The aminoacyl-tRNA synthetases (AARSs) are a diverse group of enzymes that ensure the fidelity of transfer of genetic information from DNA into protein. They catalyse the attachment of amino acids to transfer RNAs and thereby establish the rules of the genetic code by virtue of matching the nucleotide triplet of the anticodon with its cognate amino acid. Currently, 818 AARS primary structures have been reported from archaeabacteria, eubacteria, mitochondria, chloroplasts and eukaryotic cells. The database is a compilation of the amino acid sequences of all AARSs, known to date, which are available as separate entries or alignments of related proteins via the WWW at http://rose.man.poznan.pl/aars/index.html

INTRODUCTION

The genetic code is established in a single aminoacylation reaction of tRNAs catalysed by the aminoacyl-tRNA synthetases (AARSs). The conservation of the genetic code suggests that the AARSs evolved early and were probably among the first protein enzymes to emerge from the RNA world (1–2). AARSs constitute a family of 20 cellular enzymes that are responsible for the specific cestification of tRNAs with their cognate amino acids, and thus are essential in maintaining the fidelity of the protein biosynthesis process. Specific AARSs play roles in cellular fidelity, tRNA processing, RNA splicing, RNA trafficking, apoptosis and transcriptional and translational regulation (3–9). Although AARSs catalyse the same basic reaction they differ in size, amino acid sequence and subunit structure. The quaternary structures of AARSs include monomers (α), homodimers and heterotetramers (α2, α β) and heterotetramers (α β β). PheRS adopts the α β tetrameric structure in all prokaryotic and eukaryotic cytoplasmic sources known. A crystal structure of PheRS from Thermus thermophilus indicates that this tetrameric structure actually behaves as an α β structure, that is, a dimer built from two homodimers. A similar situation is observed for yeast and human mitochondrial PheRS, which may constitute a class of PheRS distinct from the enzymes found in prokaryotes and the eukaryotic cytoplasm (10–13). The ‘double sieve’ mechanism for proofreading (editing) of the aminoacylation reaction has been confirmed by solving the crystal structure of isoleucyl-tRNA synthetase from T. thermophilus. In that structure the selectivity is achieved by steric exclusion at the synthetic active site of amino acids larger than the correct one and only smaller amino acids are sterically allowed to bind at the hydrophobic editing active site. Both active sites are separate domains of IleRS and therefore they have to be brought closer together by a rotation of the editing domain upon addition of tRNA (14). Analysis of the crystal structure of Staphylococcus aureus IleRS suggests an alternative mechanism involving a tRNA-dependent shuttling of the incorrect product between the two active sites (15).

Although non-cognate amino acids are recognised as substrates by many AARSs in the first step of the aminoacylation reaction, in most cases the misactivated amino acids are not transferred to tRNA, e.g., homocysteine is mischarged by IleRS, LeuRS and MetRS at frequencies exceeding the frequency of translation errors in vivo, and are never transferred to tRNA. In a few cases where a non-cognate amino acid is transferred to tRNA, the misacylation is 10^3–10^6-fold less efficient than the correct charging, e.g., IleRS promotes one reaction of tRNA^Ile with Val per 350 000 correct acylations with Ile; ValRS promotes one aminoacylation of one molecule of tRNA^Ile with Ile and Thr per 5000 and 350 000 cognate chargings with Val, respectively. LysRS catalyses one mis-acylation of tRNA^Lys with Arg, Thr, Met, Leu, Ala, Cys and Ser per 1600, 16 000, 32 000, 132 000, 265 000, 560 000 and 750 000 correct acylations with Lys, respectively. The observed limited selectivity in tRNA aminoacylation reaction appears to be due to inefficient editing of Met, Leu, Cys, Ala and Thr or absence of editing of Arg and Ser. Purified Arg-tRNA^Lys, Thr-tRNA^Lys and Met-tRNA^Lys were essentially not deacylated by LysRS. On the other hand, LysRS possesses an efficient editing mechanism which prevents misacylation with ornithine (16).

In higher eukaryotes, nine AARSs are associated together forming a supramolecular multi-enzyme complex. The function of the whole complex which contains, in addition to the AARS three auxiliary peptides p43, p38 and p18, remains unknown. It has been shown that p43 may be the precursor of a tumor-specific cytokine, endothelial monocyte-activating polypeptide II (EMAP II) based on >80% sequence identity between the two proteins. The precursor for EMAP II is associated with the N-terminal extension of human ArgRS facilitating aminoacylation reaction and processed with releasing C-terminal cytokine domain of 23 kDa, which shares homology with the C-terminal parts of MetRS of prokaryotes, archaea and nematode, and also a yeast Arc1p/G4p, which interacts with MetRS and GlnRS (17). Human TyrRS can be split into two fragments with distinct cytokine activities. The EMAP II-like C-terminal domain has potent leukocyte and monocyte chemotaxis activity and stimulates production of myeloperoxidase, tumor necrosis factor and tissue factor. The catalytic N-terminal domain binds to the interleukin-8 type A receptor and functions as an interleukin-8-like cytokine (18). Under apoptotic conditions,
specific proteolysis of TyrRS leads to two cytokines (18,19). p38 is a non-synthetase component of the multi-enzyme complex and encodes the 320 amino acid protein with no homolog in yeast, bacteria and archaea. It is moderately hydrophobic protein and displays putative leucine-zipper motif. It associates with p43 and various AARSs (20,21).

Recently evidence has been provided for aminoacylation of tRNA in the nucleus of Xenopus laevis, which implies that AARSs exist in the nucleus (22,23). With nuclear localisation signal (NLS), AARSs bind to the importins and translocate to the nucleus through an energy dependent step. Analysis of the databases enabled us to identify NLS in 15 cytoplasmic synthetases from Saccharomyces cerevisiae. Although uncharged tRNA is also exported, it seems that nuclear acylation makes the export more efficient (24). It suggests that a proofreading mechanism ensures correct maturation and export of aminoacylated tRNA (25).

A broader evolutionary interest in the AARSs stems from their biological function being one of the core requirements from progression from the RNA world to the universal common ancestor in numerous schemes for the origin of life (26). In the Methanococcus janaschii genome, 16 AARS gene homologs have been identified. The genes of GlnRS, AsnRS, LysRS and CysRS are missing. In addition the assignment of an open reading frame encoding SerRS was based on rather low homology, while there appeared to be three reasonable candidates to encode the two subunits of PheRS. Components of the selenocysteinyl-tRNA synthesis pathway could not be identified (27).

It is generally assumed that an AARS of given specificity will always belong to the same class regardless of its biological origin, reflecting the ancient evolution of this enzyme family. The only known exceptions to this rule are the LysRSs, which are composed of two unrelated families, class I enzymes in certain archaea and bacteria and class II enzymes in all other organisms. Although the class I and class II LysRSs are similar, the same sites in their tRNA substrates, they differ in their exact mechanism of lysine activation (28,29). The class I LysRS binds to the minor groove of the helix, whereas the class II enzyme approaches from the major groove side. The class II LysRS of Escherichia coli accommodates major groove determinants encoded by G-C, but is blocked by the presence of a G-U pair. In contrast the class I LysRS of Borrelia burgdorferi accepts the minor groove determinants of G-C or G-U. The blocking determinants for a class II LysRS could explain its displacement by its class I counterpart (28,29). In some instances, the AARSs are involved in auto-regulation of their expression on the level of translation, by binding the tRNA-like structures within the mRNA. In E.coli, biosynthesis of ThrRS is autoregulated by a feedback mechanism at the translational level. The enzyme binds to the leader of its own mRNA close to the translation initiation site, thereby inhibiting ribosome attachment and, consequently, translation. The translational operator of the thrS gene contains two stem–loop structures that both mimic the anticodon arm of E.coli tRNA^Thr and interact with the monodimeric ThrRS. The fact that tRNA^Thr and the mRNA substrate share common specificity determinants argues in favor of structural analogies between the two recognition modes. The macromolecular mimicry model predicts that the two anticodon-like domains of the thrS operator recognise the anticodon-binding domain of each monomer of ThrRS. It accounts for competition between both RNAs for ThrRS binding and therefore links aminoacylation and regulation in a coherent manner (30). The other example of mimicry can be identified between E.coli ribosomal protein L25 and GlnRS. Tertiary structure comparisons indicate that the six-stranded β-barrels of L25 and the tRNA anticodon-binding domain of glutaminyl-tRNA synthetase are similar. A recent model for the evolution of the GlnRS-GluRS family suggests the acquisition of GlnRS for a subgroup of the eubacterial kingdom via horizontal gene transfer from a eukaryotic source. Therefore it is possible that the six stranded β-barrel, which constitutes an RNA-binding fold for L25, and the anticodon-domain of GlnRS, represent a case of convergent evolution (31).

Several AARSs have been found to be autoantigens for a subgroup of patients with the idiopathic inflammatory myopathies, polymyositis and dermatomyositis. Autoantibodies against synthetases are found almost exclusively in these cases, with patients generally having antibodies against only one synthetase. Most commonly, the antibodies are directed against HisRS, labelled ‘anti-Jo-1’ autoantibodies, but the antibodies to threonyl-, asparaginyl-, alanyl- or glycyl-tRNA synthetases or the multi-enzymatic complex have also been found (32).

In Drosophila melanogaster, expression of the Sex-lethal (Sxl) gene dictates the choice between male and female development. It has been shown that successful activation of Sxl requires both maternally- and zygotically-provided gene products, many of which are essential for viability and have other, non-sex-specific functions. One of a new maternally expressed gene l(2)49Dh encodes aspartyl-tRNA synthetase. Furthermore, mutations in tryptophanyl-tRNA synthetase are also dosage-sensitive maternal modifiers of Sxl. These data suggest that stable activation of Sxl in the embryo may be particularly sensitive to change in the translation machinery (33). It has been found that TrpRS is expressed to high levels in the developing Drosophila salivary gland (34). The AARS has been used as a model for screening of synthetase-directed drugs (35).

DESCRIPTION AND AVAILABILITY OF THE DATABASE

The AARS Database is a collection of amino acid sequences of all AARSs published to date (as of August 1999). The database entries are based on the EMBL/SWISS-PROT format. In addition to the amino acid sequences, they include SWISS-PROT sequence name and accession number, short description of the sequence, organism name and its taxonomic classification, as well as basic bibliographic information. Since most of the AARS primary structures were determined on the nucleic acid level the appropriate accession numbers of the related entries in nucleotide sequences databases (EMBL/GenBank, TIGR) are also included. The availability of 3D structural data is indicated by cross-references to the Brookhaven Protein Data Base. The data included in the AARS database also contain partial sequences that might be useful for some comparative and evolutionary studies. According to the original SWISS-PROT description, some of the entries have been marked as putative or probable.

Currently the database contains 819 amino acid sequences of the cytoplasmic and organelle synthetases from a variety of organisms.
The Aminoacyl-tRNA Synthetase Database can be accessed on the WWW at http://rose.man.poznan.pl/aars/index.html. To make retrieval of the data as quick as possible, each individual sequence in the database is stored as a separate file. The sequences are grouped according to the AARS amino acid specificity or organism.

ACKNOWLEDGEMENT

This work was supported by a grant from the Polish State Committee for Scientific Research.

REFERENCES