

Evolution of the Glx-tRNA synthetase family: The glutaminyl enzyme as a case of horizontal gene transfer

(gene duplication/molecular phylogeny/aminoacyl-tRNA synthetase/human cDNA)

VALÉRIE LAMOUR[†], SOPHIE QUEVILLON[‡], SYLVIE DIRIONG[§], VAN CONG N'GUYEN[§], MARC LIPINSKI[†], AND MARC MIRANDE[‡]

[†]Laboratoire de Biologie des Tumeurs Humaines, Unité de Recherche Associée 1156, and [§]Laboratoire de Cytogénétique et de Génétique Oncologiques, Unité de Recherche Associée 1158, Centre National de la Recherche Scientifique, Institut Gustave Roussy, 94805 Villejuif Cedex, France; and [‡]Laboratoire d'Enzymologie du Centre National de la Recherche Scientifique, 91190 Gif sur Yvette, France

Communicated by Russell F. Doolittle, March 10, 1994

ABSTRACT An important step ensuring the fidelity in protein biosynthesis is the aminoacylation of tRNAs by aminoacyl-tRNA synthetases. The accuracy of this process rests on a family of 20 enzymes, one for each amino acid. One exception is the formation of Gln-tRNA^{Gln} that can be accomplished by two different pathways: aminoacylation of tRNA^{Gln} with Gln by glutaminyl-tRNA synthetase (GlnRS; EC 6.1.1.18) or transamidation of Glu from Glu-tRNA^{Gln} mischarged by glutamyl-tRNA synthetase (GluRS; EC 6.1.1.17). The latter pathway is widespread among bacteria and organelles that, accordingly, lack GlnRS. However, some bacterial species, such as *Escherichia coli*, do possess a GlnRS activity, which is responsible for Gln-tRNA^{Gln} formation. In the cytoplasm of eukaryotic cells, both GluRS and GlnRS activities can be detected. To gain more insight into the evolutionary relationship between GluRS and GlnRS enzyme species, we have now isolated and characterized a human cDNA encoding GlnRS. The deduced amino acid sequence shows a strong similarity with other known GlnRSs and with eukaryotic GluRSs. A molecular phylogenetic analysis was conducted on the 14 GlxRS (GluRS or GlnRS) sequences available to date. Our data suggest that bacterial GlnRS has a eukaryotic origin and was acquired by a mechanism of horizontal gene transfer.

Aminoacyl-tRNA synthetases define a family of 20 enzymes, each specifically responsible for the esterification of a given tRNA by its cognate amino acid (1, 2). All 20 synthetases have now been cloned from *Escherichia coli* and several from eukaryotic cells. Based on the analysis of their amino acid sequences, and on the crystal structures of 5 enzymes, aminoacyl-tRNA synthetases were partitioned into two distinct structural classes (3–6). It appears to be a constant rule that an aminoacyl-tRNA synthetase with a given specificity always belongs to the same class, I or II, irrespective of its prokaryotic or eukaryotic origin. No example of class switching has been reported to date.

Among the 20 synthetases, a special issue concerns the evolutionary relationship between glutamyl-tRNA synthetase (GluRS; EC 6.1.1.17) and glutaminyl-tRNA synthetase (GlnRS; EC 6.1.1.18). These two class I synthetases have very similar primary structures, indicating that GluRS and GlnRS have evolved from a common ancestor (6, 7). In that connection, it is noteworthy that GlnRS is the only synthetase known to be dispensable for protein synthesis. In Gram-positive eubacteria, cyanobacteria and halobacteria, as well as in mitochondria and chloroplasts from eukaryotic cells, GluRS aminoacylates both tRNA^{Glu} and tRNA^{Gln} with glutamate; Gln-tRNA^{Gln} is formed by transamidation of the misacylated Glu-tRNA^{Gln} (refs. 8 and 9 and references there-

in). The prevailing hypothesis assumes that the ancestral organism from which modern bacteria and eukaryotic cells arose contained only GluRS, GlnRS appearing later on, after duplication of the GluRS gene (7, 9).

A cDNA cloned in *Drosophila* has been shown previously to encode a multifunctional protein with GluRS and ProRS carried by the N and C termini, respectively (10). The N-terminal domain showed a higher homology with *E. coli* GlnRS (37% of identical amino acids) than with *E. coli* GluRS (22% identity). However, only GluRS activity was detected when the corresponding cDNA fragment was expressed in *E. coli* (10). These observations strengthened the view according to which Glx-tRNA synthetase genes are closely related and diverged recently in evolution.

Earlier on, a human cDNA displaying an overall structure similar to that of the fly cDNA encoding the multifunctional protein with glutamyl- and prolyl-tRNA synthetase activities (GluProRS) had been reported (11). This cDNA was attributed to a GlnRS cDNA solely on the basis of sequence comparison (11). Since the *Drosophila* and human proteins were strikingly similar (66% of identical residues), we have proposed that this human cDNA would rather encode a GluProRS enzyme (10). This implied that human GlnRS remained to be cloned. In this paper, we describe the isolation of a human GlnRS cDNA that has been entirely sequenced. The identification of the actual GlnRS cDNA sheds some light into the evolutionary history of the GlxRS genes. In particular, a molecular phylogeny analysis performed with all GlxRS sequences available to date suggests an evolutionary scenario involving gene duplication of a eukaryotic GluRS, followed by a lateral gene transfer event. This scheme is consistent with the presence of GlnRS in some, but not all, prokaryotic cells.

MATERIALS AND METHODS

Isolation and Characterization of Human GlnRS cDNA. An inter-*Alu* clone, DD4, was obtained from a radiation hybrid selected for its content in human chromosome 22 fragments (12) and used to screen a human fetal brain cDNA library constructed in λ -ZAPII vector (Stratagene). The cDD4 pBluescript phagemid obtained was shown to contain a 5-kb-long insert that proved to be a chimera. The 5' extremity of cDD4, encoding human GlnRS, was used as a probe to isolate clone F27 from a human liver cDNA library made available to us by Pierre Meulien (Institut Pasteur-Mérieux, Marcy-l'Etoile, France). To reconstitute a full-length cDNA, the 5' *EcoRI* fragment of F27, containing the translation initiation codon, was subcloned into

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: GlnRS, glutaminyl-tRNA synthetase; GluRS, glutamyl-tRNA synthetase; GluProRS, multifunctional protein with glutamyl- and prolyl-tRNA synthetase activities; IPTG, isopropyl β -D-thiogalactoside.

pBluescript II SK(+) to give F27.6. The *Eag I-Eco47III* 5' fragment of cDD4 was then replaced with the corresponding fragment isolated from F27.6. The entire sequence of GlnRS cDNA was established on both strands by automated sequencing (Applied Biosystems, model 373A) using fluorescent primers or fluorescent dideoxynucleotides in combination with custom-designed primers (Bioprobe, Montreuil-sous-Bois, France).

Chromosomal Assignment. Somatic cell hybrids used to chromosomally assign the human GlnRS gene have been described (13). Southern blots of *HindIII*-digested genomic DNA from these hybrids were hybridized with a nick-translated GlnRS cDNA probe, encompassing amino acids 54–553 in Fig. 1. The aminoacylase 1 gene (*ACY1*) that maps to chromosome 3p21 was used as a reference chromosome 3 marker (14).

Western Blot Analysis and Activity Measurements. A fragment of cDD4, coding for amino acids 165–775 in Fig. 1, was cloned in the pGEX-3X vector (Pharmacia). The expression of the fusion protein was induced by isopropyl β -D-thiogalactoside (IPTG). Extracts were prepared either by sonication of a bacterial suspension (sonicated extracts) or by SDS treatment of the bacterial pellet (total extract), followed by centrifugation to remove insoluble material. Aminoacyl-tRNA synthetase activities present in the sonicated extracts of recombinant bacteria were assayed by the aminoacylation reaction (15), using beef liver tRNA and ¹⁴C-labeled amino acids (Amersham). Sonicated and total extracts were subjected to Western blot analysis (16), using antibodies directed to the aminoacyl-tRNA synthetase components of the sheep liver complex (17) and the ECL detection reagents from Amersham.

In Vitro Translation. The reconstituted full-length GlnRS cDNA was transcribed *in vitro* using the riboprobe Gemini II system (Promega). *In vitro* translation was performed as recommended by the manufacturer using the rabbit reticulocyte lysate system (Promega) and [³⁵S]methionine (Amersham).

RESULTS

Isolation of a Human cDNA Encoding GlnRS. cDD4, a 5-kb-long cDNA, was isolated from a fetal brain library

during a search for new genes mapping to a disease-associated region of human chromosome 22. Whereas the 3' end of that cDNA was found to map to the expected region of chromosome 22, the 5' end did not. Partial sequencing of cDD4 suggested that this cDNA must be a chimera. The amino acid sequences deduced from its 5' moiety were closely homologous to the GlnRS sequences stored in protein data bases.

The cDNA from cDD4 did not code for the 53 N-terminal residues of GlnRS (Fig. 1). The 5' upstream sequences were obtained by screening a human liver library with a probe corresponding to the 5' end of cDD4. The 2437-bp-long cDNA sequence (EMBL data library, accession no. X76013) displays an open reading frame of 2325 nucleotides. The ATG initiation codon is preceded by 5 nucleotides defining a suitable translation initiation site (23). The 107-bp-long 3' untranslated region contains a short poly(A) tail preceded by a canonical AATAAA polyadenylation signal. By Northern blot analysis on total RNA isolated from various tumor cell lines, a single mRNA of \approx 2500 nucleotides was observed (not shown).

Chromosomal Assignment. The gene encoding the human GlnRS was mapped to chromosome 3 by hybridization of a GlnRS probe to a panel of 25 somatic cell hybrids (13). A correlation was observed between the detection of a human-specific GlnRS signal and the presence of human chromosome 3 in the hybrid. Whereas 14 of the 15 GlnRS-positive hybrids were found to contain human chromosome 3, no chromosome 3 was detected in any of the 10 GlnRS-negative hybrids (not shown). The single hybrid that was GlnRS-positive but did not contain an intact chromosome 3, as assessed by cytogenetic analysis, nevertheless scored positive for the reference chromosome 3 marker *ACY1* (14), suggesting the presence in the hybrid of at least a fragment of human chromosome 3. The genes for 11 human aminoacyl-tRNA synthetases have been mapped so far (quoted in ref. 24). To our knowledge, the GlnRS gene is the first to be mapped to chromosome 3. In particular, the human gene encoding the multifunctional GluProRS has been mapped to chromosome 1 (25).



FIG. 1. Alignment of the amino acid sequences of GlnRSs. The complete amino acid sequences of GlnRS from *Homo sapiens* (Hs), *Saccharomyces cerevisiae* (Sc) (18), and *E. coli* (Ec) (19) were aligned using the algorithm developed by Kanehisa *et al.* (20), by weighing with the mutation data matrix of Dayhoff *et al.* (21). Amino acid residues are numbered on the left. Identities are indicated by (•), conservative substitutions by (+). The two sequence motifs (HIGH and KMSKS, or derivatives thereof) characteristic of class I aminoacyl-tRNA synthetases are boxed. The two sequence blocks (shaded) that were used for phylogenetic analysis encompass the dinucleotide binding domain of the *E. coli* GlnRS (22).

Identification of the Deduced Protein as GlnRS. The cDNA has a coding potential for a protein of 775 amino acids, with a calculated molecular mass of 87.8 kDa. The two consensus sequences characteristic of class I synthetases are found at positions 277–280 and 493–497 of the deduced protein (Fig. 1). A computer-assisted search (26) revealed homology to the other GlnRS enzymes stored in protein data libraries. Alignment of human, yeast, and bacterial GlnRS is shown in Fig. 1. The homology between the human and yeast enzymes extends over their entire lengths (41% of identical amino acid residues); both have a large N-terminal polypeptide extension as compared to the prokaryotic GlnRS. In yeast, the function of this extension is not known but it is dispensable for aminoacylation activity (27). As in the case of mammalian aspartyl- and arginyl-tRNA synthetases (28, 29), two other components of the multisynthetase complex, the extension of human GlnRS is rich in hydrophobic amino acids and thus could be involved in complex assembly.

To ascertain the assignment of the deduced protein to a GlnRS, a fragment of cDD4 coding for the polypeptide fraction that is homologous to *E. coli* GlnRS was expressed in *E. coli* as a glutathione-S-transferase fusion protein. The resulting protein was essentially insoluble. As a consequence, no GlnRS activity could be detected in the supernatant of *E. coli* crude extracts. Antigenic features of the protein were examined by Western blot analysis of total or sonicated extracts of recombinant bacteria (Fig. 2A). GlnRS-specific antibodies (17) detected large amounts of the fusion protein in a total extract from an IPTG-induced culture (Fig. 2A, lane 7). By contrast, low amounts of this protein were observed in the soluble fraction of a sonicated extract (Fig. 2A, lanes 8), similar to the basal level observed in a noninduced culture (Fig. 2A, lanes 5 and 6). To assess that a full-length cDNA has been cloned, the complete cDNA was translated *in vitro*. The major translation product (Fig. 2B,

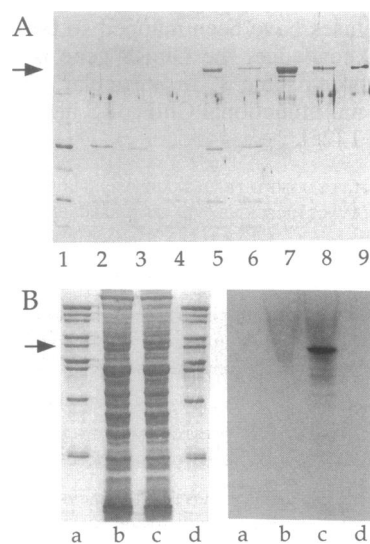


FIG. 2. Immunochemical detection of the cDNA-encoded polypeptide and assignment to a full-length GlnRS cDNA. (A) Western analysis of total (lanes 1, 3, 5, and 7) or sonicated (lanes 2, 4, 6, and 8) extracts from XL1-Blue cells transformed by the pGEX-3X control plasmid (lanes 1–4) or by the pGEX-3X recombinant vector (lanes 5–8). Lanes 3, 4, 7, and 8, extracts from IPTG-induced cultures. The arrow indicates the position of migration of the fusion protein, as revealed by Coomassie blue staining. Purified multisynthetase complex (0.1 μ g) from rat liver was loaded in lane 9. (B) The products of an *in vitro* translation in the absence (lanes b) or in the presence (lanes c) of an *in vitro* transcript of the human GlnRS cDNA were subjected to SDS/PAGE, stained with Coomassie blue (left) and exposed to x-ray films (right). Lanes a and d, rat multisynthetase complex (1 μ g); the arrow points to the GlnRS component.

lane c; calculated molecular mass of 87.8 kDa) exhibited an electrophoretic mobility identical to that of the GlnRS component of the multisynthetase complex from rat liver (Fig. 2B, lanes a and d; apparent mass of 94 kDa).

Phylogenetic Analysis of the Glx-tRNA Synthetase Family. The amino acid sequences for 14 GlxRS of prokaryotic and eukaryotic origins are available. These sequences were aligned using the MACAW program (Fig. 3). Several noncontiguous blocks of homology (black boxes in Fig. 3) are shared between all enzymes, irrespective of their GluRS or GlnRS identity. The corresponding sequences map to the dinucleotide binding domain of class I synthetases, as deduced from the crystal structure of *E. coli* GlnRS (22). Paradoxically, GluRSs from higher eukaryotes (fly and human) share additional homology blocks with the GlnRS enzymes rather than with their functional counterparts in prokaryotes (Fig. 3). Taking into account the 11 enzymes for which the complete sequences are known, the two eukaryotic GluRSs were found to share about 20–23% residues with the prokaryotic GluRS enzymes, as compared to 34–37% with the GlnRS species. These values were normalized to the shortest synthetase—namely, GluRS from *T. thermophilus*.

The amino acid sequences corresponding to the most conserved region of the GlxRS enzymes were used to build an evolutionary tree by the maximum parsimony approach (Fig. 4). An essentially similar tree was also inferred by using the fundamentally different, unweighted pair-group method (37). The resulting phylogenetic tree showed two clusters of sequences (Fig. 4). The eubacterial GluRSs, including the yeast mitochondrial enzyme, built one major branch of the tree. In addition to synthetases with Gln specificity, the other branch contained the fly and human GluRS. The two branches have approximately the same size, exemplifying the tight grouping of eukaryotic GluRS and of GlnRS of various origins.

The statistical significance of this anomalous location of the fly and human GluRS sequences was assessed by using the bootstrap test, a random resampling method (38). The grouping of GlnRS and eukaryotic GluRS sequences was supported by 100% of 1000 bootstrap replications of a branch-

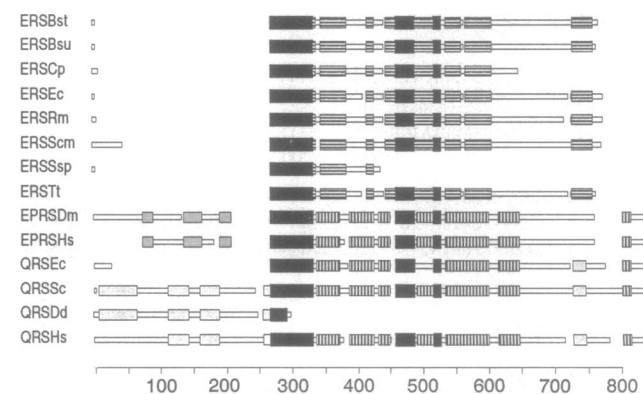


FIG. 3. Schematic alignment of proteins from the Glx-tRNA synthetase family. Protein sequences were compared by using the MACAW program (30). Blocks of similarity are indicated by distinct shaded areas. Sequences were as follows: GluRS (ERS) from *Bacillus stearothermophilus* (Bst) and *Bacillus subtilis* (Bsu) (31); *Thermus thermophilus* (Tt) (32); *E. coli* (Ec) (7); *Rhizobium meliloti* (Rm) (33); *Synechococcus sp.* (Ssp) (GenBank SYOPSAEGLT); *Chlamydia psittaci* (Cp) (34); *S. cerevisiae* mitochondria (Scm) (A. Tzagoloff, personal communication); *Drosophila melanogaster* (Dm; gluprolyl enzyme EPRS) (10); *H. sapiens* (Hs; gluprolyl enzyme EPRS) (11). GlnRS (QRS) from *E. coli* (Ec) (19); *S. cerevisiae* cytoplasm (Sc) (18); *Dictyostelium discoideum* (Dd) (35); *H. sapiens* (Hs) (this study). Sequences for ERSSsp, ERSCp, and QRSDd are partial sequences. For EPRSDm and EPRSHs, only the GluRS region is shown.

ERSSep 3 VVRIAPSPGTNLHICTARTAVFNWLFANHHGGTFVLRVEDTDLERSKFEVTENIKTKLQMLGLH
 ERSQp 8 VVRIAPSPGTDPHVGTAYMALFNEIKARRNFKGKMLRIEDTDQTRSDVDEKNI FSA LKWKGLQ 115 QVLIKSDGFPYTHFANVVDHLMGITHVLRGEWLSSTPKHLLLYEAFGWAEPTFLDMPDLLNPDGTKLSKRRK
 ERSTt 2 VVTRIPSPGTDPHVGTAYIALFNVAWARRNGGRFVRIEDTDRARYVFGAERI LAALKWLGLS 109 VVLLKSDGVPYTHLANVVDHLMGIVTDVIRAEWLVSTPIHVLLYRAFGWEAPRFTYHMLPLLNPDKTKLSKRRK
 ERSRm 6 VVRIAPSPGTGPHVGTAYIALFNVAWARRNGGRFVRIEDTDRARYVFGAERI LAALKWLGLS 115 QVLLKADGMPYTHLANVVDHLMGITHVLRGEWLSSTPKHLLLYEAFGWAEPTFLDMPDLLNPDGTKLSKRRK
 ERSBet 5 VRVGYAPSPGTGHLHIGGARTALFNVAWARRNGGRFVRIEDTDRARYVFGAERI LAALKWLGLS 116 VVIWKANGIPYTHFANVVDHLMGITHVLRGEWLSSTPKHLLLYEAFGWAEPTFLDMPDLLNPDGTKLSKRRK
 ERSBeu 5 VVRIAPSPGTGHLHIGGARTALFNVAWARRNGGRFVRIEDTDRARYVFGAERI LAALKWLGLS 115 FVIVKADGMPYTHFANVVDHLMGITHVLRGEWLSSTPKHLLLYEAFGWAEPTFLDMPDLLNPDGTKLSKRRK
 ERSEc 3 IKTRFAPSPGTGLVGGARTALYSWLFARNHGGGFVLRIEDTDLERSTPEAIEAIDMGMMWLSLE 102 LIIIRRTDGSPTYNFCVVDMDMEITHEVIRGEDEHINPTQKQIMVYQAFGWMDIPOPQGHMTLIVNESRKKLSKRRK
 ERSScm 45 VTRRFAPSPGTGLVGGARTALYSWLFARNHGGGFVLRIEDTDLERSTPEAIEAIDMGMMWLSLE 102 LIIIRRTDGSPTYNFCVVDMDMEITHEVIRGEDEHINPTQKQIMVYQAFGWMDIPOPQGHMTLIVNESRKKLSKRRK
 EPRSDm 204 VVRFPEASGLYHIGHAKAALLNGHYALVCCQTLIRFDDTNPKEKEDEPKVILIEDVAMLIHRIK 114 LILVRSKDLPTTHLANVVDHLMGITHVIRGEWLPSTPHALYNAFGWACPKFTIYPLITVGDKLSKRRK
 EPRSHs 126 VVRFPEASGLYHIGHAKAALLNGHYALVCCQTLIRFDDTNPKEKEDEPKVILIEDVAMLIHRIK 103 RGTGKVIPTDYDFACPIVDSIEGVTHALRTTEYHEDRDEQPMFIDALKLR-KPTIWSYSRINANTVLSKRRK
 QRSEc 27 VTRFPPEPNGLYHIGHAKSICLNFGIAQDYRGQCNLRFDDTNPKEKEDEPKVILIEDVAMLIHRIK 108 QTGNKWCIVPMYDFTHCISDALEGI THSLCTLEFQDMRRLIYVNVLDNITIPVHPRQVEYSRINANTVLSKRRK
 QRSSc 252 VTRFPPEPNGLYHIGHAKSICLNFGIAQDYRGQCNLRFDDTNPKEKEDEPKVILIEDVAMLIHRIK 112 RTGKWCIVPTDYDFTHCISDALEGI THSLCTLEFQDMRRLIYVNVLDNITIPVHPRQVEYSRINANTVLSKRRK
 QRSHs 264 VTRFPPEPNGLYHIGHAKSICLNFGIAQDYRGQCNLRFDDTNPKEKEDEPKVILIEDVAMLIHRIK 98 RTGKWCIVPTDYDFTHCISDALEGI THSLCTLEFQDMRRLIYVNVLDNITIPVHPRQVEYSRINANTVLSKRRK

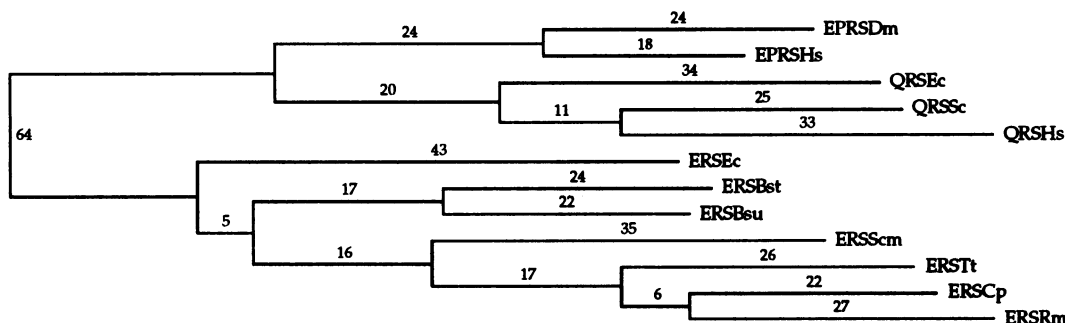


Fig. 4. Phylogenetic tree of Glx-tRNA synthetases. The homologous regions from the dinucleotide binding domain of available GlxRS sequences are aligned. Synthetases are named as in Fig. 3. The most parsimonious tree was searched using the exhaustive method of the PAUP program (36). Branch lengths represent the number of amino acid changes required.

and-bound search (36). By contrast, a lower confidence estimate was obtained for the relative branching order shown in Fig. 4 for the various prokaryotic GluRSs, which should be considered as tentative. When independent analyses were conducted on the first and second halves of the dinucleotide binding domain, a difference was observed in the relative branching order of yeast mitochondrial and Gram-positive GluRS. The analysis conducted with the first half of this region localized the *Synechococcus* sp. GluRS, from the phylum of cyanobacteria, as an individual branch between the *E. coli* and *Bacillus* branches.

DISCUSSION

In human, as well as in all metazoan species studied so far, the nine aminoacyl-tRNA synthetases specific for amino acids Glu, Pro, Ile, Leu, Met, Gln, Lys, Arg, and Asp are associated within a multisynthetase complex (2). In this paper, we have reported the cloning and sequencing of a cDNA encoding human GlnRS. The assignment to GlnRS is supported by several lines of evidence. (i) The deduced translation product is homologous to the other known GlnRS species (Figs. 1, 3, and 4). (ii) The encoded protein is recognized by GlnRS-specific antibodies and (iii) has an apparent molecular mass identical to that of the GlnRS component of the multisynthetase complex (Fig. 2). Since the protein expressed in *E. coli* was insoluble, the catalytic activity could not be directly assayed.

The isolation of a bona fide GlnRS cDNA confirms that the cDNA previously reported and ascribed to human GlnRS (11) actually encodes GluRS, as suggested by its similarity with the fly GluRS (10). However, this initial misleading identification is relevant to the evolutionary scheme that prevails within the GlxRS family. In the light of the present study, the paradoxical observation (10) that eukaryotic GluRS share more extensive sequence similarities with prokaryotic and eukaryotic GlnRS than with the functionally homologous prokaryotic GluRS proves to have evolutionary significance. The phylogenetic tree built using all GlxRS sequences available (Fig. 4) confirms the anomalous branching of eukaryotic GluRS with all the GlnRS species and strengthens the view according to which they arose from a common ancestral enzyme.

The absence of a Gln-specific enzyme in many prokaryotic cells could be a landmark of the ancestral state. Two possible evolutionary schemes can be invoked *a priori* to account for the particular branching of the GlnRS species. Both rely on the assumption that in the universal ancestor of the present day primary kingdoms (39), an early GluRS protein aminoacylated tRNA^{Glu} and tRNA^{Gln} (8, 9). Alternatively, the possibility that the amino acid Gln appeared later on in evolution (40) cannot be dismissed. The evolution of the biosynthetic pathways and the emergence of a new amino acid would have required a new enzyme for the selective attachment of Gln to its cognate tRNA. Such an enzyme would have appeared in some, but not all, eubacteria, those lacking GlnRS adopting, or retaining, the transamidation pathway.

In the first of the two models shown in Fig. 5A, a single ancestral GluRS gene (E) would have evolved along three pathways leading to (i) a nondiscriminative enzyme (E*) that accommodates both tRNA^{Glu} and tRNA^{Gln} (top), (ii) a specialized one (E) that recognizes only tRNA^{Glu} (middle), and (iii), following a gene duplication event, two sister genes (E and Q) that diverged to acquire distinct Glu and Gln specificities, the latter being laterally transferred from eukaryote to eubacteria (bottom). From the length of the branches in the phylogenetic tree shown in Fig. 4, it appears that the diver-

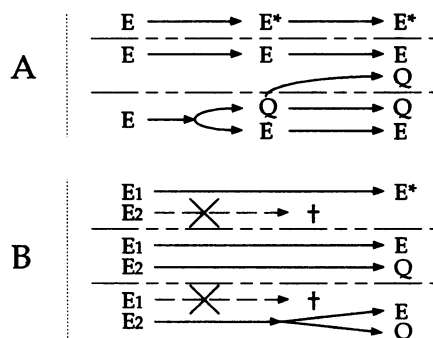


FIG. 5. Schematic putative evolutionary routes for GlxRS species. The two evolutionary schemes (A and B) proposed are detailed in the main text. In each scheme, the pathways account for the situations observed in eukaryotes (bottom), and prokaryotes with (middle) or without (top) a GlnRS.

gence between GlnRS and eukaryotic GluRS should have occurred more recently than the radiation between prokaryotic and eukaryotic GluRS. The respective usage of the transamidation and GlnRS pathways within the eubacterial kingdom is not precisely known. In the eubacterial tree (39), cyanobacteria and Gram-positive bacteria, which branch close to the phylum of purple bacteria, lack GlnRS (8, 9). This contrasts with *E. coli*, a representative of the γ subdivision of purple bacteria, which harbors a GlnRS (19). Therefore, the GlnRS pathway might be restricted, within the eubacterial kingdom, to the phylum of purple bacteria. Moreover, since eukaryotic mitochondria, which arose from an endosymbiont of the α subdivision of purple bacteria, also lack GlnRS, the horizontal genetic transfer might specifically apply to the β and γ , or possibly only γ , subdivisions of purple bacteria. The lack of GlnRS in halobacteria (41) further suggests that the transamidation pathway might also be widespread within the archaeobacterial kingdom.

In the second model that could account for the inferred phylogeny of the GlxRS family, it is assumed that two GluRS genes, E1 and E2, preexisted (Fig. 5B). Then, each one of them would have undergone several independent molecular events of gene extinction, specialization, and duplication in different phyla. According to the model, E2 would have given rise to all GlnRS species and to the eukaryotic GluRS, the other GluRS deriving from E1. This model implies a multiplicity of events that makes it less likely than the previous one. In addition, according to the universal phylogenetic tree designed by Woese (39), this scheme further implies that several phyla from the eubacterial kingdom—i.e., cyanobacteria and Gram-positive bacteria—Independently lost E2. The model predicts that residual traces of E1 and E2 genes should be detectable in the genome of eukaryotes and in some prokaryotic organisms lacking GlnRS, respectively. The search for vestigial enzymes and the knowledge of additional GlxRS sequences will help to clarify this issue.

Possible cases of horizontal gene transfer from eukaryotes to prokaryotes have been reported and their likelihood of occurrence has been discussed (42, 43). However, the postulate of genetic exchange between primary kingdoms remains a matter of debate. We have proposed an evolutionary scheme for the GlxRS family that requires such a gene transfer event. If the GlnRS observed in eubacteria are indeed of eukaryotic origin, it is predicted that the presence of GlnRS in prokaryotes will be restricted to few phyla. As our knowledge on the spreading of GlnRS among species expands, and with the multiplication of available GlxRS sequences, a more precise view of their evolutionary relationship should emerge.

We thank Philippe Dessen for his fruitful suggestion, Prof. Alexander Tzagoloff for communicating the sequence of yeast mitochondrial GluRS, Pierre Meulien and Anne Dejean for the liver cDNA library, and Manolo Gouy for fruitful discussions. This work was supported by grants from the Association pour la Recherche sur le Cancer (M.L. and M.M.), the Association Française contre les Myopathies (M.L.), and the Groupement de Recherches et d'Études sur les Génomes (M.L.). V.L. and S.Q. are Fellows from the Ministère pour la Recherche et l'Enseignement Supérieur.

- Schimmel, P. (1987) *Annu. Rev. Biochem.* **56**, 125–158.
- Mirande, M. (1991) *Prog. Nucleic Acid Res. Mol. Biol.* **40**, 95–142.
- Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. (1990) *Nature (London)* **347**, 203–206.
- Burbaum, J. J. & Schimmel, P. (1991) *J. Biol. Chem.* **266**, 16965–16968.
- Cusack, S., Härtlein, M. & Leberman, R. (1991) *Nucleic Acids Res.* **19**, 3489–3498.
- Nagel, G. N. & Doolittle, R. F. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 8121–8125.
- Breton, R., Sanfaçon, H., Papayannopoulos, I., Biemann, K. & Lapointe, J. (1986) *J. Biol. Chem.* **261**, 10610–10617.
- Lapointe, J., Duplain, L. & Proulx, M. (1986) *J. Bacteriol.* **165**, 88–93.
- Schön, A., Kannangara, G., Gough, S. & Söll, D. (1988) *Science* **331**, 187–190.
- Cerini, C., Kerjan, P., Astier, M., Gratecos, D., Mirande, M. & Séméria, M. (1991) *EMBO J.* **10**, 4267–4277.
- Fett, R. & Knippers, R. (1991) *J. Biol. Chem.* **266**, 1448–1455.
- Lamour, V., Lévy, N., Desmaze, C., Baud, V., Lécluse, Y., Delattre, O., Bernheim, A., Thomas, G., Aurias, A. & Lipinski, M. (1993) *Hum. Mol. Genet.* **2**, 535–540.
- N'Guyen, V. C., Weil, D., Finaz, C., Cohen-Haguenaer, O., Gross, M. S., Jégou-Foubert, C., de Tand, M.-F., Cochet, C., de Grouchy, J. & Frézal, J. (1986) *Ann. Genet.* **29**, 20–26.
- Human Gene Mapping 11 London Conference (1991) *Cytogenet. Cell Genet.* **58**, 1–2197.
- Mirande, M., LeCorre, D. & Waller, J.-P. (1985) *Eur. J. Biochem.* **147**, 281–289.
- Harlow, E. & Lane, D. (1988) *Antibodies: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
- Mirande, M., Cirakoglu, B. & Waller, J.-P. (1982) *J. Biol. Chem.* **257**, 11056–11063.
- Ludmerer, S. W. & Schimmel, P. (1987) *J. Biol. Chem.* **262**, 10801–10806.
- Yamao, F., Inokuchi, H., Cheung, A., Ozeki, H. & Söll, D. (1982) *J. Biol. Chem.* **257**, 11639–11643.
- Kanehisa, M., Klein, P., Greif, P. & DeLisi, C. (1984) *Nucleic Acids Res.* **12**, 417–428.
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
- Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989) *Science* **246**, 1135–1142.
- Kozak, M. (1991) *J. Cell Biol.* **115**, 887–903.
- Cruzen, M. E., Bengtsson, U., McMahon, J., Wasmuth, J. J. & Arfin, S. M. (1993) *Genomics* **15**, 692–693.
- Kunze, N., Bittler, E., Fett, R., Schray, B., Hameister, H., Wiedorn, K. H. & Knippers, R. (1990) *Hum. Genet.* **85**, 527–530.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
- Ludmerer, S. W., Wright, D. J. & Schimmel, P. (1993) *J. Biol. Chem.* **268**, 5519–5523.
- Mirande, M., Lazard, M., Martinez, R. & Latreille, M.-T. (1992) *Eur. J. Biochem.* **203**, 459–466.
- Lazard, M. & Mirande, M. (1993) *Gene* **132**, 237–245.
- Schuler, G. D., Altschul, S. F. & Lipman, D. J. (1991) *Proteins* **9**, 180–190.
- Breton, R., Watson, D., Yaguchi, M. & Lapointe, J. (1990) *J. Biol. Chem.* **265**, 18248–18255.
- Nureki, O., Susuki, K., Hara-Yokoyama, M., Kohno, T., Matsuzawa, H., Ohta, T., Shimizu, T., Morikawa, K., Miyazawa, T. & Yokoyama, S. (1992) *Eur. J. Biochem.* **204**, 465–472.
- Laberge, S., Gagnon, Y., Bordeleau, L. M. & Lapointe, J. (1989) *J. Bacteriol.* **171**, 3926–3932.
- Wichlan, D. G. & Hatch, T. P. (1993) *J. Bacteriol.* **175**, 2936–2942.
- Singleton, C. K., Manning, S. S. & Ken, R. (1989) *Nucleic Acids Res.* **17**, 9679–9692.
- Swofford, D. L. (1993) PAUP: Phylogenetic Analysis Using Parsimony (Illinois Natural History Survey, Champaign), Version 3.1.
- Nei, M., Stephens, J. C. & Saitou, N. (1985) *Mol. Biol. Evol.* **2**, 66–85.
- Felsenstein, J. (1988) *Annu. Rev. Genet.* **22**, 521–565.
- Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221–271.
- Wong, J. T. F. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1909–1912.
- Gupta, R. (1984) *J. Biol. Chem.* **259**, 9461–9471.
- Sprague, G. F., Jr. (1991) *Curr. Opin. Genet. Dev.* **1**, 530–533.
- Smith, M. W., Feng, D. F. & Doolittle, R. F. (1992) *Trends Biochem. Sci.* **17**, 489–493.