A chimaeric glutamyl:glutaminyl-tRNA synthetase: implications for evolution

Rajesh SAHA*†, Saumya DASGUPTA*, Gautam BASU* and Siddhartha ROY‡[1]

*Department of Biophysics, Bose Institute, P 1/12 C.I.T. Scheme VII M, Calcutta 700054, India, †Dr. B.C. Guha Centre for Genetic Engineering and Biotechnology, Department of Biochemistry, University College of Science, University of Calcutta, 35 Ballygunge Circular Road, Calcutta 700019, India, and ‡Indian Institute of Chemical Biology, CSIR, 4, Raja S.C. Mullick Road, Kolkata 700032, India

aaRSs (aminoacyl-tRNA synthetases) are multi-domain proteins that have evolved by domain acquisition. The anti-codon binding domain was added to the more ancient catalytic domain during aaRS evolution. Unlike in eukaryotes, the anti-codon binding domains of GluRS (glutamyl-tRNA synthetase) and GlnRS (glutaminyl-tRNA synthetase) in bacteria are structurally distinct. This originates from the unique evolutionary history of GlnRSs. Starting from the catalytic domain, eukaryotic GluRS evolved by acquiring the archaea/eukaryote-specific anti-codon binding domain after branching away from the eubacteria family. Subsequently, eukaryotic GlnRS evolved from GluRS by gene duplication and horizontally transferred to bacteria. In order to study the properties of the putative ancestral GluRS in eukaryotes, formed immediately after acquiring the anti-codon binding domain, we have designed and constructed a chimaeric protein, cGluGlnRS, consisting of the catalytic domain, *Ec* GluRS (*Escherichia coli* GluRS), and the anti-codon binding domain of *Ec*GlnRS (*E. coli* GlnRS). In contrast to the isolated *Ec*N-GluRS, cGluGlnRS showed detectable activity of glutamylation of *E. coli* $tRNA^{glu}$ and was capable of complementing an *E. coli* ts (temperature-sensitive)-GluRS strain at non-permissive temperatures. Both cGluGlnRS and *Ec*N-GluRS were found to bind *E. coli* $tRNA^{glu}$ with native *Ec*GluRS-like affinity, suggesting that the anticodon-binding domain in cGluGlnRS enhances k_{cat} for glutamylation. This was further confirmed from similar experiments with a chimaera between *Ec*N-GluRS and the substrate-binding domain of *Ec*DnaK (*E. coli* DnaK). We also show that an extended loop, present in the anticodon-binding domains of GlnRSs, is absent in archaeal GluRS, suggesting that the loop was a later addition, generating additional anti-codon discrimination capability in GlnRS as it evolved from GluRS in eukaryotes.

Key words: chimaera, complementation, evolution, specificity, synthetase, tRNA.

INTRODUCTION

A major mechanism of generation of protein diversity appears to be by shuffling and acquisition of domains followed by mutational changes [1–4]. Although this principle appears to be well accepted, many unresolved questions remain. Chimaeric proteins that initially form from such recombination events are not found, as expected, in extant organisms because of further mutational changes. A better understanding of the nature of such chimaeric proteins, formed immediately after domain shuffling and/or acquisition, may offer a deeper appreciation of selective advantages of such events as well as biochemical mechanisms of further evolution. In addition, principles learnt from such studies may provide us with a better way to evolve proteins *in vitro*.

aaRSs (aminoacyl-tRNA synthetases) are multi-domain proteins [5–7]. It is believed that the catalytic, class-defining domain of aaRS may be the most ancient [8]. It was suggested that other domains, such as the anti-codon binding and the editing domains of aaRS, were added later to augment the specificity of aminoacylation for enhanced adaptability [5]. One such domain-acquisition event probably occurred during the evolution of GluRS (glutamyl-tRNA synthetase) and GlnRS (glutaminyl-tRNA synthetase) after divergence of the archaeal and eubacterial kingdoms from the last common universal ancestor [9]. One possible scenario for evolution of GlnRS and GluRS across all kingdoms of life has been outlined by Mirande and co-workers [9]. They have argued that GluRS in the last common universal ancestor consisted of only the class-defining catalytic domain. After eubacteria and archaea/eukaryotes split, the two lineages independently acquired two structurally distinct anti-codon binding domains: (1) the bacterial kingdom acquired the helical-cage anti-codon binding domains found in extant eubacterial GluRS; (2) archaea/eukaryotes acquired the β-barrel anti-codon binding domains found in all GlnRSs and in all known archaeal and eukaryotic GluRSs. The initial event towards the development of a GlnRS may have been the creation of a chimaera of the class-defining catalytic domain of GluRS and the β-barrel anti-codon binding domain in the archaea/eukaryote lineage. Although this putative chimaeric protein has overall domain structure similar to extant GlnRSs and GluRSs of archaeal/eukaryotic origin, the domain interfaces may be substantially different. The extant aaRSs have undergone further evolution and optimization of the domain interfaces as well as acquisition of additional structural elements, such as loops. Thus extant aaRSs are not good models for this putative evolutionary intermediate protein.

To recapitulate such a possible evolutionary process and to examine the properties of the putative chimaeric protein, a chimaeric protein (from now on called cGluGlnRS, for chimaeric glutamyl-glutaminyl tRNA synthetase, to distinguish it from other cases where the nomenclature GluGlnRS was used in the literature) was created by joining residues 1–314 of *Ec*GluRS (*Escherichia coli* GluRS) and residues 335–547 of *Ec*GlnRS (*E. coli*

Abbreviations used: aaRS, aminoacyl-tRNA synthetase; *Ec*DnaK, *Escherichia coli* DnaK; GluRS, glutamyl-tRNA synthetase; GlnRS, glutaminyl-tRNA synthetase; *Ec*GluRS, *Escherichia coli* GluRS; *Ec*GlnRS, *Escherichia coli* GlnRS; *Ec*N-GluRS, *Escherichia coli* N-terminal GluRS; *Ec*C-GlnRS, *Escherichia coli* C-terminal GlnRS; cGluGlnRS, chimaeric GluRS (*Ec*N-GluRS–*Ec*C-GlnRS); ts, temperature-sensitive; *Tth*GluRS, *Thermus thermophilus* GluRS.

[1] To whom correspondence should be addressed (email sidroykolkata@gmail.com).

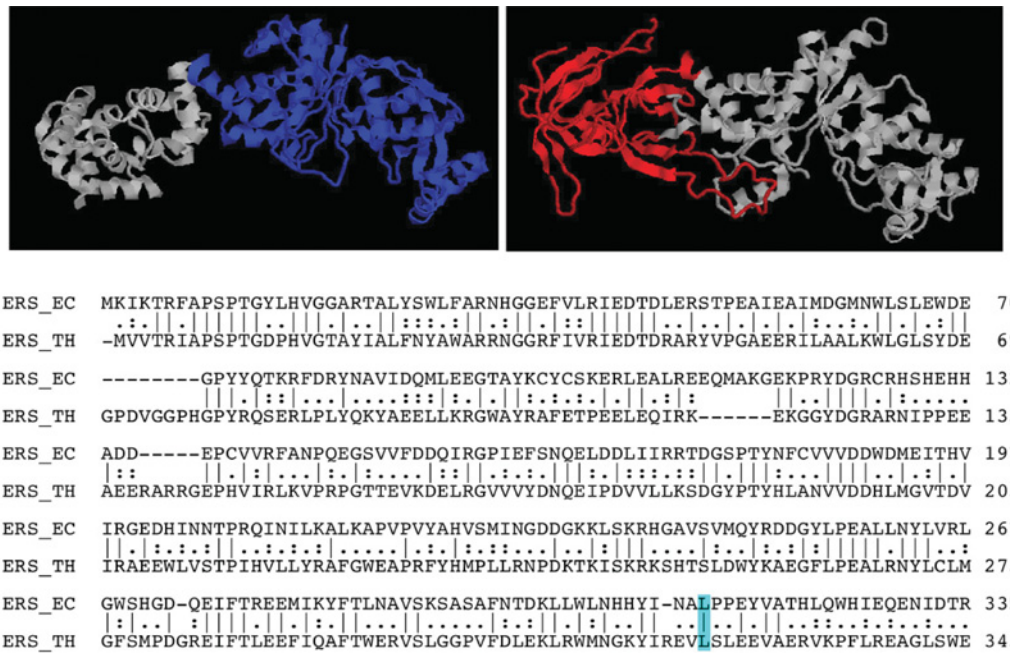


Figure 1 Crystal structures of *Tth*GluRS (top left; PDB code: 1J09) and *Ec*GlnRS (top right; PDB code: 1NYL)

The catalytic N-terminal domain of *Tth*GluRS is coloured blue and the C-terminal anti-codon binding domain of *Ec*GlnRS is coloured red. Pairwise sequence alignment of the catalytic domains of *Ec*GluRS (ERS_EC; P04805) and *Tth*GluRS (ERS_TH; 1J09) is shown in the bottom panel. The highlighted leucine residue is the extent of the catalytic domains.

GlnRS) with two extra amino acids in between. cGluGlnRS created this way has the same overall domain structure as that of the extant GlnRSs and GluRSs of archaeal/eukaryotic origin. However, the anti-codon/catalytic domain interface is likely to be non-optimized. A long loop that originates from the β-barrel anti-codon binding domain and interacts with the catalytic domain in GlnRS, may not interact with the catalytic domain of cGluGlnRS as the interaction interface is quite different (see below). Thus, cGluGlnRS may be a more realistic model for the putative evolutionary intermediate chimaeric protein. This article reports the properties of this chimaeric protein and its significance in the evolution of GlnRS from GluRS.

EXPERIMENTAL

Materials

E. coli $tRNA^{glu}$ was purchased from Sigma Chemical Company and dialysed extensively against sterile water before use. All other materials were of analytical grade.

Design and construction of cGluGlnRS

In the absence of an available crystal structure of *Ec*GluRS, residues spanning the catalytic domain of *Ec*GluRS were determined indirectly from a close sequence homologue with known structure, *Tth*GluRS (*Thermus thermophilus* GluRS). By visual inspection of the *Tth*GluRS crystal structure [10], we deduce that the N-terminal catalytic domain spans from residue 1 to 322 (Figure 1, top panel). Although the N-terminal half of GluRS or GlnRS may consist of more than one domain as defined by standard domain definition, we will define residues 1–322 in *Tth* GluRS or equivalent residues in other GluRSs or GlnRSs as the N-terminal domain. Similar nomenclature was used for the C-terminal half. The end residue of the N-terminal domain is 322 in *Tth*GluRS, which is equivalent to residue 314 in *Ec*GluRS (Figure 1, bottom panel; marked cyan). Thus, residues 1–314 of *Ec*GluRS were chosen as the N-terminal domain (*Ec*N-GluRS) to be incorporated in cGluGlnRS. In order to design the restriction sites needed to fuse the two fragments and keep mutations to a minimum, we decided to join residues 1–314 of *Ec*GluRS with residues 335–554 in *Ec*GlnRS (*Ec*C-GlnRS) with two additional residues in between [11]. The construction of the cGluGlnRS gene was performed by amplifying *Ec*GluRS (1–314) and *Ec*GlnRS (335–554) genes by PCR. The primers were chosen in a manner that HindIII sites occurs in-frame at the end and the beginning of the two respective fragments. The two amplified fragments, after appropriate restriction digestions, were first cloned separately in pUC18 (*Ec*N-GluRS in EcoRI and HindIII site and *Ec*C-GlnRS in HindIII and BamHI site respectively). Then the plasmids were purified, digested with appropriate restriction enzymes, gel purified, ligated through their HindIII sites and cloned in pBR322 to generate the cGluGlnRS gene. [cGluGlnRS carried a two amino acid insertion, underlined; C-terminal end of the *Ec*GluRS (1–314) in italics and N-terminal end of *Ec*GlnRS (335–554) in bold italics; *KLLWLNHHYINAL***KL***APRAMAVIDPV*]. The resultant *Ec*GluRS (1–314)-KL-*Ec*GlnRS (335–554) gene (cGluGlnRS) was sequenced and sub-cloned into the pET28a vector with an N-terminal His-tag. Similar protocols were followed for construction of *Ec*GluRS-DnaK. Site-directed mutagenesis was performed by the overlap extension procedure [12].

Purification of enzymes

pET28A plasmid containing the cGluGlnRS gene was transformed into either the *E. coli* strain BL21(DE3) or the ts JP1449 (DE3). From the transformed plate, a single colony was picked and allowed to grow in LB (Luria–Bertani) medium at 37 °C containing 35 μg/ml of kanamycin until it attained an attenuance of 0.2. For the JP1449(DE3) host strain, the initial growth temperature was 32 °C. Then the cells were transferred to 16 °C and kept there for approx. 30 min without shaking to allow

the cells to come to thermal equilibrium. After 30 min, 0.5 mM IPTG (isopropyl β-D-thiogalactoside) was added and shaking was resumed at 16 °C overnight. Cells were then harvested and the cell pellet was resuspended and sonicated in lysis buffer (50 mM potassium phosphate buffer, pH 7.5, containing 100 mM KCl, 1 mM 2-mercaptoethanol, 20 % glycerol, PMSF and 5 mM imidazole). The sonicated cells were spun at 21 400 ***g***, and the resultant supernatant was loaded slowly on to a Ni-NTA agarose resin equilibrated with the lysis buffer but without 2-mercaptoethanol. After the loading was complete, the column was washed with 10 column volumes of lysis buffer, followed by 5 column volumes of lysis buffer without 2-mercaptoethanol, but containing 20 mM imidazole. The protein was eluted by lysis buffer without 2-mercaptoethanol but containing 500 mM imidazole. The eluted protein was analysed on SDS/PAGE and the major fractions were pooled and then dialysed in 50 mM potassium phosphate buffer, pH 7.5, containing 100 mM KCl, 10 mM 2-mercaptoethanol and 20 % glycerol. Purification of *Ec*N-GluRS was performed in the same way. Purification of GluRS was performed as described before [13,14].

Fluorescence methods

Steady state fluorescence was measured in a Hitachi F3010 spectrofluorimeter. The experiments were performed at 25 °C. The excitation and emission bandpasses were 5 nm each, unless stated otherwise. Spectra of appropriate buffers were subtracted from the fluorescence spectra. Binding of *E. coli* $tRNA^{glu}$ to cGluGlnRS, *Ec*GluRS and *Ec*N-GluRS was performed in 20 mM Hepes buffer, pH 7.2, containing 20 % glycerol and 5 mM $MgCl_2$. The excitation wavelength was 295 nm and emission wavelength was 340 nm. Each point was performed separately as described before [15]. After transferring the protein from a stock solution to the cuvette, the fluorescence intensity was measured. A predetermined concentration of *E. coli* $tRNA^{glu}$ was then added and after 1 min the fluorescence was determined again. The inner filter effect correction was done using the formula:

$$F_{\text{corrected}} = F_{\text{observed}} \times [(A_{\text{ex}} + A_{\text{em}})/2].$$

where F is the fluorescence intensity and A is the absorbance. The ratios of corrected fluorescence intensity at 340 nm were used to measure the quenching.

Determination of enzyme activity *in vitro*

In all *in vitro* assays, enzymes were prepared from strain JP1449 (DE3) bearing appropriate plasmids. The enzyme assays were performed as described before, except the assay temperature was 42 °C to remove the contribution of endogenous *Ec*GluRS [16]. In the assays, the concentrations of cGluGlnRS and *Ec*N-GluRS were each kept at 1 μg per assay point, whereas *Ec*GluRS was kept at 1 ng per point.

Determination of K_m of cGluGlnRS with respect to *E. coli* $tRNA^{glu}$

The assay of cGluGlnRS was performed as described by Bhattacharyya et al. [16]. The changes that were made to calculate the K_m were as follows: the enzyme concentration was kept at 1.2 μg per assay point; the concentration of L-glutamate was kept at 500 μM; the concentration of *E. coli* $tRNA^{glu}$ was increased from 0 to 10 μM; the slope obtained in each case was plotted against the respective *E. coli* $tRNA^{glu}$ concentration and the required K_m was obtained by curve fitting. The assay temperature was 37 °C.

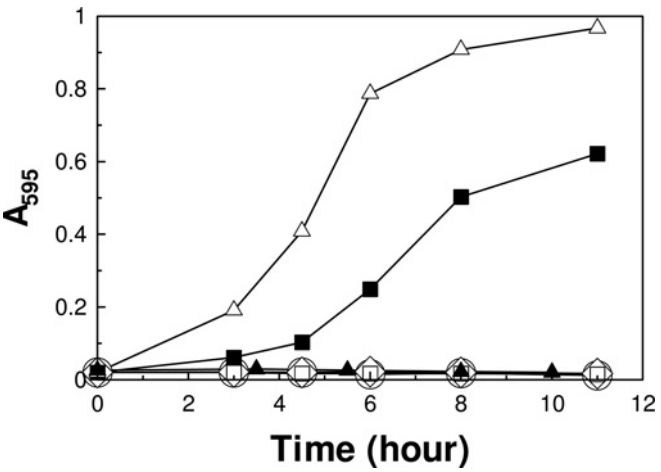


Figure 2 ***In vivo*** **complementation of a ts-GluRS mutant strain**

JP1449 (DE3) was transformed at 42 °C with different plasmids harbouring genes for different proteins and growth curves were determined at non-permissive 42 °C. △, native *Ec*GluRS; ■, cGluGlnRS; ○, pET28a; ◇, H16AcGluGlnRS; □, *Ec*N-GluRS; ▲, GluRS-DnaK chimaera. Plasmids encoding genes for *Ec*GluRS, cGluGlnRS, H16AcGluGlnRS, *Ec*N-GluRS, pET28A and GluRS-DnaK chimaera were transformed in JP1449 (DE3) at 32 °C. Small cultures were grown at 32 °C from colonies obtained from the transformed plates. Overnight cultures were diluted into fresh media in such a manner as to keep the initial attenuance the same. Then all of them were allowed to grow at 42 °C and their growth was monitored at different time intervals.

In vivo complementation assay

pET28A plasmids encoding genes for *Ec*GluRS, cGluGlnRS, H16AcGluGlnRS, *Ec*GluRS-DnaK or *Ec*N-GluRS were transformed into JP1449 (DE3) at 32 °C. From the respective transformed plates, overnight small cultures were grown at 32 °C. Overnight cultures were diluted into fresh media in such a manner as to keep the initial attenuance the same. Then all of the cultures were allowed to grow at 42 °C and their growth was monitored at different time intervals. All the growth curves were performed in media containing the appropriate antibiotic.

Sequence alignments

Pair-wise and multiple sequence alignments were carried out using web-based programs Needle (http://www.ebi.ac.uk/emboss/align/) and ClustalW (http://www.ebi.ac.uk/clustalw/) respectively.

RESULTS

One of the ways to judge as to how effective a protein is for the survival of an organism is through complementation assays. Whether cGluGlnRS is capable of effectively transferring L-glutamate *in vivo* to *E. coli* $tRNA^{glu}$ was judged by *in vivo* complementation assay. The host strain JP1449 (DE3), carrying a ts (temperature-sensitive) mutation in the chromosomal copy of the *Ec*GluRS, does not grow at 42 °C [17]. pET28a vectors carrying different genes were used to transform the strain and the transformed strains were grown at 42 °C. Figure 2 shows the growth curve of JP1449 (DE3) transformed with plasmids bearing different genes at the non-permissive temperature. JP1449 (DE3) transformed with plasmid bearing the wild-type *Ec*GluRS gene grew normally, whereas JP1449 (DE3), complemented with the gene for cGluGlnRS, grew slower but reasonably well. pET28a bearing JP1449 (DE3) did not grow at all. The support of growth in the case of cGluGlnRS is not likely to occur through chromosome-plasmid recombination as the ts mutation is at

residue 351 of *Ec*GluRS, absent in all cGluGlnRS constructs [18] (http://cgsc2.biology.yale.edu/Strain.php?ID = 6992).

In order to make sure that the growth of JP1449 (DE3), complemented by the cGluGlnRS gene, is due to the glutamylation activity of cGluGlnRS, mutations in the conserved HIGH signature sequence were introduced in the cGluGlnRS gene (H16A mutation) [19]. The H16AcGluGlnRS gene was incapable of complementing the ts-*Ec*GluRS at the non-permissive temperature, indicating that glutamylation activity of cGluGlnRS was the actual source of growth at the non-permissive temperature. In order to see if the C-terminal anti-codon binding domain played any role in the glutamylation activity, a plasmid bearing the N-terminal domain of *Ec*GluRS (1–314) gene, *Ec*N-GluRS, was introduced in JP1449 (DE3). This transformed strain was also unable to grow at the non-permissive temperature indicating that the C-terminal anti-codon binding domain of *Ec*GlnRS plays a crucial role in augmenting the glutamylation activity of cGluGlnRS. One possible role of the anti-codon binding domain may be a general conformational stabilization of the N-terminal domain bearing the active site. To find out whether the role of the C-terminal anti-codon binding domain is to interact specifically with the anti-codon loop of *E. coli* $tRNA^{glu}$ or to induce mere non-specific conformational stabilization, we have created another chimaeric protein in which the N-terminal domain of *Ec*GluRS (1–314) was fused with the substrate-binding domain of *Ec*DnaK (*E. coli* DnaK) (384–638) (a domain of similar size; the construction was similar to cGluGlnRS and contained a KL residue between the two domains). The plasmid bearing the gene of the resultant *Ec*GluRS (1–314)-KL-*Ec*DnaK (384–638) chimaera (from now on called GluRS-DnaK) was transformed into JP1449 (DE3) and a transformant colony was grown at the non-permissive temperature. The GluRS-DnaK was also unable to support growth at 42 °C, indicating that the role of the C-terminal domain of *Ec*GlnRS in the cGluGlnRS is specific for this domain. The anti-codon of *E. coli* $tRNA^{glu}$ differs by a single base from *E. coli* $tRNA^{gln2}$ at position 36. Substitution of *E. coli* $tRNA^{glu}$ at this position reduces k_{cat} and modestly increases K_m for *Ec*GlnRS, but retains decent overall activity [20]. This may indicate that the β-barrel domain may be capable of recognizing the anti-codon loop of *E. coli* $tRNA^{glu}$ in the context of cGluGlnRS and may interact with the anti-codon nucleotides of *E. coli* $tRNA^{glu}$.

In order to better understand the role played by the C-terminal anti-codon binding domain in cGluGlnRS, we have purified cGluGlnRS and the corresponding control protein *Ec*N-GluRS (cGluGlnRS without the C-terminal domain) and determined their catalytic and substrate-binding properties. The His-tagged cGluGlnRS from JP1449 (DE3) was purified by Ni-NTA affinity chromatography to homogeneity. The tertiary structure of the protein was evaluated by fluorescence spectroscopy. Figure 3 shows the fluorescence emission spectrum of cGluGlnRS. The observed emission maximum was 339.4 nm, indicating solvent-buried tryptophan side chains characteristic of a folded protein. The emission maximum was very close to that of wild-type *Ec*GluRS, which was found to be 337 nm. For comparison, the emission maximum of *Ec*N-GluRS was measured, and remained unchanged at 340 nm. The CD spectra of the *Ec*GluRS and *Ec*N-GluRS were also typical of proteins with large α-helical content, characterized by two distinct minima around 208 and 222 nm (results not shown).

cGluGlnRS, purified from JP1449 (DE3) background, was assayed for *E. coli* $tRNA^{glu}$ glutamylation activity at 42 °C to avoid any possible contamination from wild-type GluRS activity. Figure 4(A) shows the glutamylation activity of the cGluGlnRS, *Ec*GluRS and *Ec*N-GluRS. Clearly, cGluGlnRS has

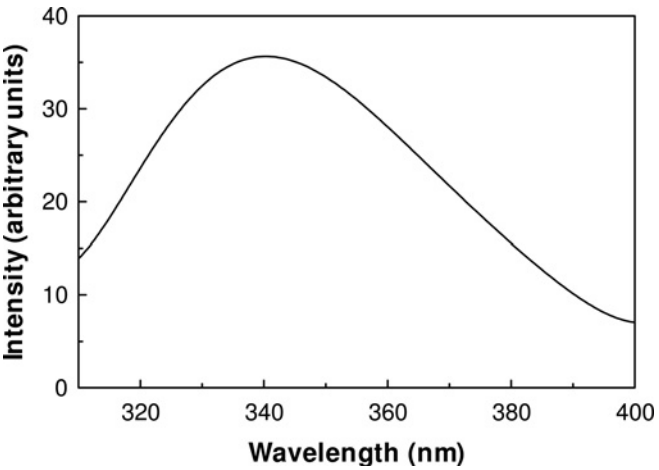


Figure 3 Fluorescence emission spectrum of cGluGlnRS

The excitation wavelength was 295 nm. The buffer was 50 mM Tris/HCl, pH 7.8, containing 100 mM KCl, 10 mM 2-mercaptoethanol and 20 % glycerol. Temperature was 25 ± 1 °C.

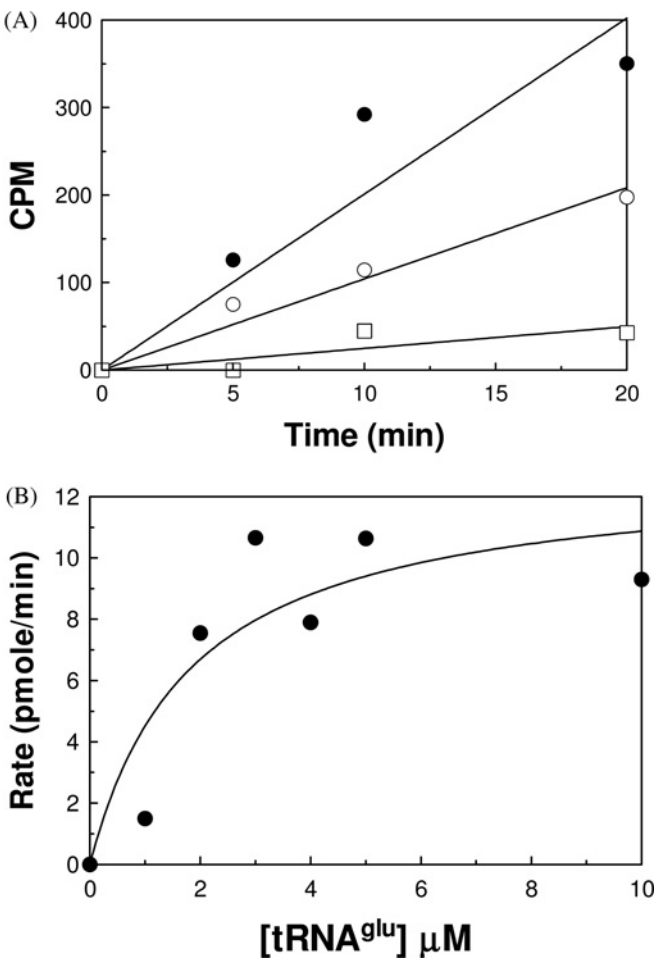


Figure 4 *In vitro* aminoacylation assay at 42 °C

(**A**) Aminoacylation assay of cGluGlnRS (●), *Ec*N-GluRS (□) and *Ec*GluRS (○) performed with *E. coli* $tRNA^{glu}$ and L-glutamate at 42 °C. For cGluGlnRS and *Ec*N-GluRS, the amount of enzyme added per assay point was 1 μg. In the case of *Ec*GluRS, the amount of protein added per point was 1 ng. (**B**) Determination of K_m of cGluGlnRS. cGluGlnRS (1.2 μg per point) was assayed with increasing concentrations of *E. coli* $tRNA^{glu}$. Aminoacylation assays are described in Bhattacharyya et al. [16].

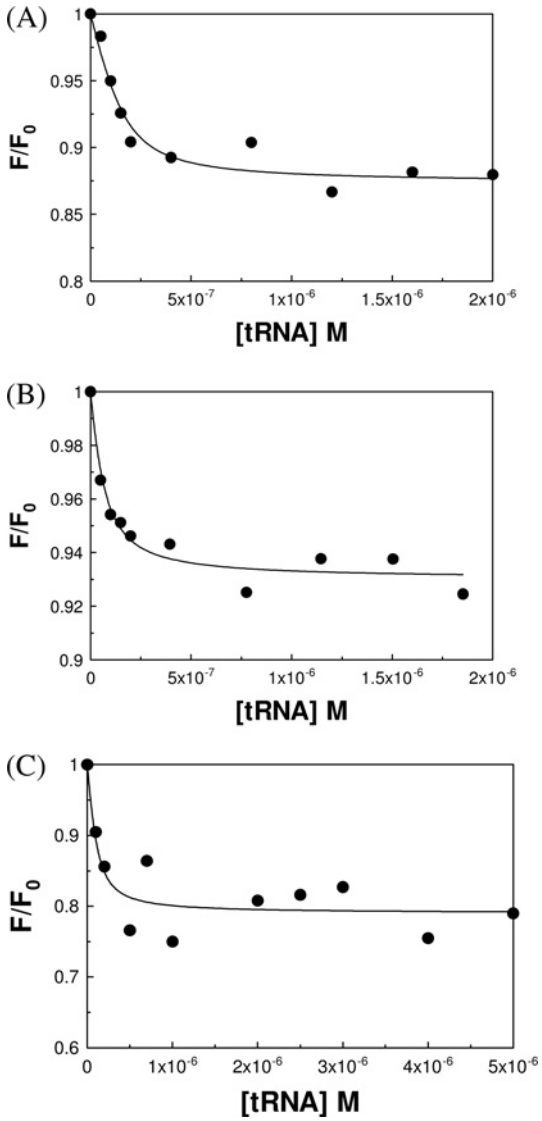


Figure 5 Titration of (A) *Ec*N-GluRS, (B) *Ec*GluRS and (C) cGluGlnRS with *E. coli* tRNA^{glu}

The buffer used is 20 mM Hepes, pH 7.2, containing 5 mM $MgCl_2$ and 20 % glycerol. The fits were done to a single site binding equation with protein concentration kept as a variable parameter.

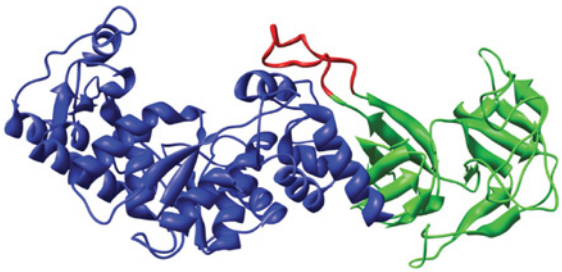
significant activity and *Ec*N-GluRS has negligible activity. The K_m and k_{cat} values were determined by activity assays at different tRNA^{glu} concentrations. The plot of rate against *E. coli* tRNA^{glu} concentration is shown in Figure 4(B). The plot yields a K_m of 1.86 μM and a k_{cat} value of 0.35 min^{-1}. The k_{cat} value of cGluGlnRS is approx. 500-fold lower than GluRS, whereas the K_m value is only moderately higher at the same solution conditions.

A substantially increased activity of cGluGlnRS, when compared with that of *Ec*N-GluRS, which only showed a trace amount of activity at best, suggests that the C-terminal domain plays an important role in enhancing the activity of the N-terminal domain that contains the catalytic site. Since this effect is not due to stabilization of the N-terminal domain (as inferred from lack of activity and complementation of GluRS-DnaK chimaera), it may originate either from an alteration of substrate binding or from an alteration of k_{cat}. Some idea may be obtained by studying the *E. coli* tRNA^{glu} binding to cGluGlnRS and *Ec*N-GluRS. Figure 5 shows binding isotherms of *E. coli* tRNA^{glu} to cGluGlnRS, *Ec*GluRS and *Ec*N-GluRS obtained by quenching of tryptophan fluorescence [21]. The isotherms were fitted to a single site binding equation to obtain the dissociation constants. The values obtained are 42 nM, 49 nM and 49.1 nM for *Ec*GluRS, cGluGlnRS and *Ec*N-GluRS respectively. The dissociation constant values are similar, indicating little additional apparent binding energy provided by the C-terminal β-barrel domain. It is likely that the additional binding energy may have been used to cause a conformational change resulting in the enhanced k_{cat} [22]. The difference between the K_d and K_m values are most probably due to higher ionic strength of the assay mixture.

DISCUSSION

It is generally believed that all GlnRSs evolved from an ancestral GluRS gene [23,24]. One possible clue as to how such an event took place lies in the fact that the anti-codon binding domains of bacterial GluRS and GlnRS are structurally completely different: an α-cage in GluRS and a β-barrel in GlnRS [11,25]. Homology studies of C-terminal domains of extant bacterial GluRS and GlnRS from different organisms led to the conclusion that archaeal and eukaryotic C-terminal domains of GluRS are structurally similar to the β-barrel domain of GlnRS [9,23], whereas bacterial GluRS anti-codon domains are all α-helical, similar to that seen in the *Tth*GluRS crystal structure [10]. Based on this observation, Mirande and co-workers have proposed that, before the divergence of bacteria and archaea/eukaryotes, GluRS consisted of only the class-defining catalytic domain that also glutamylated tRNA^{gln} and was able to incorporate glutamine through the transamidation pathway [9]. After the divergence of eubacteria and archaea/eukaryotes, the former acquired the α-helical anti-codon binding domain, whereas the latter acquired the β-barrel domain. Subsequent gene duplication in eukaryotes gave rise to GlnRS from GluRS. The GlnRS gene was then transmitted to certain species of the bacterial kingdom [23].

An important question in this regard is how GlnRS evolved in the eukaryotic branch and not in any other kingdoms of life. An interesting difference between the archaeal GluRS and eukaryotic GluRS and GlnRS appears to be the presence of the long loop (474–495 in *Ec*GlnRS) that extends from the β-barrel anti-codon binding domain to the class-defining catalytic domain (Figure 6). This loop appears to be crucial for specificity of this synthetase [20]. We have attempted to find out the existence of this loop in different kingdoms of life by multiple sequence alignment of GlnRS and GluRS from different sources. We examined 28 archaeal GluRS sequences and found the loop to be absent in all (Figure 6). In contrast, bacterial and eukaryotic GlnRS, represented by *Ec*GlnRS and a eukaryotic GluRS (*Saccharomyces cerevisiae)*, have this loop (Figure 6). In fact, mutagenesis of this loop resulted in somewhat relaxed anti-codon specificity in *Ec*GlnRS [26]. It may be postulated (Figure 7) that insertion of this loop occurred in the eukaryotic branch after divergence of eukaryotes and archaea and may have played a crucial role towards the development of sufficient anti-codon discrimination ability to develop GluRS and GlnRS [23]. We created cGluGlnRS to study the properties of the putative chimaeric protein produced by the postulated recombination with the isolated N-terminal class-defining domain in the early eukaryote/archaea branch. This early chimaeric enzyme may have either lacked the loop, spanning the anti-codon binding domain and the class-defining domain containing the active site, or was characterized by a reduced coupling between the N- and



```
Q_ECOLI   AHALPVEIRLYDRLFSVPNPGAADDFLSVINPESLVIKQGFAEPS 507
E_YEAST   KDVVPVDLVDFDHLITKDRLEEDESFEDFLTPQTEFHTDAIADLN 660
E_METJA   KDSVKVKVLMPD----------------------GEIKEGFAEKD 518
E_METM5   KDSVPVTVIDEN----------------------AEKIEGFAEKD 520
E_METTH   ESALEAEVIMPD----------------------ASRVRGVIEAD 519
E_METKA   DQAVRCEVVRPD----------------------GSVESGYAEIN 536
E_PYRFU   GKECEVIIPEGE-E--------------------LIIRKGLLEKN 538
E_SULSO   NEKVPVMVEKAERDE-------------------IKMINGYAEKI 540
E_THEVO   N-SREFQVLKPD----------------------GSIDKGLIEPA 538
E_HALMA   DGPALRLRTMDG-----------------------DVSGVAEPG 541
E_ECOLI   KVGMPLRVAVTG-----------------------AGQSPALDVT 444
```

Figure 6 Crystal structure of *E. coli* GlnRS and multiple sequence alignment of GluRSs

Top, crystal structure of *E. coli* GlnRS (1NYL) with the extended anti-codon binding domain loop (residues 474–495) coloured red. Bottom, multiple sequence alignment of *Saccharomyces cerevisiae* GluRS (E_YEAST; P46655) and representative archaeal GluRS sequences: *Methanococcus jannaschii* (E_METJA; Q58772), *Methanococcus maripaludis* (E_METM5; A4FXG8), *Methanobacterium thermoautotrophicum* (E_METTH; O26157), *Methanopyrus kandleri* (E_METKA; Q8TXB7), *Pyrococcus furiosus* (E_PYRFU; Q8U064), *Sulfolobus solfataricus* (E_SULSO; P95968), *Thermoplasma volcanium* (E_THEVO; Q979Q0), *Halobacterium marismortui* (E_HALMA; Q5V5N9), *E. coli* (E_ECOLI; P04805) and in the region of the *E. coli* GlnRS (Q_ECOLI; P00962) loop (474–495).

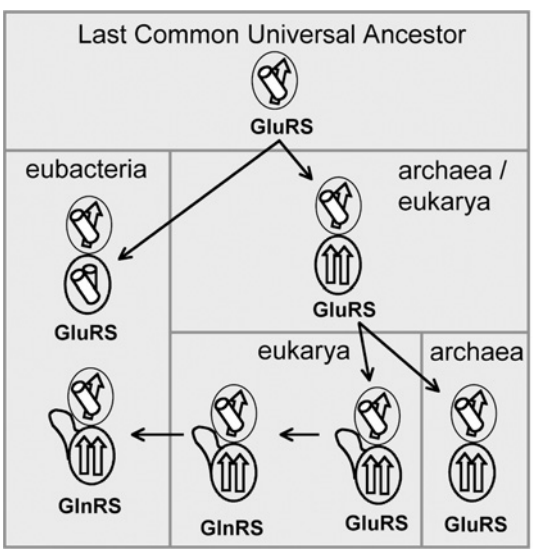


Figure 7 A schematic representation of the modular evolution of the Glx-tRNA synthetase family [9]

The β-barrel anti-codon binding domain loop (see Figure 6) is shown to have been acquired after the eukaryote/archaea branching.

the C-terminal domains. Although cGluGlnRS has this loop, the residues in *Ec*GlnRS that interact with the loop (residues 262, 264–267, 269, 272, 275–276, 279, 281 and 289) are either absent or mutated in *Ec*GluRS (based on alignment; results not shown). This suggests that the loop in cGluGlnRS may not interact or interacts weakly with the N-terminal class-defining domain. Thus, cGluGlnRS may represent the early chimaeric GluRS in the archaea/eukaryote branch after the separation of the bacterial kingdom and before the development of GlnRS. The chimaeric enzyme constructed here thus structurally resembles the putative GluRS of early archaea/eukaryotes.

The N-terminal domain of *Ec*GluRS, *Ec*N-GluRS, showed negligible activity at best; insufficient to complement a ts defect in the *Ec*GluRS gene at non-permissive temperatures. In contrast, cGluGlnRS is at least several-fold more active *in vitro* and was able to complement a ts defect in the *Ec*GluRS gene at non-permissive temperatures. Clearly, the presence of the β-barrel domain increases the catalytic activity of the active site of *Ec*GluRS (in the context of cGluGlnRS) situated several tens of angstroms away. For *Ec*N-GluRS, the binding affinity of *E. coli* $tRNA^{glu}$ was almost *Ec*GluRS-like, indicating that the anti-codon binding domain in EcGluRS might primarily contribute to the enhancement of k_{cat}. Similarly, the *E. coli* $tRNA^{glu}$ binding affinity to cGluGlnRS is also almost *Ec*GluRS-like. As stated before, due to similarity of the anti-codon binding loops of the *E. coli* $tRNA^{glu}$ and $tRNA^{gln}$, it is likely that the β-barrel domain, in the context of cGluGlnRS, interacts with the anti-codon loop of *E. coli* $tRNA^{glu}$. Thus, the additional postulated interaction energy of the β-barrel domain in the context of cGluGlnRS may be used to cause a conformational change that increases the k_{cat}. Taken together, all of this evidence suggests that the addition of a β-barrel anti-codon binding domain may have created an enzyme with a better k_{cat}. A question then is how this anti-codon binding information is transmitted to the distant active site in the context of cGluGlnRS?

It has been shown that mutation of identity elements in the anti-codon loop of *E. coli* $tRNA^{glu}$ predominantly reduces k_{cat} in *Ec*GluRS [27]. This suggests existence of pathways that transmit this binding information to the active site. Similar, but more detailed, conclusions have been drawn for *Ec*GlnRS. Recent work on *Ec*GlnRS suggests that there may be two pathways for the transmission of the correct anti-codon binding information to the active site. One involves the residues that are responsible for interacting with the inside of the L-shaped tRNA [28,29]. The other involves the loop (474–495 in *Ec*GlnRS) mentioned above [30,31]. However, these extant enzymes are evolutionarily optimized, whereas cGluGlnRS is not. Recognition of correct tRNA itself is a complex process involving many identity elements and coupled conformational changes, neither of which is well understood. In several cases it has been demonstrated that the binding of a cognate tRNA leads to the ordering of the active site, including the associated water molecules [32]. As far as it is known today, protein-mediated pathways play important roles in transmitting the effect of distant identity elements to the active site [28]. tRNA-mediated pathways, if any, may be unimportant [33]. It is likely that the k_{cat} effect seen in cGluGlnRS may derive from a non-optimized protein-mediated rudimentary anti-codon binding information transmission pathway. Existence of such rudimentary transmission pathways after a domain-shuffling event or insertion of loops may create proteins which may confer some evolutionary advantages followed by further optimization.

Other chimaeric enzymes have also been created by domain exchange [34,35]. When the domain exchange involves related proteins, the resultant chimaera tends to have significant activity [36]. This is probably a result of divergence of domain interfaces from the optimized one. During evolution, many domains undergo recombination processes to generate new domain combinations. However, not all combinations are found in extant organisms. The basis of such selectivity is unclear. One possibility that emerges from this study is that only those chimaeric proteins that exhibit enhanced activity, perhaps due to some degree of complementary

interface, prove advantageous for the organism. This may be followed by further evolution to a more efficient entity.

ACKNOWLEDGEMENTS

JP1449 (DE3) was a gift from Professor Jacques Lapointe (Departement of Biochemistry and Microbiology, Université Laval, Québec, Canada). Plasmid containing substrate-binding domain of DnaK gene was a gift from Professor Tanaka (Kyoto Institute of Technology, Matsugasaki, Sakyo, Kyoto, Japan).

FUNDING

This work was funded by Council of Scientific and Industrial Research (India). S.D. and G.B. would like to acknowledge additional financial support from CSIR (Council of Scientific and Industrial Research), India.

REFERENCES

1 Doolittle, R. F. (1995) The multiplicity of domains in proteins. Annu. Rev. Biochem. **64**, 287–314
2 Babushok, D. V., Ostertag, E. M. and Kazazian, Jr, H. H. (2007) Current topics in genome evolution: molecular mechanisms of new gene formation. Cell. Mol. Life Sci. **64**, 542–554
3 Gilbert, W., de Souza, S. J. and Long, M. (1997) Origin of genes. Proc. Natl. Acad. Sci. U.S.A. **94**, 7698–7703
4 Patthy, L. (1987) Intron-dependent evolution: preferred types of exons and introns. FEBS Lett. **214**, 1–7
5 Alexander, R. W. and Schimmel, P. (2001) Domain-domain communication in aminoacyl-tRNA synthetases. Prog. Nucleic Acid Res. Mol. Biol. **69**, 317–349
6 Buechter, D. D. and Schimmel, P. (1993) Aminoacylation of RNA minihelices: implications for tRNA synthetase structural design and evolution. Crit. Rev. Biochem. Mol. Biol. **28**, 309–322
7 Moras, D. (1992) Structural and functional relationships between aminoacyl-tRNA synthetases. Trends Biochem. Sci. **17**, 159–164
8 Pham, Y., Li, L., Kim, A., Erdogan, O., Weinreb, V., Butterfoss, G. L., Kuhlman, B. and Carter, Jr, C. W. (2007) A minimal TrpRS catalytic domain supports sense/antisense ancestry of class I and II aminoacyl-tRNA synthetases. Mol. Cell **25**, 851–862
9 Siatecka, M., Rozek, M., Barciszewski, J. and Mirande, M. (1998) Modular evolution of the Glx-tRNA synthetase family–rooting of the evolutionary tree between the bacteria and archaea/eukarya branches. Eur. J. Biochem. **256**, 80–87
10 Nureki, O., Vassylyev, D. G., Katayanagi, K., Shimizu, T., Sekine, S., Kigawa, T., Miyazawa, T., Yokoyama, S. and Morikawa, K. (1995) Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. Science **267**, 1958–1965
11 Rould, M. A., Perona, J. J. and Steitz, T. A. (1991) Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. Nature **352**, 213–218
12 Urban, A., Neukirchen, S. and Jaeger, K. E. (1997) A rapid and efficient method for site-directed mutagenesis using one-step overlap extension PCR. Nucleic Acids Res. **25**, 2227–2228
13 Lin, S. X., Brisson, A., Liu, J., Roy, P. H. and Lapointe, J. (1992) Higher specific activity of the *Escherichia coli* glutamyl-tRNA synthetase purified to homogeneity by a six-hour procedure. Protein Expr. Purif. **3**, 71–74
14 Banerjee, R., Dubois, D. Y., Gauthier, J., Lin, S. X., Roy, S. and Lapointe, J. (2004) The zinc-binding site of a class I aminoacyl-tRNA synthetase is a SWIM domain that modulates amino acid binding via the tRNA acceptor arm. Eur. J. Biochem. **271**, 724–733
15 Banerjee, R., Mandal, A. K., Saha, R., Guha, S., Samaddar, S., Bhattacharyya, A. and Roy, S. (2003) Solvation change and ion release during aminoacylation by aminoacyl-tRNA synthetases. Nucleic Acids Res. **31**, 6035–6042
16 Bhattacharyya, T., Bhattacharyya, A. and Roy, S. (1991) A fluorescence spectroscopic study of glutaminyl-tRNA synthetase from *Escherichia coli* and its implications for the enzyme mechanism. Eur. J. Biochem. **200**, 739–745
17 Lapointe, J. and Delcuve, G. (1975) Thermosensitive mutants of *Escherichia coli* K-12 altered in the catalytic subunit and in a regulatory factor of the glutamyl-transfer ribonucleic acid synthetase. J. Bacteriol. **122**, 352–358
18 Russell, R. R. and Pittard, A. J. (1971) Mutants of *Escherichia coli* unable to make protein at 42 C. J. Bacteriol. **108**, 790–798
19 Racher, K. I., Kalmar, G. B. and Borgford, T. J. (1991) Expression and characterization of a recombinant yeast isoleucyl-tRNA synthetase. J. Biol. Chem. **266**, 17158–17164
20 Ibba, M., Hong, K. W., Sherman, J. M., Sever, S. and Soll, D. (1996) Interactions between tRNA identity nucleotides and their recognition sites in glutaminyl-tRNA synthetase determine the cognate amino acid affinity of the enzyme. Proc. Natl. Acad. Sci. U.S.A. **93**, 6953–6958
21 Roy, S. (2004) Fluorescence quenching methods to study protein-nucleic acid interactions. Methods Enzymol. **379**, 175–187
22 Jencks, W. P. (1975) Binding energy, specificity, and enzymic catalysis: the circe effect. Adv. Enzymol. Relat. Areas Mol. Biol. **43**, 219–410
23 Brown, J. R. and Doolittle, W. F. (1999) Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. J. Mol. Evol. **49**, 485–495
24 Lamour, V., Quevillon, S., Diriong, S., N'Guyen, V. C., Lipinski, M. and Mirande, M. (1994) Evolution of the Glx-tRNA synthetase family: the glutaminyl enzyme as a case of horizontal gene transfer. Proc. Natl. Acad. Sci. U.S.A. **91**, 8670–8674
25 Nureki, O., Vassylyev, D. G., Katayanagi, K., Shimizu, T., Sekine, S., Kigawa, T., Miyazawa, T., Yokoyama, S. and Morikawa, K. (1995) Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. Science **267**, 1958–1965
26 Weygand-Durasevic, I., Nalaskowska, M. and Soll, D. (1994) Coexpression of eukaryotic tRNASer and yeast seryl-tRNA synthetase leads to functional amber suppression in *Escherichia coli*. J. Bacteriol. **176**, 232–239
27 Sekine, S., Nureki, O., Tateno, M. and Yokoyama, S. (1999) The identity determinants required for the discrimination between tRNAGlu and tRNAAsp by glutamyl-tRNA synthetase from *Escherichia coli*. Eur. J. Biochem. **261**, 354–360
28 Rogers, M. J., Adachi, T., Inokuchi, H. and Soll, D. (1994) Functional communication in the recognition of tRNA by *Escherichia coli* glutaminyl-tRNA synthetase. Proc. Natl. Acad. Sci. U.S.A. **91**, 291–295
29 Sherman, J. M., Thomann, H. U. and Soll, D. (1996) Functional connectivity between tRNA-binding domains in glutaminyl-tRNA synthetase. J. Mol. Biol. **256**, 818–828
30 Weygand-Durasevic, I., Rogers, M. J. and Soll, D. (1994) Connecting anticodon recognition with the active site of *Escherichia coli* glutaminyl-tRNA synthetase. J. Mol. Biol. **240**, 111–118
31 Uter, N. T. and Perona, J. J. (2004) Long-range intramolecular signaling in a tRNA synthetase complex revealed by pre-steady-state kinetics. Proc. Natl. Acad. Sci. U.S.A. **101**, 14396–14401
32 Guha, S., Sahu, K., Roy, D., Mondal, S. K., Roy, S. and Bhattacharyya, K. (2005) Slow solvation dynamics at the active site of an enzyme: implications for catalysis. Biochemistry **44**, 8940–8947
33 Ghosh, A. and Vishveshwara, S. (2007) A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. Proc. Natl. Acad. Sci. U.S.A. **104**, 15711–15716
34 Diaz, E., Lopez, R. and Garcia, J. L. (1990) Chimeric phage-bacterial enzymes: a clue to the modular evolution of genes. Proc. Natl. Acad. Sci. U.S.A. **87**, 8125–8129
35 SternJohn, J., Hati, S., Siliciano, P. G. and Musier-Forsyth, K. (2007) Restoring species-specific posttransfer editing activity to a synthetase with a defunct editing domain. Proc. Natl. Acad. Sci. U.S.A. **104**, 2127–2132
36 Nixon, A. E., Warren, M. S. and Benkovic, S. J. (1997) Assembly of an active enzyme by the linkage of two protein modules. Proc. Natl. Acad. Sci. U.S.A. **94**, 1069–1073

Received 9 April 2008/22 September 2008; accepted 26 September 2008
Published as BJ Immediate Publication 26 September 2008, doi:10.1042/BJ20080747