAd ligand was performed with the ValLigURL server (70). Figures were prepared with PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC).

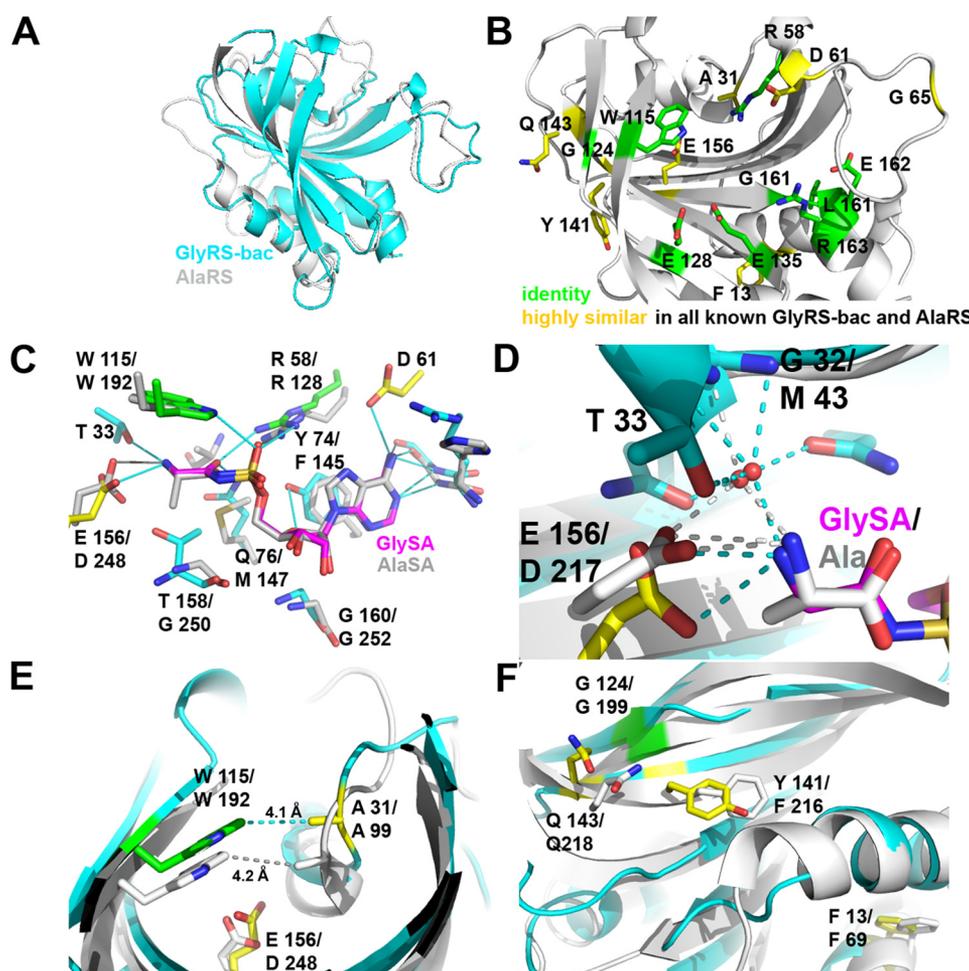
### Thermal Shift Assays

According to the protocol of Ref. 71, purified  $\alpha$ -AaGlyRS in 25 mM  $\text{NaH}_2\text{PO}_4$ , 25 mM KCl, and 10 mM  $\text{MgCl}_2$ , at a final concentration of 1 mg/ml, was mixed with different concentrations of ligands and a 1:100 dilution of SYPRO Orange dye (Invitrogen) in a final volume of 10  $\mu\text{l}$ . The dye was excited at 490 nm and the emitted light intensity was recorded at 575 nm. Data were collected at 1  $^\circ\text{C}$  intervals from 25 to 99  $^\circ\text{C}$  on

a StepOnePlus real time PCR system and analyzed using the Protein Thermal Shift software from Applied Biosystems. No significant differences were detected using three different values of pH (5, 7, and 8). Assays were performed in triplicate.

### Measurements of SAXS Data

Data collection for SAXS of  $\alpha$ -AaGlyRS was performed following standard procedures in the European Molecular Biology Laboratory (EMBL) on the storage ring Petra III (DESY, Hamburg, Germany) on the P12 beamline (72). The scattered intensity was recorded as a function of the scattering vector  $s$  using a wavelength of 0.124 nm at 23  $^\circ\text{C}$ . The measurements were carried out in 30 mM HEPES, pH 7.4, 25 mM NaCl, and 1 mM DTT, with exposure times of  $20 \times 0.05$  s. The average of the data was normalized and subtracted the scattering attributed to the solvent using automatic procedures (73). The SAXS data were processed using PRIMUS (73), where the values of the forward scattering intensity  $I_0$  and radii of gyration  $R_g$  were evaluated from the experimental SAXS patterns using the Guinier approximation; these parameters were also computed from the entire scattering curve using Porod's law by the calculation of the distance distribution function  $P(r)$  using the program



**FIGURE 11. Bacterial GlyRS and AlaRS share several features.** *A*, superposition of the catalytic core, according to Dali, of bacterial GlyRS (residues 1–169) and AlaRS (PDB code 2zzg, residues 58–260). The Z-score is 19.5 for the 160 residues aligned, with r.m.s.d. of 2.3 and 22% identity. *B*, residues that are highly (yellow) or absolutely conserved (green) and also shared by bacterial GlyRS and AlaRS, according to the structural and sequence alignment of all available sequences, are shown on the structure of bacterial GlyRS. The following panels show the role and/or location of some of these residues in more detail. *C*, superposition of the active site region. Some shared residues (Trp-115/Trp-192) and Glu-156/Asp-248 are related with the cognate amino acid recognition. Thr-158/Gly-250 are proposed to be related with amino acid discrimination. Some other residues (Arg-58/Arg-128), for example, are common class II residues aimed to recognize the ATP molecule. *D*, amino group of the cognate amino acid is in part recognized by a solvent molecule and an acidic residue (Glu-156/Asp-217) in both bacterial GlyRS and AlaRS (PDB code 1yfs, in complex with Ala). A positive electron density peak is also seen in the same position of the solvent molecule for PDB code 2zzg, the model that was used for all the figures of this enzyme shown in this work. *E*, Trp-115/Trp-192 makes a cation- $\pi$  interaction on the cognate amino acid that helps to close the active site cavity, together with Ala-31/Ala-99 located in the vicinity. *F*, four other residues are absolutely or highly conserved in both enzymes and are not related to substrate recognition.

GNOM (74). Three different approaches were used to determine the molecular weight as follows: the  $I_0$  value, the Porod volume (from the *Pr* function), and the excluded volume (from DAMMIF calculations). The low resolution shape was reconstructed *ab initio* by the DAMMIF method (75). The theoretical SAXS patterns of dimer and monomer from the crystallographic structure were predicted using CRYSOLOG (76) and were compared graphically with the experimental data using Sasplot (73). The missing flexible residues of the His tag/tobacco etch virus were modeled using Coral (77). The crystallographic structure with the flexible residues was superposed with the *ab initio* envelope using Supcomb (78).

#### Activity Assays

**Glycine Activation**—The first step of the aminoacylation reaction was measured using a previously described method to quantify the synthesis of aminoacyl-adenylate (79). Reaction conditions were initially screened for optimization of

pH, enzyme, ATP, and glycine concentrations. The final reaction mix contained the  $\alpha$ -subunit of Aa-GlyRS at 40  $\mu$ M in 50 mM Tris, pH 8.0, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 0.5 mM ATP, 0.25  $\mu$ Ci of [ $\alpha$ -<sup>32</sup>P]ATP and varying concentrations of glycine. Reactions mixes were pre-incubated for 5 min at 45 °C and started by the addition of enzyme. 1- $\mu$ l time points were quenched in 4  $\mu$ l of buffer containing 0.1% SDS and 400 mM sodium acetate, pH 5.2. 2- $\mu$ l aliquots were spotted on pre-washed PEI-cellulose thin layer chromatography (TLC) plates and developed in 0.1 M ammonium acetate, 5% acetic acid solution. The TLC plates were dried and exposed to a phosphor-screen overnight to monitor Gly-[<sup>32</sup>P]AMP, [<sup>32</sup>P]AMP, and [<sup>32</sup>P]ATP.

Images were obtained on an Amersham Biosciences Storm 820 Imager and quantified with ImageQuant TL version 2005. The total fraction of [<sup>32</sup>P]AMP was plotted against time to obtain initial velocities; the slopes from the linear data were then plotted against glycine concentration and fit to a hyper-

## Insights into the Polyphyletic Origins of GlyRSs

bolic Michaelis-Menten equation using Prism6 software. All experiments were performed in triplicate.

**Aminoacylation Reactions**—According to Refs. 80–82, the aminoacylation of *A. aeolicus* tRNA<sup>Gly</sup> was assayed using identical buffer conditions as for the glycine activation reaction, with the exception of using 5 mM ATP and the omission of labeled ATP. Instead *in vitro* transcribed tRNA<sup>Gly</sup> was radiolabeled at the 3'-end using [ $\alpha$ -<sup>32</sup>P]ATP and the *E. coli* tRNA nucleotidyltransferase, and labeled tRNA was purified using a QIAquick nucleotide removal kit (Qiagen), followed by buffer exchange using Amicon Ultra 3-kDa filter (Millipore). All reactions were initiated by the addition of enzyme, and 1- $\mu$ l aliquots were quenched in 10  $\mu$ l of buffer containing 0.1% SDS and 400 mM sodium acetate, pH 5.2, and 0.1 mM PI nuclease SIGMA (N8630-1VL) or 0.1% SDS, 400 mM sodium acetate, pH 5.2. 0.5  $\mu$ l of the quenched mix were spotted on pre-washed PEI-cellulose TLC plates and developed and treated like the ones for the glycine activation reaction.

### Structural and Sequence Analysis

**STAMP**—A collection of 111 crystal structures from class II aaRSs was chosen according to their diversity in specificity, species, and ligands bound in the active site to take into account conformational changes. The catalytic cores of these structures was selected by including the residues located between motifs 1 and 3 and excluding additional insertions or domains not comprising common structural motifs for most aaRSs. An example of the trimmed catalytic cores is found in Fig. 6. The PDB entries of the structures used and a partial description of them are found in [supplemental Fig. S1](#). The approximate length of each structure of the working set was 200 residues. An overall multiple structural alignment was made using the STAMP algorithm (42) as implemented in MultiSeq in VMD (51, 52). The program generates trees according to four different criteria as follows: Qres (structure similarity per residue); Qh (structural homology); percent identity, and r.m.s.d. Bacterial GlyRS and AlaRS appeared in a separate group in the trees calculated according to all four parameters. Although all the trees were highly similar to each other, the percent identity and r.m.s.d. trees were in full agreement with previous subclass definitions. The same results were obtained when the algorithm was applied to just a sample of representative aaRSs.

Structure-based multiple sequence alignments were made for the cases of the two types of GlyRSs and AlaRS. All the sequences contained in Pfam families PF01411 (AlaRS), PF02091 (bacterial GlyRS) and PF00587 (subclass IIa, but trimmed for eukaryal GlyRS) were analyzed. The catalytic cores were selected and aligned with the ClustalW plug-in in MultiSeq. Non-redundant sets were generated using the sequence QR tool in MultiSeq (83). A profile alignment was calculated taking into account the structures aligned for each type of aaRS.

**DALI**—A sample of structural cores as defined above was used to perform an all-against-all three-dimensional structural comparison with the Dali pairwise comparison server. A matrix was generated based on the resulting Z-score values. Although no tree was derived from this matrix, four clear groups are defined, taking into account a cutoff of 18.5. The Z-score value for structures belonging to other subgroups was between 11.9

and 16.5. In most cases other than PylRS, a clear gap is seen between members and non-members of the subclass.

**T-Coffee Espresso**—A non-redundant set of sequences of the catalytic core of 63 structures from different species used for the alignment with STAMP were aligned using structural information with the T-Coffee Espresso server (43). The obtained alignment was further validated by means of the suggested TCS method, which identifies the most correct positions on the alignment (54). An estimated phylogenetic tree was built with PhyML (53) according to the tcs\_weighted obtained file, where the sequences are ordered according to their TCS score. The tree was visualized with Archaeopteryx (84).

**Root Mean Square Deviation and PCA with Bio3D**—The analysis was performed using the Bio3D package in R (44). 39 structures derived from the STAMP analysis were selected, three of different species and in complex with different ligands for each type of aaRS. The aligned structures were used in Bio3D to define an invariant core with the minimal structural variance among all the protein structures. This core of 80 C $\alpha$  atoms was the basis for the r.m.s.d. dendrogram calculation and PCA. The structures were superimposed onto this core, and the variances shown by PC1 and PC2 were used to define the four different subgroups.

---

**Author Contributions**—M. I. V. S., A. T. L., A. C. D. B., and D. M. conceived the project. M. I. V. S. and H. A. S. S. purified  $\alpha$ -AaGlyRS and performed thermal shift assays. A. R. H. and L. B. C. performed the activity experiments. M. I. V. S. purified, crystallized, and solved the structure of the  $\alpha$ -AaGlyRS-GlyAMS complex. M. I. V. S., H. M., and D. S. performed and analyzed the SAXS experiments. R. F. and M. G. synthesized the Gly-SA compound. D. M., B. B., M. A., M. I. V. S., and A. T. L. performed the structural and sequence analysis. All authors analyzed the data and contributed to manuscript preparation. M. I. V. S. and A. T. L. wrote the manuscript.

---

**Acknowledgments**—We thank Elizabeth Nallely Cabrera González and Nevena Cvetesic for scientific and technical guidance, Prof. John J. Perona for the gift of gene coding for the tRNA nucleotidyltransferase, and Martha Contreras and Beatriz Aguirre López for technical support. The critical reading by Jonathan Ipsaro and Rajan Sankaranarayanan is greatly acknowledged. Use of the Advanced Photon Source, an Office of Science User Facility operated for the United States Department of Energy Office of Science at Argonne National Laboratory, was supported by the United States Department of Energy under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corp. and Michigan Technology Tri-Corridor Grant 085P1000817.

---

### References

1. Berg, P., and Offengand, E. J. (1958) An enzymatic mechanism for linking amino acids to RNA. *Proc. Natl. Acad. Sci. U.S.A.* **44**, 78–86
2. Ibba, M., and Soll, D. (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69**, 617–650
3. Pang, Y. L., Poruri, K., and Martinis, S. A. (2014) tRNA synthetase: tRNA aminoacylation and beyond. *Wiley Interdiscip. Rev. RNA* **5**, 461–480
4. Berg, P. (1956) Acyl adenylate—synthesis and properties of adenyl acetate. *J. Biol. Chem.* **222**, 1015–1023
5. Preiss, J., Berg, P., Ofengand, E. J., Bergmann, F. H., and Dieckmann, M. (1959) The chemical nature of the RNA-amino acid compound formed by amino acid-activating enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **45**, 319–328

6. Berg, P., Bergmann, F. H., Ofengand, E. J., and Dieckmann, M. (1961) Enzymic synthesis of amino acyl derivatives of ribonucleic acid. *J. Biol. Chem.* **236**, 1735–1740
7. Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347**, 203–206
8. Cusack, S., Berthet-Colominas, C., Härtle, M., Nassar, N., and Leberman, R. (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature* **347**, 249–255
9. Burbaum, J. J., Starzyk, R. M., and Schimmel, P. (1990) Understanding structural relationships in proteins of unsolved three-dimensional structure. *Proteins* **7**, 99–111
10. Burbaum, J. J., and Schimmel, P. (1991) Structural relationships and the classification of aminoacyl-transfer RNA-synthetases. *J. Biol. Chem.* **266**, 16965–16968
11. Wolf, Y. I., Aravind, L., Grishin, N. V., and Koonin, E. V. (1999) Evolution of aminoacyl-tRNA synthetases—analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res.* **9**, 689–710
12. O'Donoghue, P., and Luthey-Schulten, Z. (2003) On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol. Mol. Biol. Rev.* **67**, 550–573
13. Schimmel, P. (2008) Development of tRNA synthetases and connection to genetic code and disease. *Protein Sci.* **17**, 1643–1652
14. Shiba, K., Schimmel, P., Motegi, H., and Noda, T. (1994) Human glycyl-tRNA synthetase. Wide divergence of primary structure from bacterial counterpart and species-specific aminoacylation. *J. Biol. Chem.* **269**, 30049–30055
15. Fournier, G. P., Andam, C. P., and Gogarten, J. P. (2015) Ancient horizontal gene transfer and the last common ancestors. *BMC Evol. Biol.* **15**, 70
16. Chien, C. I., Chen, Y. W., Wu, Y. H., Chang, C. Y., Wang, T. L., and Wang, C. C. (2014) Functional substitution of a eukaryotic glycyl-tRNA synthetase with an evolutionarily unrelated bacterial cognate enzyme. *PLoS ONE* **9**, e94659
17. Logan, D. T., Mazauric, M. H., Kern, D., and Moras, D. (1995) Crystal structure of glycyl-transfer-RNA synthetase from *Thermus thermophilus*. *EMBO J.* **14**, 4156–4167
18. Mazauric, M. H., Reinbolt, J., Lorber, B., Ebel, C., Keith, G., Giegé, R., and Kern, D. (1996) An example of non-conservation of oligomeric structure in prokaryotic aminoacyl-tRNA synthetases. Biochemical and structural properties of glycyl-tRNA synthetase from *Thermus thermophilus*. *Eur. J. Biochem.* **241**, 814–826
19. Freist, W., Logan, D. T., and Gauss, D. H. (1996) Glycyl-tRNA synthetase. *Biol. Chem. Hoppe-Seyler* **377**, 343–356
20. Okamoto, K., Kuno, A., and Hasegawa, T. (2005) Recognition sites of glycine tRNA for glycyl-tRNA synthetase from hyperthermophilic archaeon, *Aeropyrum pernix* K1. *Nucleic Acids Symposium Series* **49**, 299–300
21. Ostrem, D. L., and Berg, P. (1974) Glycyl transfer ribonucleic acid synthetase from *Escherichia coli*: purification, properties, and substrate binding. *Biochemistry* **13**, 1338–1348
22. Mazauric, M. H., Keith, G., Logan, D., Kreutzer, R., Giegé, R., and Kern, D. (1998) Glycyl-tRNA synthetase from *Thermus thermophilus*—wide structural divergence with other prokaryotic glycyl-tRNA synthetases and functional inter-relation with prokaryotic and eukaryotic glycylation systems. *Eur. J. Biochem.* **251**, 744–757
23. Tang, S. N., and Huang, J. F. (2005) Evolution of different oligomeric glycyl-tRNA synthetases. *FEBS Lett.* **579**, 1441–1445
24. Tan, K., Zhou, M., Zhang, R., Anderson, W. F., and Joachimiak, A. (2012) The crystal structures of the  $\alpha$ -subunit of the  $\alpha(2)\beta(2)$  tetrameric glycyl-tRNA synthetase. *J. Struct. Funct. Genomics* **13**, 233–239
25. Woese, C. R., Olsen, G. J., Ibba, M., and Söll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**, 202–236
26. Perona, J. J., and Hadd, A. (2012) Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. *Biochemistry* **51**, 8705–8729
27. Englert, M., Moses, S., Hohn, M., Ling, J., O'Donoghue, P., and Söll, D. (2013) Aminoacylation of tRNA 2'- or 3'-hydroxyl by phosphoseryl- and pyrrolysyl-tRNA synthetases. *FEBS Lett.* **587**, 3360–3364
28. Ribas de Pouplana, L., and Schimmel, P. (2001) Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. *Cell* **104**, 191–193
29. Nawaz, M., and Martinis, S. A. (2008) *Wiley Encyclopedia of Chemical Biology*. 10.1002/9780470048672.webc008. John Wiley & Sons, Inc., New York
30. Giege, R., and Springer, M. (2013) Aminoacyl tRNA synthetases in the bacterial world. *EcoSal. Plus*. 10.1128/ecosalplus.ESP-0002–201
31. Klipcan, L., Saforo, M., and Schimmel, P. (2013) Anticodon G recognition by tRNA synthetases mimics the tRNA core. *Trends Biochem. Sci.* **38**, 229–232
32. Caetano-Anollés, G., Wang, M., and Caetano-Anollés, D. (2013) Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. *PLoS ONE* **8**, e72225
33. Toth, M. J., and Schimmel, P. (1990) A mutation in the small ( $\alpha$ ) subunit of glycyl-transfer RNA-synthetase affects amino-acid activation and subunit association parameters. *J. Biol. Chem.* **265**, 1005–1009
34. Toth, M. J., and Schimmel, P. (1990) Deletions in the large ( $\beta$ ) subunit of a hetero-oligomeric aminoacyl-transfer RNA-synthetase. *J. Biol. Chem.* **265**, 1000–1004
35. Ostrem, D. L., and Berg, P. (1970) Glycyl-transfer RNA synthetase—an oligomeric protein containing dissimilar subunits. *Proc. Natl. Acad. Sci. U.S.A.* **67**, 1967–1974
36. McDonald, T., Breite, L., Pangburn, K. L., Hom, S., Manser, J., and Nagel, G. M. (1980) Overproduction, purification, and subunit structure of *Escherichia coli* glycyl transfer ribonucleic-acid synthetase. *Biochemistry* **19**, 1402–1409
37. Nagel, G. M., Cumberledge, S., Johnson, M. S., Petrella, E., and Weber, B. H. (1984) The  $\beta$  subunit of *E. coli* glycyl-tRNA synthetase plays a major role in tRNA recognition. *Nucleic Acids Res.* **12**, 4377–4384
38. Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* **378**, 634–652
39. Arnez, J. G., Dock-Bregeon, A. C., and Moras, D. (1999) Glycyl-tRNA synthetase uses a negatively charged pit for specific recognition and activation of glycine. *J. Mol. Biol.* **286**, 1449–1459
40. Schimmel, P., Giegé, R., Moras, D., and Yokoyama, S. (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8763–8768
41. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
42. Russell, R. B., and Barton, G. J. (1992) Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins* **14**, 309–323
43. Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., and Notredame, C. (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.* **34**, W604–W608
44. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A., and Caves, L. S. (2006) Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696
45. Guo, M., Chong, Y. E., Shapiro, R., Beebe, K., Yang, X. L., and Schimmel, P. (2009) Paradox of mistranslation of serine for alanine caused by AlaRS recognition dilemma. *Nature* **462**, 808–812
46. Sankaranarayanan, R., Dock-Bregeon, A. C., Rees, B., Bovee, M., Caillet, J., Romby, P., Francklyn, C. S., and Moras, D. (2000) Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase. *Nat. Struct. Biol.* **7**, 461–465
47. Finarov, I., Moor, N., Kessler, N., Klipcan, L., and Saforo, M. G. (2010) Structure of human cytosolic phenylalanyl-tRNA synthetase: evidence for kingdom-specific design of the active sites and tRNA binding patterns. *Structure* **18**, 343–353
48. Bilokapic, S., Maier, T., Ahel, D., Gruic-Sovulj, I., Söll, D., Weygand-Durasevic, I., and Ban, N. (2006) Structure of the unusual seryl-tRNA

- synthetase reveals a distinct zinc-dependent mode of substrate recognition. *EMBO J.* **25**, 2498–2509
49. Lichtarge, O. (2001) Getting past appearances: the many-fold consequences of remote homology. *Nat. Struct. Biol.* **8**, 918–920
  50. Majorek, K. A., Dunin-Horkawicz, S., Steczkiewicz, K., Muszewska, A., Nowotny, M., Ginalski, K., and Bujnicki, J. M. (2014) The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res.* **42**, 4160–4179
  51. Roberts, E., Eargle, J., Wright, D., and Luthey-Schulten, Z. (2006) MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC Bioinformatics* **7**, 382
  52. Eargle, J., Wright, D., and Luthey-Schulten, Z. (2006) Multiple alignment of protein structures and sequences for VMD. *Bioinformatics* **22**, 504–506
  53. Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321
  54. Chang, J. M., Di Tommaso, P., and Notredame, C. (2014) TCS: A new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction. *Mol. Biol. Evol.* **31**, 1625–1637
  55. Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Söll, D., and Steitz, T. A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11268–11273
  56. Fournier, G. P., Huang, J., and Gogarten, J. P. (2009) Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 2229–2239
  57. Swairjo, M. A., Otero, F. J., Yang, X. L., Lovato, M. A., Skene, R. J., McRee, D. E., Ribas de Pouplana, L., and Schimmel, P. (2004) Alanyl-tRNA synthetase crystal structure and design for acceptor-stem recognition. *Mol. Cell* **13**, 829–841
  58. Smith, T. F., and Hartman, H. (2015) The evolution of class II aminoacyl-tRNA synthetases and the first code. *FEBS Lett.* **589**, 3499–3507
  59. Beebe, K., Ribas De Pouplana, L., and Schimmel, P. (2003) Elucidation of tRNA-dependent editing by a class II tRNA synthetase and significance for cell viability. *EMBO J.* **22**, 668–675
  60. Schimmel, P. (2011) Mistranslation and its control by tRNA synthetases. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2965–2971
  61. Gabaldón, T., and Koonin, E. V. (2013) Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* **14**, 360–366
  62. Redwan, I. N., Ingemyr, H. J., Ljungdahl, T., Lawson, C. P., and Grotli, M. (2012) Solid-Phase synthesis of 5'-O-[N-(acyl)sulfamoyl]adenosine derivatives. *Eur. J. Org. Chem.* **2012**, 3665–3669
  63. Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 271–281
  64. Evans, P. R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 282–292
  65. Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 125–132
  66. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
  67. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
  68. Joosten, R. P., Long, F., Murshudov, G. N., and Perrakis, A. (2014) The PDB\_REDO server for macromolecular structure model optimization. *IUCr. J.* **1**, 213–220
  69. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
  70. Kleywegt, G. J., and Harris, M. R. (2007) ValLigURL: a server for ligand-structure comparison and validation. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 935–938
  71. Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221
  72. Blanchet, C. E., Hermes, C., Svergun, D. I., and Fiedler, S. (2015) A small and robust active beamstop for scattering experiments on high-brilliance undulator beamlines. *J. Synchrotron Radiat.* **22**, 461–464
  73. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282
  74. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503
  75. Franke, D., and Svergun, D. I. (2009) DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346
  76. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOLE—A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773
  77. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350
  78. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41
  79. Gruic-Sovulj, I., Uter, N., Bullock, T., and Perona, J. J. (2005) tRNA-dependent aminoacyl-adenylate hydrolysis by a nonediting class I aminoacyl-tRNA synthetase. *J. Biol. Chem.* **280**, 23978–23986
  80. Ledoux, S., and Uhlenbeck, O. C. (2008) [<sup>32</sup>P]-labeling tRNA with nucleotidyltransferase for assaying aminoacylation and peptide bond formation. *Methods* **44**, 74–80
  81. Rodríguez-Hernández, A., Bhaskaran, H., Hadd, A., and Perona, J. J. (2010) Synthesis of Glu-tRNA(Gln) by engineered and natural aminoacyl-tRNA synthetases. *Biochemistry* **49**, 6727–6736
  82. Rodríguez-Hernández, A., and Perona, J. J. (2011) Heat maps for intramolecular communication in an RNP enzyme encoding glutamine. *Structure* **19**, 386–396
  83. O'Donoghue, P., and Luthey-Schulten, Z. (2005) Evolutionary profiles derived from the QR factorization of multiple structural alignments gives an economy of information. *J. Mol. Biol.* **346**, 875–894
  84. Han, M. V., and Zmasek, C. M. (2009) phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* **10**, 356