Architectures of Class-Defining and Specific Domains of Glutamyl-tRNA Synthetase

Osamu Nureki, Dmitry G. Vassylyev, Katsuo Katayanagi, Toshiyuki Shimizu, Shun-ichi Sekine, Takanori Kigawa, Tatsuo Miyazawa, Shigeyuki Yokoyama,* Kosuke Morikawa*

The crystal structure of a class I aminoacyl–transfer RNA synthetase, glutamyl-tRNA synthetase (GluRS) from *Thermus thermophilus*, was solved and refined at 2.5 Å resolution. The amino-terminal half of GluRS shows a geometrical similarity with that of *Escherichia coli* glutaminyl-tRNA synthetase (GlnRS) of the same subclass in class I, comprising the class I–specific Rossmann fold domain and the intervening subclass-specific α/β domain. These domains were found to have two GluRS-specific, secondary-structure insertions, which then participated in the specific recognition of the D and acceptor stems of $tRNA^{Glu}$ as indicated by mutagenesis analyses based on the docking properties of GluRS and tRNA. In striking contrast to the β-barrel structure of the GlnRS carboxyl-terminal half, the GluRS carboxyl-terminal half displayed an all-α-helix architecture, an α-helix cage, and mutagenesis analyses indicated that it had a role in the anticodon recognition.

Aminoacyl-tRNA synthetases (aaRS's) strictly recognize and ligate their specific tRNA and amino acid, thus contributing to the fidelity of translation of genetic information. In spite of the common features of the aminoacylation reaction, the 20 aaRS's exhibit broad structural diversity. In 1990, Eriani *et al.* proposed, on the basis of the adenosine triphosphate (ATP)–binding motifs, that the 20 aaRS's are divided into two classes, each consisting of 10 members (*1*). This classification correlates well with that based on the specificity toward either the 2′-OH or the 3′-OH of the tRNA terminal adenosine as the amino acid attachment site (*1*). The members of classes I and II have been subdivided into three and four subclasses, respectively; the amino acid substrates for a given subclass show a certain similarity in chemical properties (*1*). The crystal structures of five aaRS's determined thus far (*2*–*6*) have provided a structural basis for this classification. TyrRS, MetRS, and GlnRS, which belong to class I, have a canonical Rossmann fold (five or six parallel β sheets flanked by α helices), including the characteristic ATP-binding sequences, HIGH (His-Ile-Gly-His) and KMSK (Lys-Met-Ser-Lys) (*7*), and at the end an α-helical subdomain (*2*–*4*). In contrast, SerRS and AspRS, which belong to class II, exhibit a central seven-stranded, antiparallel β sheet for three motifs for ATP binding (*5*, *6*). Into the common scaffolds of class-defining domains, different sequences and domains, which possibly define the subclasses, intervene. These functional and structural features suggest that the two sets of 10 aaRS's have evolved from distinct origins (*1*).

In addition to the class- or subclass-defining domains, aaRS's appear to have nonconserved domains, which are structurally diversified. For example, the nonconserved domains of GlnRS (class I) (*4*) and AspRS (class II) (*6*) show antiparallel β-barrel architectures, whereas those of TyrRS (class I) (*2*), MetRS (class I) (*3*), and SerRS (class II) (*5*) show all-α-helix constructions. It has been proposed that the two parts of the aaRS, the class or subclass-defining domains and the nonconserved domains, interact with two distinct domains of the tRNA molecule, for example, the acceptor-TΨC helix and the dihydrouridine-anticodon stem biloops, respectively (*8*). Because the above five aaRS's belong to different subclasses, their specific structural elements, such as the nonconserved domains participating in the tRNA recognition, are too diversified to indicate how these elements were established in aaRS evolution. Therefore, we expected that a comparison of two related aaRS's in the same subclass could provide a structural basis for the origin of aaRS's in terms of the strict tRNA recognition.

The glutaminyl- and glutamyl-tRNA synthetases (GlnRS and GluRS, respectively) are closely related and show extensive sequence similarity (*9*). Gram-negative bacteria and the cytoplasm of eukaryotes have GlnRS which aminoacylates glutamine transfer RNA ($tRNA^{Gln}$) with glutamine. In contrast, when GlnRS is missing, GluRS aminoacylates both $tRNA^{Glu}$ and $tRNA^{Gln}$ with glutamate, and a transamidase converts the Glu-$tRNA^{Gln}$ to Gln-$tRNA^{Gln}$, in other systems (*10*, *11*). Recently, as a result of primary structure analysis, it was proposed that the prokaryotic GlnRS of Gram-

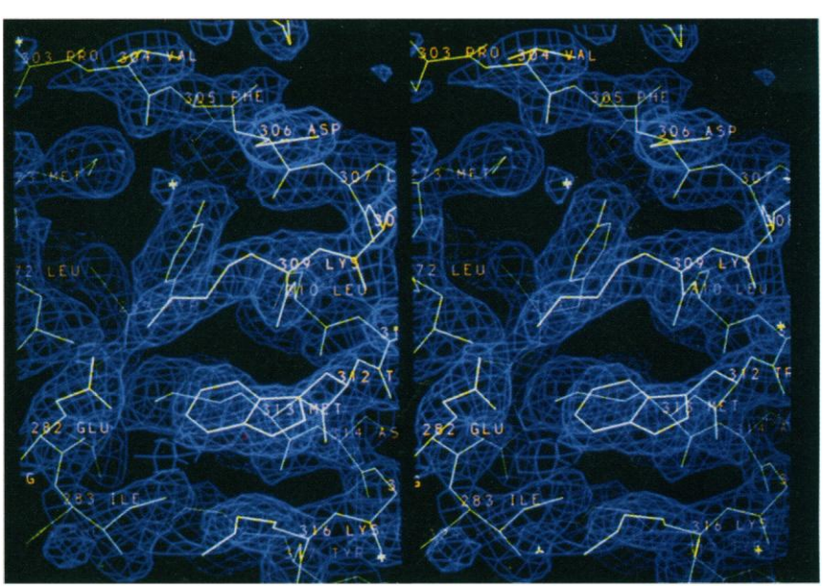

Fig. 1. A stereo drawing of electron density contoured at 1σ with ($|2F_o| - |F_c|$) amplitudes. The final 2.5 Å refined model is superimposed on the electron density. A view of the tRNA-recognition surface is shown.

O. Nureki, S.-i. Sekine, T. Kigawa, and S. Yokoyama are at the Department of Biophysics and Biochemistry, School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan. D. G. Vassylyev, K. Katayanagi, T. Shimizu, T. Miyazawa, and K. Morikawa are at the Protein Engineering Research Institute, 6-2-3 Furuedai, Suita, Osaka 565, Japan.

*To whom correspondence should be addressed.

negative bacteria was of eukaryotic origin and was acquired by horizontal gene transfer (*12*). Here we describe the x-ray crystallographic three-dimensional (3D) structure of GluRS from the extreme thermophile *Thermus thermophilus*. On the basis of a model of the GluRS·tRNA complex, we then performed site-directed mutagenesis of GluRS to identify amino acid residues that take part in the specific recognition of the tRNA.

Structure determination. *T. thermophilus* GluRS consists of a 468-residue, monomeric subunit, with a molecular mass of 53,900 daltons; its gene has been cloned and sequenced (*13*). The GluRS protein has been purified and crystallized (*13*). Four different crystal forms were produced, depending on variations of pH, ionic strength, and concentrations of protein and precipitant (*14*). Crystals suitable for x-ray analyses were grown by macro-seeding, and one crystal form was selected among them. The best crystals belong to the orthorhombic space group $P2_12_12_1$, with unit cell parameters of $a = 75.8$, $b = 110.1$, and $c = 67.6$ Å. The asymmetric unit contains a single subunit, resulting in a solvent content of 52.9 percent.

Initial phases were calculated from two heavy atom derivatives, namely, methyl mercuric chloride and $(NH_3)_4Pt/PtCl_4$ (Table 1). The wild-type protein contains a single Cys residue. To create an additional Hg-binding site, we introduced a second Cys residue at Ala^{22}, where the corresponding position in the homologous *Escherichia coli* GlnRS was Cys^{48} (on the first α helix of the Rossmann fold), a major Hg-binding site (*4*). This GluRS mutant (A22C) retained full activity and produced isomorphous crystals after the native crystals were seeded. The mercury derivative of the mutant contained an additional, clear peak in the difference Patterson map that improved the phasing statistics (Table 1). In addition to this derivative, four other derivatives were used to produce the multiple isomorphous replacement (MIR) map for model building (Table 1). The final mean figure of merit from the seven heavy atom derivatives, augmented with anomalous dispersion, was 0.76 at 3.0 Å resolution (Table 1).

The graphic program FRODO (*15*) was used to obtain the initial model and to manually fit it to the 3.0 Å MIR electron density map. This map was improved by means of the solvent flattening program (*16*). As an additional density modification procedure, the program SQUASH (*17*), was used, which improved the quality of the map, so that the complete polypeptide chain could be unambiguously traced, including a region (residues 100 to 186) that was poorly defined in the solvent-flattened map.

The model was refined with the use of alternating rounds of the X-PLOR (*18*) simulated annealing refinement program and manual rebuilding, with a gradual extension of the resolution from 3.0 to 2.5 Å. Further refinement with the program PROLSQ (*19*) improved the geometry of the model and reduced the crystallographic *R* factor by

Table 1. Summary of crystallographic analysis. Intensity data to 2.8 Å were collected at 4°C with an automated oscillation camera system DIP100 (Mac Science) equipped with a Rigaku RU300 rotating anode generator. The data were processed with the program ELMS (*37*). Intensity data for the Hg-aniline derivative were collected on a FAST area detector (Enraf-Nonius), and processed with the program MADNESS (*38*). Higher resolution data set for the native and derivative crystals were collected to 2.5 Å with a Weissenberg camera for macromolecules (*39*) installed on the beamline 6A2 at the Photon Factory (Tsukuba, Japan). The data were processed by the program WEIS (*40*). Scaling, phasing, and heavy atom parameter refinement were performed with the program PROTEIN (*41*) or programs in the CCP4 package (*42*). Heavy atom sites were determined by isomorphous difference Patterson maps and confirmed by cross-difference Fourier maps. MIR phases were determined essentially by CH_3HgCl and $(NH_3)_4Pt/PtCl_4$ derivatives of the wild-type and a CH_3HgCl derivative of A22C mutant; Hg-aniline binds to the same site as CH_3HgCl, and cis-[Pt(methyl)-$NH_2)_2Cl_2$] and $Pt(NH_3)_2Cl_2$ share two sites with $(NH_3)_4Pt/PtCl_4$. Platinum tends to bind to multiple sites, inducing nonisomorphism, which will explain the poor phasing power of $Pt(NH_3)_2Cl_2$ and K_2[Pt(oxalate)$_2$]. Anomalous scattering data from all derivatives but Hg-aniline were used. Solvent flattening (eight cycles, 40 percent solvent content) was performed (*16*). Phases were further improved by the program Histogram Matching in the SQUASH system (*17*) so that the electron density distribution could match the ideal one. Crystallographic refinement was carried out with the X-PLOR package (*18*); the initial model was subjected to conventional energy minimization and simulated annealing with molecular dynamics (overall *B* factor), followed by iterative refinement of atomic *B* factors and positions, and manual rebuilding. At convergence of this process, refinement was concluded with PROLSQ (*19*).

Derivative*	Native	Native	CH_3HgCl	Hg-aniline	CH_3HgCl (A22C)	$(NH_3)_4$ Pt/ $PtCl_4$	*cis*-[Pt(methyl-$NH_2)_2Cl_2$]	$Pt(NH_3)_2Cl_2$	K_2[Pt (oxalate)$_2$]
Instruments	DIP100	PF	DIP100	FAST	PF	PF	PF	DIP100	DIP100
Resolution (Å)	2.8	2.5	2.8	3.0	2.5	2.5	2.5	2.8	2.8
Unique reflections (*N*)	13.867	29,264	12,779	10,579	16,384	14,756	15,566	11,171	12,518
(% complete)	(90.9)	(86.2)	(86.4)	(77.0)	(81.0)	(73.0)	(77.0)	(76.2)	(84.6)
R_{merge} (%)†	4.22	5.38	7.29	10.64	5.52	5.81	5.03	8.17	6.61
Number of heavy atom sites			1	1	2	3	2	3	4
R_{iso} (%)‡			18.0	26.0	17.3	22.5	19.2	34.8	17.8
MIR analysis (50 to 3.0 Å)									
Overall phasing power§			1.44	1.14	1.07	1.31	0.79	0.58	0.55
R_{cullis}‖			0.72	0.71	0.74	0.76	0.83	0.91	0:88
Resolution	50–16.1	16.1–9.91	9.91–7.16	7.16–5.61	5.61–4.61	4.61–3.91	3.91–3.39	3.39–3.00	Total
Mean figure of merit versus resolution	0.918	0.931	0.925	0.906	0.861	0.787	0.710	0.663	0.762
Refinement (6.0 to 2.5 Å)									
R factor (%)	18.5								
Reflections ($\|F\| > 2\sigma_F$)	13,867								
Number of atoms	3,975								
Solvent molecules	94								
rms bond length (Å)	0.016								
rms bond angle (°)	0.039								

*CH_3HgCl, crystals soaked in 0.5 mM CH_3HgCl for 2 days; Hg-aniline, crystals soaked in 1 mM Hg-aniline for 2 days; CH_3HgCl (A22C), crystals of A22C mutant soaked in 0.5 mM CH_3HgCl for 2 days; $(NH_3)_4Pt/PtCl_4$, crystals soaked in half-saturated $(NH_3)_4Pt/PtCl_4$ for 2 days; *cis*-[Pt(methyl-$NH_2)_2Cl_2$], crystals soaked in 2 mM cis-[Pt(methyl-$NH_2)_2Cl_2$] for 2 days; $Pt(NH_3)_2Cl_2$, crystals soaked in 1 mM $Pt(NH_3)_2Cl_2$ for 2 days; K_2[Pt(oxalate)$_2$], crystals soaked in 1 mM K_2[Pt(oxalate)$_2$] for 2 days. †$R_{merge} = \Sigma_h\Sigma_i|I_{hi} - \langle I_h\rangle|/\Sigma_h\Sigma_i|I_{hi}|$ where *h* indicates unique reflection indices, and *i* indicates symmetry equivalent indices. ‡$R_{iso} = 347|F_{PH} - F_P|/\Sigma|F_P|$, where $|F_P|$ and $|F_{PH}|$ refer to the measured structure factor amplitudes of the native and the derivative. §Phasing power = f_{rms}/E_{rms}, where $f_{rms} = [(\Sigma f_H^2)/n]^{1/2}$ and $E_{rms} = [\Sigma(F_{PH} - |F_P + f_H|^2/n]^{1/2}$. ‖$R_{cullis} = \Sigma||F_H| - (|F_{PH}| - |F_P|)|/\Sigma(|F_{PH}| - |F_P|)$ (only for centric reflections), where $|F_H|$ represents the calculated heavy atom structure factor.

1.6 percent. The final model yielded an R factor of 18.5 percent for 13,867 reflections with $|F| > 2\sigma_F$ between 6.0 and 2.5 Å resolution, and included 3975 protein atoms and 94 solvent molecules with individual isotropic temperature factors (Table 1). The final $2F_o - F_c$ map was of high quality (Fig. 1).

Architectures of the class-defining, subclass-defining, and specific domains. *Thermus thermophilus* GluRS is a bent cylinder, with overall dimensions of 80 by 40 by 30 Å (Fig. 2). The molecule consists of four domains, which are tandemly arranged to form an elongated and curved shape with an axial ratio of 3.4 to 1. The NH_2-terminal half, which is composed of domains 1 and 2, primarily folds into an α/β structure (12 α helices and 12 β strands) (Fig. 3A). Even at first glance, we see that these NH_2-terminal domains of *T. thermophilus* GluRS bear a striking resemblance to those of *Escherichia coli* GlnRS, whose crystal structure has been solved in complex with $tRNA^{Gln}$ (*4*) (Fig. 3A). In contrast, the GluRS COOH-terminal half, consisting of domains 3 and 4, exhibits an all α topology (nine α helices), which is fundamentally different from the all β topology of the COOH-terminal domains of GlnRS (Fig. 3A).

The NH_2-terminal domain (domain 1) consists of two peptide segments, residues 1 to 70 and 187 to 322 (Fig. 2). Most of this domain forms a typical Rossmann fold; the NH_2-terminal 70 residues comprise three β strands (from β1 to β3) and two α helices (αA and αB), and residues 187 to 237 form two β strands (β10 and β11) and two flanking α helices (αG and αH) (Fig. 3A). The putative ATP-binding motifs, His^{15}-Val^{16}-Gly^{17}-Thr^{18} and Lys^{243}-Ile^{244}-Ser^{245}-Lys^{246}, are located in loops between the β1 strand and the αA helix, and between the β11 strand and the αI helix, respectively (Fig. 3A). These two loops are close to each other (Fig. 3A) in a deep cleft of the NH_2-terminal half of GluRS (Fig. 2). A crystal-soaking experiment revealed that ATP is bound to these motifs (*20*). This Rossmann fold is followed by a subdomain (residues 238 to 322) folded into four α helices (from αI to αL) (Fig. 3A). All these structural features have been identified in other class I aaRS's (*2*–*4*), an indication that their NH_2-terminal domains have evolved from a common ancestor (*1*).

The secondary structural elements in this class-defining domain of *T. thermophilus* GluRS correspond remarkably well to their counterparts in *E. coli* GlnRS (Fig. 3A), whereas domain 1 of the GluRS exhibits two insertions (Ins Glu-1 and Ins Glu-3) and a deletion (Ins Gln-4) as compared to GlnRS (Fig. 3B). Except for these inserted and deleted sequences, the polypeptide segments of the class-defining domain fold into three conserved tertiary-structure blocks within which the Cα atoms superimpose well between the GluRS and GlnRS structures by least-squares fitting (Table 2). A striking similarity [the root-mean-square (rms) deviation of 1.19 Å, on average] is observed in their geometry, for as many as 158 residues. This structural similarity between GluRS and GlnRS is remarkable, in contrast to the similarity between GluRS and TyrRS (class I) from *Bacillus stearothermophilus* (*2*); among the 206 Cα atoms in the Rossmann fold, 96 Cα atoms can be superimposed, with a rms deviation of 2.18 Å. Thus, the close fitting of the class-defining domains between GluRS and GlnRS suggests that these two aaRS's have diverged from each other recently in the evolution of class I aaRS's.

The 116-residue polypeptide segment intervening in domain 1 (the Rossmann fold) folds into another structural domain (domain 2), which consists of two α helices (αD and αF) and one 3_{10} helix (αE), all of which are roughly arranged about an apparent internal threefold symmetry, and of a four-stranded antiparallel β sheet (β4, β5, β8, and β9) (Fig. 3A). An insertion com-

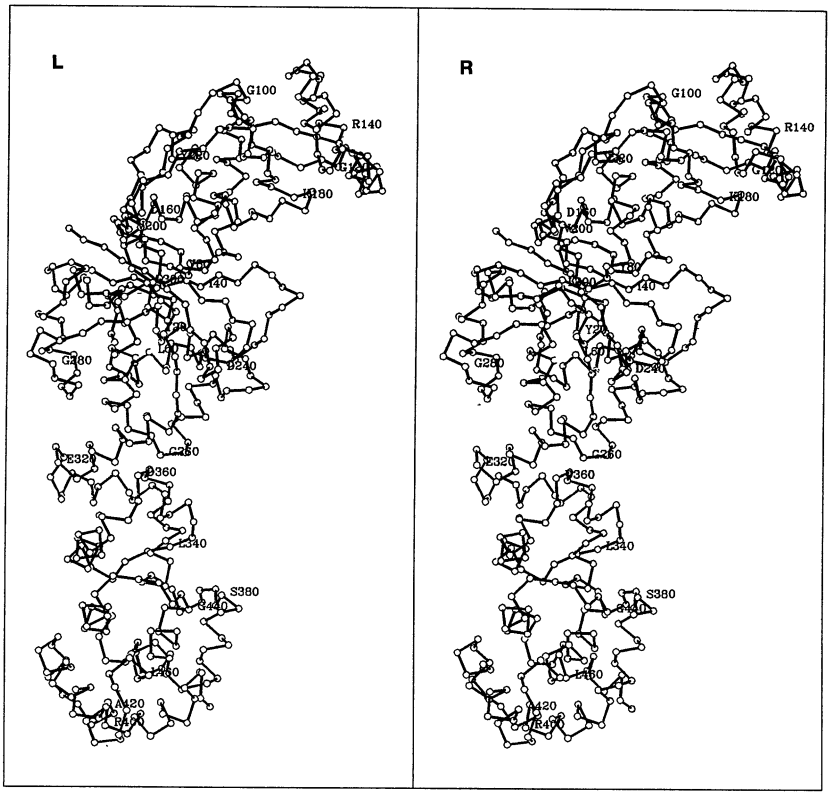

Fig. 2. A stereo drawing of the α-carbon backbone of *T. thermophilus* GluRS. The NH_2- and COOH-termini are labeled.

Table 2. Comparison of *T. thermophilus* GluRS structure with that of *E. coli* GlnRS. The superimposition was calculated by minimizing the distance between corresponding Cα positions for four segments as followed in the NH_2-terminal halves.

Domain	Residues in		Superimposed Cα atoms	Secondary structures	rms deviation (Å)
	GluRS	GlnRS			
1	1–68	26–93	68	α-A, -B β-1, -2, -3	1.06
2	85–107	104–124	23	α-C β-4	0.92
2	144–186*	170–210*	17	β-5, -8, -9	1.07
1	187–237†	211–262†	48	α-G, -H β-10, -11	1.07
1	252–319‡	292–336‡	42	α-I, -J, -L	1.53

*Except for residues 151 to 173 and 181 to 183 in GluRS and for residues 177 to 185 and 193 to 207 in GlnRS. †Except for residues 224 to 226 in GluRS and for residues 248 to 251 in GlnRS. ‡Except for residues 274 to 299 in GluRS and for residues 314 to 316 in GlnRS.

monly occurs in the Rossmann fold of class I aaRS's (*21*). However, the intervening sequences show great diversity in their lengths and sequences (*22*). The domain 2 structures of *T. thermophilus* GluRS and *E. coli* GlnRS are mostly conserved; 7 of the 10 secondary structural elements of GluRS correspond well to their counterparts in GlnRS (Fig. 3A). In the GluRS structure, however, one β strand is replaced by a 3_{10} helix (αE), and a long antiparallel β sheet (β6 and β7, Ins Glu-2) is inserted, in con-

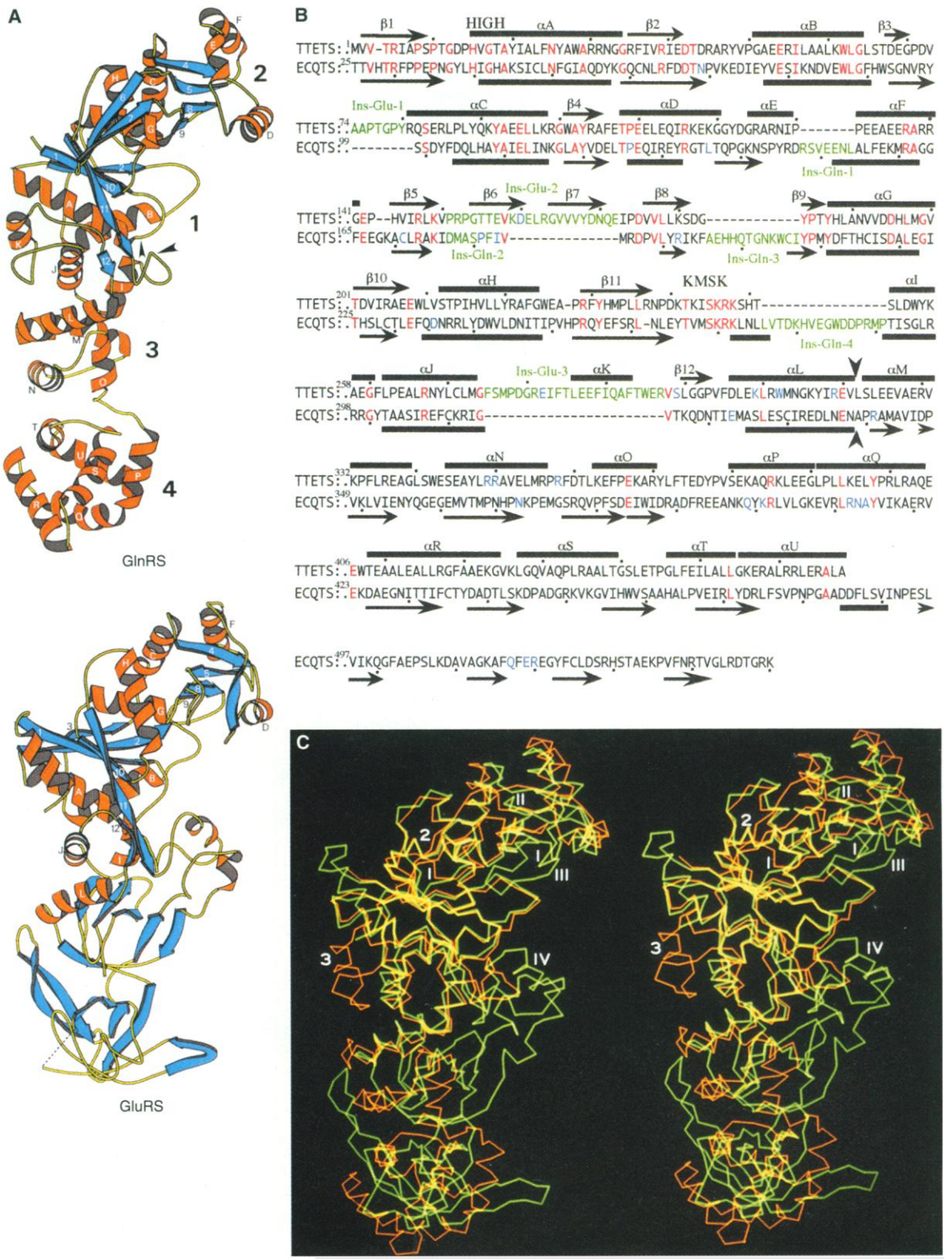

Fig. 3. (**A**) Ribbon diagram (*45*) displaying the overall folding of *T. thermophilus* GluRS and *E. coli* GlnRS (*4*). The α helices are shown in orange, the β strands are cyan, and the random coils are yellow. The GluRS structure consists of four tandemly arranged domains, labeled 1 through 4. The α helices (αA through αU) and β strands (β1 through β12) in GluRS are labeled. To highlight the striking similarity between GluRS and GlnRS in the NH_2-terminal halves, the corresponding secondary structures in GlnRS are identically marked. Locations of characteristic motifs for ATP-binding, HIGH and KMSK, are indicated by small and large arrowheads, respectively. (**B**) Primary sequence alignment of *T. thermophilus* GluRS (TTETS) (*13*) and *E. coli* GlnRS (ECQTS) (*46*), based on the structural alignment (Fig. 3C). Identical amino acids are colored in red. Secondary structural elements of each protein are indicated with thick bars for α helices and arrows for β sheets. Regions considered to be insertions (Ins Glu-1 through Ins Glu-3) and deletions (Ins Gln-1 through Ins Gln-4) in the GluRS structure are indicated in green. The boundaries between the NH_2- and COOH-terminal halves are indicated by arrowheads. Two specific ATP-binding motifs, HIGH and KMSK, are labeled. The amino acid residues of GluRS and GlnRS that specifically interact with their cognate tRNAs (*4*, *27*, and our data) are denoted by blue letters. (**C**) Superposition of *T. thermophilus* GluRS, in red, on *E. coli* GlnRS (*4*), in green. The two structures are manually superimposed to minimize the distance between the corresponding Cα positions for 198 residues in the NH_2-terminal half. Structural insertions in GluRS (Ins Glu-1 through Ins Glu-3B) and those in GlnRS (Ins Gln-1 through Ins Gln-4B) are labeled by Arabic and Roman numerals, respectively. This view is a similar orientation to that in (A).

trast to the GlnRS structure (Fig. 3, A and B). As for the conserved secondary structural elements of domain 2, two tertiary structural blocks, consisting of 40 residues, superimpose well upon the Cα atoms (rms deviation is 0.98 Å, on average) between the GluRS and GlnRS structures (Table 2). In contrast, the corresponding intervening domain of TyrRS folds in a completely different manner (2) from those of GluRS and GlnRS.

Superposition between GluRS and GlnRS is successful within each of the five blocks in NH_2-terminal half (Table 2). However, when we tried simultaneously to fit these five blocks of GluRS to their counterparts in GlnRS, slight but significant displacements remained, including the flanking region rich in α helices of domain 2 (from αD to αF) and the helical subdomain that occurs just after the Rossmann fold (Fig. 3C), both of which participate in tRNA recognition by *E. coli* GlnRS (4) and *T. thermophilus* GluRS, as described below. The GlnRS structure was chosen from the complex with $tRNA^{Gln}$, so that the displacements in the superimposed structures between GluRS and GlnRS are likely to be derived from changes in the relative orientations of the segments in GlnRS, which are induced by tRNA binding.

In contrast to the similarity in the NH_2-terminal halves, the all-α architecture of the GluRS COOH-terminal half is completely different from the β-barrel architecture of the GlnRS COOH-terminal half (Fig. 3, A and B). The three α helices (from αM to αO) of GluRS domain 3 (residues 323 to 370) form a structure similar to a helix bundle, which comes in contact with αL of the α-helical subdomain from the NH_2-terminal half (Fig. 3A). The COOH-terminal domain (domain 4, residues 371 to 468) consists of six α helices (from αP to αU), which fold into a hemispherical shape like an "α-helix cage" (Figs. 2 and 3A). This folding topology is distinctive, as compared with any other structures thus far determined. The two helices (αR and αS) form a helix-turn-helix structure similar to those found in repressor DNA-binding domains (23). For example, we could superimpose residues 415 to 436 of GluRS on residues 15 to 36 of lambda repressor (24) with an rms deviation of 0.93 Å, and residues 415 to 438 of GluRS on residues 67 to 90 of *trp* repressor (25) with an rms deviation of 0.88 Å. Twenty-one Leu residues appear at every third or fourth positions in the 98-residue sequence of domain 4 (Fig. 4). This arrangement aligns the Leu side chains along each helical axis and enhances the hydrophobic interhelical contacts (Fig. 4). This structural feature is distinct from the well-known leucine zipper motif (26); the slightly irregular occurrence of the Leu residues results in the formation of an extensive internal hydrophobic core surrounded by six helices (Fig. 4). In addition to these topological differences, the spatial arrangement of the third and fourth structural domains of GluRS differs from that of GlnRS (Fig. 3A). Thus, this particular contrast between GluRS and GlnRS indicates

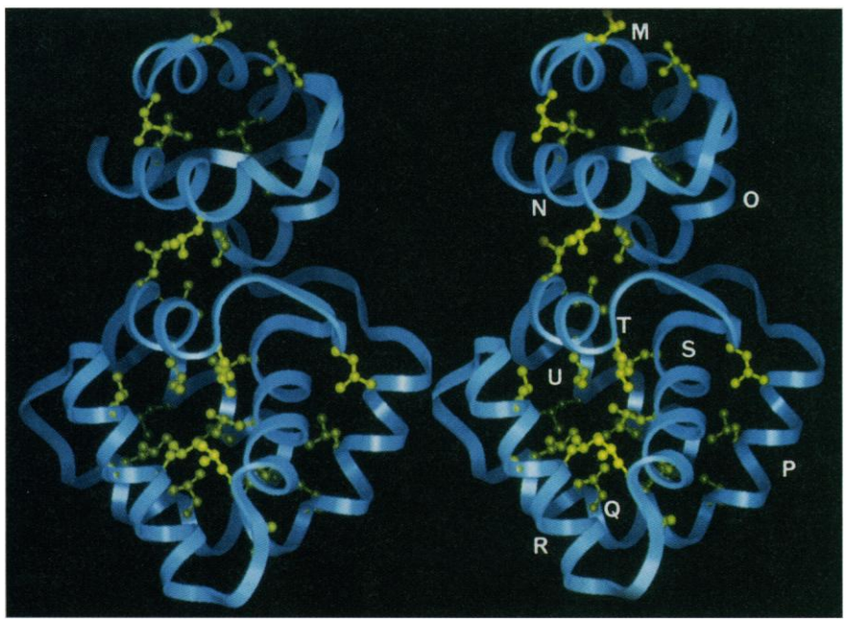

Fig. 4. Stereoview of the GluRS COOH-terminal half consisting of helix-bundle and an "α-helix cage." This view highlights the hydrophobic network of interhelical side chain interactions between leucines in the "α-helix cage" (domain 4). Leu side chains are colored yellow. The labelings of helices correspond to those in Fig. 3A.

Table 3. Catalytic properties of *T. thermophilus* GluRS mutants. Mutant *gltX* genes were generated by oligonucleotide-directed mutagenesis. Wild-type (WT) and mutant GluRS's were purified as described (13). The aminoacylation reactions were performed at 65°C in a buffer containing 50 mM Hepes (pH 7.5), 16 mM $MgCl_2$, 4 mM ATP, 130 μM [U-^{14}C]Glu, and 10 μM *E. coli* $tRNA^{Glu}$. *Thermus thermophilus* $tRNA^{Glu}$ shares 66 nucleotides with *E. coli* $tRNA^{Glu}$ including all the identity elements (43), and the Michaelis constants for *T. thermophilus* $tRNA^{Glu}$ and for *E. coli* $tRNA^{Glu}$ by *T. thermophilus* GluRS are identical (44). Kinetic parameters were calculated from Lineweaver-Burk plots. The concentration ranges for tRNA, Glu, and ATP were 2 to 20 μM, 20 to 200 μM, and 20 to 500 μM, respectively. The concentration of GluRS was optimized for each mutant in the range of 0.05 to 1 μM.

Variant	K_m (μM)			k_{cat} (s^{-1})	Loss of specificity*
	$tRNA^{Glu}$	L-Glu	ATP		
WT	2.73	120	23.0	2.39	1.00
Domain 2, Ins Glu-2					
D160A	172.4	81.5	41.7	0.659	229
Domain 1, Ins Glu-3					
S276A	24.7	129	46.1	0.945	22.9
E282A	422.4	166	72.3	1.06	365
S299A	2.70	127	58.1	0.00727	325
L300S	6.10	28.6	77.5	1.36	3.93
Domain 1, α-L					
K309Q	n.d.†	n.d.†	n.d.†	n.d.†	n.d.†
W312Y	21.0	8.00	65.4	1.87	9.83
W312C	3.43	131	132	0.0312	96.2
R319Q	40.7	83.8	36.2	3.13	11.4
Domain 3, α-N					
R349Q	59.1	53.3	27.5	1.28	40.4
R350Q	21.5	32.1	53.1	0.957	19.7
R358Q	27.5	103	112	3.03	7.95
Domain 4, helix(R)-turn-helix(S)					
R426Q	55.0	45.2	39.8	2.76	17.4

*Relative Km^{tRNA}/k_{cat}. †Not detectable.

that the two COOH-terminal halves evolved from different ancestors.

Structural elements for tRNA recognition. The detailed comparison of the tertiary structures of *T. thermophilus* GluRS and *E. coli* GlnRS (Fig. 3A) allowed a precise sequence alignment that highlights the structural differences between these two proteins, such as the sequence deletion and insertion (Ins), both in their NH_2- and COOH-terminal halves (Fig. 3B). The NH_2-terminal half of GluRS lacks four peptide motifs found in GlnRS (Fig. 3, B and C): residues 148 to 154 (Ins Gln-1 in domain 2) near the site of the replacement of the β strand with 3_{10} helix described above, residues 177 to 183 and 196 to 207 (Ins Gln-2 and Ins Gln-3, respectively, in domain 2), which form the protruding loops, and residues 276 to 291 (Ins Gln-4 in domain 1), which fold into a protruding α-helical structure. In contrast, *T. thermophilus* GluRS showed three characteristic motifs, Ins Glu-1, -2, and -3 (Fig. 3, B and C). Ins Glu-1, consisting of residues 74 to 80 (Fig. 3B), is looped out at the junction between domains 1 and 2 (Fig. 3A). Ins Glu-2 (residues 151 to 172) forms a long (27 Å) antiparallel β sheet (β6 and β7) and extends from domain 2 (Fig. 3A). Ins Glu-3 (residues 275 to 297) folds into an α helix (αK) and its accompanying loops, which protrude from domain 1 (Fig. 3A). The all-α architecture of the GluRS COOH-terminal half is totally different from the β-barrel architecture of the GlnRS COOH-terminal half (Fig. 3A).

In *E. coli* GlnRS, two of the four GlnRS-specific motifs and the COOH-terminal β-barrel domains participate in the recognition of the identity determinants in the acceptor stem and the anticodon loop of $tRNA^{Gln}$, respectively (*4*, *27*). Therefore, we examined whether the GluRS-specific structures, the three inserted motifs, and the COOH-terminal α-helical domains are also involved in tRNA recognition.

On the basis of the striking geometrical similarity between the NH_2-terminal halves of GluRS and GlnRS, we built a preliminary model of the complex of GluRS and $tRNA^{Glu}$ as follows. First, we converted the atomic coordinates of GluRS so as to superimpose well the α carbons of residues 1 to 319 of GluRS onto those of residues 26 to 336 of GlnRS in the complex with $tRNA^{Gln}$ (*4*). Then, we took the phosphate positions of the GlnRS-bound $tRNA^{Gln}$ (*4*) and placed them on the GluRS structure (Fig. 5A). In this process, the conformations of any parts of the tRNA were not changed. There is no serious clash between the tRNA and the GluRS in the model (Fig. 5A).

The crystal structure of the *E. coli* GlnRS·$tRNA^{Gln}$ complex (*4*, *27*) together with the results of biochemical studies (*28*) revealed that the terminal three base pairs (U1·A72, G2·C71, and G3·C70) of the $tRNA^{Gln}$ acceptor stem are specifically recognized by two protruding loops, Ins Gln-2 and Ins Gln-3, in the "subclass-defining" domain. However, in the GluRS structure, these loop motifs are missing (Fig. 3B). In the model of the GluRS·$tRNA^{Glu}$ complex (Fig. 5A), the antiparallel β sheet specific for GluRS (Ins Glu-2, Fig. 3B) is extended and comes into contact with the middle of the $tRNA^{Glu}$ acceptor stem. Therefore, in our study, we replaced Asp^{160} by Ala in Ins Glu-2. This GluRS variant (D160A) had a high Michaelis constant (K_m) for $tRNA^{Glu}$ in the aminoacylation reaction (Table 3). This result corresponds well to our chemical footprinting analysis of the *E. coli* $tRNA^{Glu}$-GluRS complex; phosphate groups around the middle of the $tRNA^{Glu}$ acceptor stem are in contact with GluRS (*29*). Thus, it is suggested that the β-strand insertion, Ins Glu-2, of GluRS recognizes identity determinants in the acceptor stem of $tRNA^{Glu}$.

In our model of the GluRS·$tRNA^{Glu}$ complex, the GluRS-specific 23-residue motif, including αK (Ins Glu-3 in Fig. 3B, and the red colored portion in Fig. 5A) is in contact with the middle of the D stem, and hence we replaced several amino acid residues in this motif by Ala to investigate the functions of their side chains. The mutation of Ser^{299}, which is in the vicinity of Ψ13 of the tRNA in the model (Fig. 5A), reduced the k_{cat} value (Table 3). In contrast, the mutation of Glu^{282}, which is close to A24 in the model (Fig. 5A), increased the K_m value for $tRNA^{Glu}$ (Table 3). Further kinetic experiments on GluRS revealed that Lys^{309} and Trp^{312}, which closely reside with Glu^{282} on the molecular surface (Fig. 1A), are crucial for GluRS activity (Table 3). Thus, these four residues, Glu^{282}, Ser^{299}, Lys^{309}, and Trp^{312}, are now shown to play a pivotal role in the specific recognition of the D arm of $tRNA^{Glu}$. The four residues are highly conserved among the GluRS's investigated thus far. These results correspond to our kinetic results of $tRNA^{Glu}$

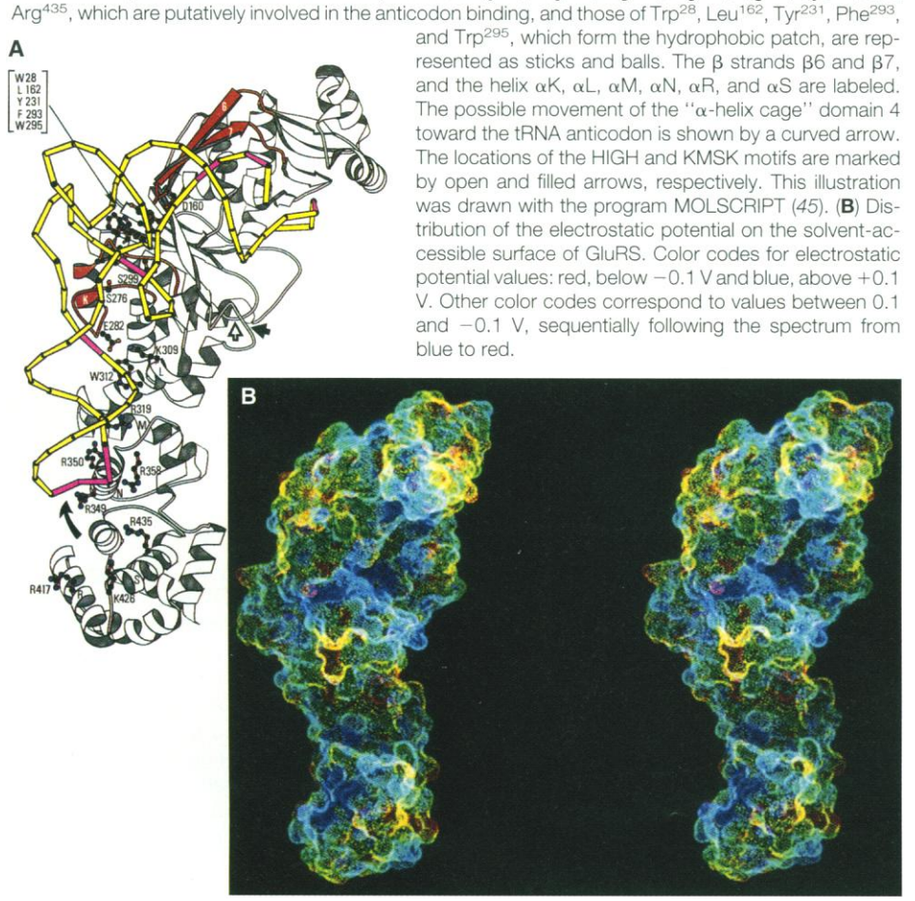

Fig. 5. (**A**) A model of the GluRS·$tRNA^{Glu}$ complex. The tRNA coordinates were obtained from the GlnRS·$tRNA^{Gln}$ complex (*4*). Nucleotides U2, C4, Ψ13, A24, U34, U35, C36, and G76 of $tRNA^{Glu}$ are colored magenta on the tRNA body (yellow). The helical (αK) and β-stranded (β6 and β7) structures specific to GluRS (Ins Glu-2 and -3) are colored in red. These two "GluRS-specific" motifs are linked by a hydrophobic patch. The side chains of Asp^{160}, Ser^{276}, Glu^{282}, Ser^{299}, Lys^{309}, and Trp^{312} for the binding with the acceptor and D stems of $tRNA^{Glu}$, those of Arg^{319}, Arg^{349}, Arg^{350}, Arg^{358}, Arg^{417}, Lys^{426}, and Arg^{435}, which are putatively involved in the anticodon binding, and those of Trp^{28}, Leu^{162}, Tyr^{231}, Phe^{293}, and Trp^{295}, which form the hydrophobic patch, are represented as sticks and balls. The β strands β6 and β7, and the helix αK, αL, αM, αN, αR, and αS are labeled. The possible movement of the "α-helix cage" domain 4 toward the tRNA anticodon is shown by a curved arrow. The locations of the HIGH and KMSK motifs are marked by open and filled arrows, respectively. This illustration was drawn with the program MOLSCRIPT (*45*). (**B**) Distribution of the electrostatic potential on the solvent-accessible surface of GluRS. Color codes for electrostatic potential values: red, below −0.1 V and blue, above +0.1 V. Other color codes correspond to values between 0.1 and −0.1 V, sequentially following the spectrum from blue to red.

variants: U11·A24, Ψ13·G22, and A21 in the D arm are strong identity determinants for *E. coli* $tRNA^{Glu}$ (*29*).

Genetic and biochemical studies (*30*) indicate that three nucleotides of the anticodon serve as the identity determinants of *E. coli* $tRNA^{Glu}$, as in the case of *E. coli* $tRNA^{Gln}$ (*28*). However, in that the structures of the GluRS COOH-terminal domains are totally distinct from those of GlnRS, the recognition mechanism for the anticodon must be different for each of the two enzymes. In the hypothetical model (Fig. 5A), the anticodon is located near the cleft between the two helical domains 3 and 4 of GluRS. The electrostatic potentials on the solvent-accessible surface (*31*) of GluRS indicate that positive charges are clustered around the cleft between the domains (Fig. 5B). These positive charges are derived from Arg^{319} on the αL of domain 1, and Arg^{349}, Arg^{350}, and Arg^{358} on the αN of domain 3 (Fig. 5A). Therefore, we produced GluRS variants in which these Arg residues are replaced by Gln, and we then examined whether their positive charges are essential for the $tRNA^{Glu}$ recognition. All these mutations resulted in substantial increases in the K_m values for $tRNA^{Glu}$ (Table 3), suggesting that these basic residues Arg^{319}, Arg^{349}, Arg^{350}, and Arg^{358} interact with the anticodon nucleotides (Fig. 5A). The positively charged pocket expands to the α-helix cage domain 4 (Fig. 5B). This positively charged cavity involves Arg^{417}, Lys^{426}, and Arg^{435} on the aforementioned helix-turn-helix structure (Fig. 5A). The mutation of Lys^{426} to Gln in the turn appreciably increased the K_m value for $tRNA^{Glu}$ (Table 3). However, in our model for the complex, domain 4 is too distant to interact with the anticodon of $tRNA^{Glu}$. Domain 4 is linked to domain 3 through a mobile hinge (Fig. 3A). Therefore, we assume that domain 4, with the "α-helix cage," tilts to enable the positively charged cavity to fit to the tRNA anticodon, as shown by a curved arrow in Fig. 5A.

Implication for GluRS and GlnRS evolution. The present structural and functional studies on *T. thermophilus* GluRS indicated that the motifs inserted into the NH_2-terminal class- and subclass-defining domains participate in the recognition of the D arm and the acceptor stem of $tRNA^{Glu}$, while the COOH-terminal α-helical domains have a role in the anticodon recognition. Then, in terms of our primary structure alignment on the basis of the secondary and tertiary structures, *T. thermophilus* GluRS was compared with six other GluRS's and three GlnRS's (Fig. 6). The prokaryotic GluRS's from *Rhizobium meliloti*, *E. coli*, *Bacillus subtilis*, and *B. stearothermophilus* show high sequence similarities with *T. thermophilus* GluRS; the three GluRS-specific motifs in the NH_2-terminal half and the α-helical architecture of the COOH-terminal half are conserved. GlnRS exists in Gram-negative bacteria such as *E. coli* and *R. meliloti* as well as in eukaryotic cytoplasm, while GlnRS is missing and GluRS glutamylates both $tRNA^{Glu}$ and $tRNA^{Gln}$ in various prokaryotes (Gram-positive bacteria, such as *B. subtilis* and *B. stearothermophilus*, and cyanobacteria), in prokaryotic organelles of eukaryotes (chloroplasts and mitochondria), and in archaebacteria or archaea (halobacteria) (*10, 11*). With regard to the question of how the GluRS's from the latter group can aminoacylate $tRNA^{Gln}$ as well as $tRNA^{Glu}$, it has been pointed out that the two *Bacillus* GluRS's have a 30 amino acid segment of a high homology with the anticodon binding region of *E. coli* GlnRS (*32*) (Fig. 6), which adopts a loop conformation (*4, 27*). The GlnRS-like 30 amino acid segment of the *Bacillus* GluRS's replaces a segment from the middle of αL to the beginning of αN of the *T. thermophilus* GluRS, and may therefore come close to the anticodon of tRNA as in our docking model (Fig. 5A). Thus, it is possible that this GlnRS-specific segment can function in the recognition of the $tRNA^{Gln}$ anticodon, both on the all-β and all-α frameworks.

In contrast, the eukaryotic GluRS's from human and *Drosophila* show higher sequence homologies to GlnRS's not only from human but also from Gram-negative bacteria, even as compared with the prokaryotic GluRS's (*12, 33,* and *34*). In fact, the present alignment based on the 3D structure indicates that the eukaryotic GluRS's have all the four GlnRS-specific insertions, but none of the GluRS-specific ones, in the NH_2-terminal half (Fig. 6). A question arises how the eukaryotic GluRS recognizes the cognate $tRNA^{Glu}$. A possible answer is that eukaryotic GluRS assigns much more contribution to the anticodon than to the D arm. We found that, between human GluRS and GlnRS, the sequence similarity in the COOH-terminal half (24 percent identity) is much lower than that in the NH_2-terminal half (42 percent identity), and that most of the similarities in the COOH-terminal half were observed corresponding to the connecting loops of the *E. coli* GlnRS β-barrel structure (*4, 27*). In contrast, the eukaryotic GluRS's also show slightly lower but considerable sequence homology with prokaryotic GluRS's. In particular, several Leu residues involved in hydrophobic interhelical contacts in our GluRS structure (Fig. 4) are conserved in the eukaryotic GluRS's. Therefore, it is possible that eukaryotic GluRS's adopt an α-helical architecture in the COOH-terminal half for recognition of the $tRNA^{Glu}$ anticodon. There-

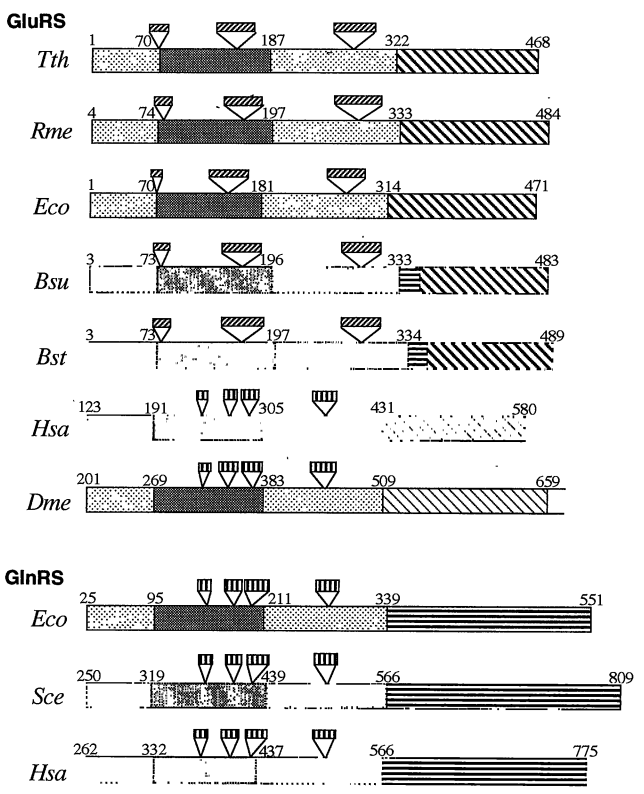

Fig. 6. Schematic alignment of GluRS and GlnRS sequences. For each enzyme, the large bar indicates the main framework, while the smaller bars indicate the inserted sequences. In the framework bar, the class-defining and subclass-defining domains are lightly and heavily shaded, respectively. Nonconserved anti-codon-binding domains of prokaryotic GluRS's (*13, 32, 47, 48*), eukaryotic GluRS's (*33, 34*), and GlnRS's (*12, 46, 49*) are hatched with thick oblique lines, thin oblique lines, and horizontal lines, respectively. Boundaries between respective domains are numbered. As for the inserted sequences, the "GluRS-specific" and "GlnRS-specific" motifs are hatched with oblique lines and vertical lines, respectively. Abbreviations are: *Tth*, *Thermus thermophilus*; *Rme*, *Rhizobium meliloti*; *Eco*, *E. coli*; *Bsu*, *B. subtilis*; *Bst*, *B. stearothermophilus*; *Hsa*, *Homo sapiens*; *Dme*, *Drosophila melanogaster*; and *Sce*, *Saccharomyces cerevisiae*.

fore, it would be important to determine the 3D structure of eukaryotic GluRS.

For the origins of GluRS and GlnRS, Mirande *et al.* have proposed a surprising hypothesis of horizontal gene transfer, namely, that the prokaryotic GlnRS has a eukaryotic origin and was acquired by genetic exchange between the two primary Kingdoms (*12*). According to this hypothesis, we describe here a possible scenario of the GluRS and GlnRS evolution from our 3D structural viewpoint. In the first place, it should be assumed that the original ancestral cells had the combination of GluRS and the transamidase. This still remains to be the case for many systems such as Gram-positive bacteria and mitochondria that arose from an endosymbiont of purple bacteria (the ancestor of Gram-negative bacteria) (*12*). Then, this prototypical GluRS could have evolved into prokaryotic and eukaryotic GluRS's and acquired the GluRS-specific and GlnRS-specific insertions, respectively, in their NH_2-terminal class- and subclass-defining domains. In the cytoplasm of the ancestral eukaryote, this GluRS further diverged into the present-day eukaryotic GluRS and GlnRS by gene duprication (*12*). The gene for the GlnRS could have been transferred laterally from eukaryote to prokaryote, resulting in the Gram-negative bacteria that have both GluRS and GlnRS (*12*). During this course, the GluRS lost its ability to aminoacylate $tRNA^{Gln}$ upon accumulation of mutations into the 30 amino acid segment for recognition of the $tRNA^{Gln}$ anticodon. Actually, the amino acid sequence of the segment of *T. thermophilus* GluRS (*13*) has lost nearly half of the residues conserved between *Bacillus* GluRS and *E. coli* GlnRS (*32*).

If we assume that the GlnRS and the bacterial GluRS have evolved from the common ancestor, their difference in the architecture of the COOH-terminal half appears to be drastic. It is possible that the prototypical GluRS consisting of the class- and subclass-defining NH_2-terminal half recruited the all-α domains in prokaryote, and independently the all-β domains in eukaryote, for better tRNA recognition. However, in the case that the present-day eukaryotic and archaebacterial GluRS's have the α-helical architecture for the COOH-terminal half, the total switch of architecture may have occurred in parallel to the derivation of GlnRS from the eukaryotic GluRS, for example, by accumulation of mutations or by an exon shuffling. As to the origins of such domains, the nonconserved anticodon-binding domains of GlnRS (class I) and AspRS (class II) share a β-barrel architecture (*4*, *6*). We therefore performed a 3D-profile database search (*35*) using the α-helix cage of domain 4 from *T. thermophilus* GluRS as the reference structure. A substantial structural similarity was detected between *T. thermophilus* GluRS of class I and *E. coli* AlaRS of class II; AlaRS carries the heptad leucine repeat in its sequence (*36*), which is presumed to form an α-helical domain, similar to domain 4 of the GluRS. The nonconserved domains might have arisen from several ancestors and have evolved independently of their catalytic domains defining the aaRS class and subclass.

REFERENCES AND NOTES

1. G. Eriani, M. Delarue, O. Poch, J. Gangloff, D. Moras, *Nature* **347**, 203 (1990).
2. P. Brick, T. N. Bhat, D. M. Blow, *J. Mol. Biol.* **208**, 83 (1989).
3. S. Brunie, C. Zelwer, J. L. Risler, *ibid.* **216**, 411 (1990).
4. M. A. Rould, J. J. Perona, D. Söll, T. A. Steitz, *Science* **246**, 1135 (1989).
5. S. Cusack *et al.*, *Nature* **347**, 249 (1990); M. Fujinaga *et al.*, *J. Mol. Biol.* **234**, 222 (1993); V. Biou, A. Yaremchuk, M. Tukalo, S. Cusack, *Science* **263**, 1404 (1994).
6. M. Ruff *et al.*, *Science* **252**, 1682 (1991); J. Cavarelli *et al.*, *Nature* **362**, 181 (1992); M. Delarue *et al.*, *EMBO J.* **13**, 3219 (1994).
7. Abbreviations for the amino acid residues are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q. Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.
8. P. Schimmel, R. Giegé, D. Moras, S. Yokoyama, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8763 (1993).
9. R. Breton *et al.*, *J. Biol. Chem.* **261**, 10610 (1986).
10. A. Schön, C. G. Kannangara, S. Gough, D. Söll, *Nature* **331**, 187 (1988).
11. M. Wilcox and M. Nirenberg, *Proc. Natl. Acad. Sci. U.S.A.* **61**, 229 (1968); B. N. White and S. T. Bayley, *Can. J. Biochem.* **50**, 600 (1972).
12. V. Lamour *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8670 (1994).
13. O. Nureki *et al.*, *Eur. J. Biochem.* **204**, 465 (1992).
14. _____, in preparation.
15. T. A. Jones, *J. Appl. Crystallogr.* **11**, 268 (1978).
16. B.-C. Wang, *Methods Enzymol.* **115**, 90 (1985).
17. K. Y. J. Zhang, *Acta Crystallogr.* **D49**, 213 (1993).
18. A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987).
19. W. A. Hendrickson, *Methods Enzymol.* **115**, 252 (1985).
20. O. Nureki *et al.*, in preparation.
21. R. M. Starzyk, T. A. Webster, P. Schimmel, *Science* **237**, 1614 (1987).
22. K. Shiba and P. Schimmel, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1880 (1992).
23. T. A. Steitz *et al.*, *ibid.* **79**, 3097 (1982); P. F. Johnson and S. L. McKnight, *Annu. Rev. Biochem.* **58**, 799 (1989); C. O. Pabo and R. T. Sauer, *ibid.* **61**, 1053 (1992); S. C. Harrison and A. K. Aggarwal, *ibid.* **59**, 933 (1990).
24. S. R. Jordan and C. O. Pabo, *Science* **242**, 893 (1988).
25. R. W. Schevitz, Z. Otowinowski, A. Joachimiak, C. L. Lawson, P. B. Sigler, *Nature* **317**, 782 (1985); Z. Otowinowski *et al.*, *ibid.* **335**, 321 (1988).
26. E. K. O'Shea, J. D. Klemm, P. S. Kim, T. Alber, *Science* **254**, 539 (1991); T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell* **71**, 1223 (1992).
27. M. A. Rould, J. J. Perona, T. A. Steitz, *Nature* **352**, 213 (1991).
28. M. Jahn, S. English, D. Söll, *ibid.*, p. 258.
29. S. Sekine *et al.*, in preparation.
30. L. A. Sylvers, K. C. Rogers, M. Shimizu, E. Ohtsuka, D. Söll, *Biochemistry* **32**, 3836 (1993).
31. H. Nakamura and S. Nishida, *J. Phys. Soc. Japan* **56**, 1609 (1987).
32. R. Breton, D. Watson, M. Yaguchi, J. Lapointe, *J. Biol. Chem.* **265**, 18248 (1990).
33. R. Fett and R. Knippers, *ibid.* **266**, 1448 (1991).
34. C. Cerini *et al.*, *EMBO J.* **10**, 4267 (1991).
35. J. U. Bowie, R. Lüthy, D. Eisenberg, *Science* **253**, 164 (1991).
36. S. D. Putney *et al.*, *ibid.* **213**, 1497 (1981); A. Lupas, M. Van Dyke, J. Stock, *ibid.* **252**, 1162 (1991).
37. I. Tanaka *et al.*, *J. Appl. Crystallogr.* **23**, 334 (1990).
38. A. Messerschmidt and J. W. Pflugrath, *ibid.* **20**, 436 (1987).
39. N. Sakabe, *Nuclear Instrum. Methods Phys. Res. A* **303**, 448 (1991).
40. T. Higashi, *J. Appl. Cryst.* **22**, 9 (1989).
41. W. Steigemann, thesis, Technische Universität, Munchen (1974).
42. CCP4, programme package Daresbury Laboratory, Warrington W4A 4AD, England.
43. M. Hara-Yokoyama, thesis, University of Tokyo, Tokyo (1985).
44. M. Hara-Yokoyama, S. Yokoyama, T. Miyazawa, *J. Biochem. (Tokyo)* **96**, 1599 (1984).
45. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946–950 (1991).
46. F. Yamao, H. Inokuchi, A. Cheung, H. Ozeki, D. Söll, *J. Biol. Chem.* **257**, 11639, (1982).
47. S. Laberge, Y. Gagnon, L. M. Bordeleau, J. Lapointe, *J. Bacteriol.* **171**, 3926 (1989).
48. R. Breton, H. Sanfacon, I. Papayannopoulos, K. Biemann, J. Lapointe, *J. Biol. Chem.* **261**, 10610 (1986).
49. S. W. Ludmerer and P. Schimmel, *ibid.* **262**, 10801 (1987).

50. We thank K. Y. J. Zhang and D. Eisenberg for providing the program SQUASH and the 3D profile search program; N. Sakabe, A. Nakagawa, and M. Konno for support in data collection at KEK, Japan; H. Nakamura for calculating the electrostatic potential; and K. Nishikawa for 3D-profile analysis. The coordinates have been deposited in the Brookhaven Protein Data Bank (tracking number: T5089). Supported in part by the grants-in-aid for Scientific Research on Priority Areas from the Ministry of Education, Science, and Culture of Japan (S.Y.) and a research grant from the International Human Frontier Science Program Organization (S.Y.)

15 August 1994; accepted 1 February 1995