The Aminoacyl-tRNA Synthetase Data Bank (AARSDB)

Maciej Szymanski and Jan Barciszewski*

Institute of Bioorganic Chemistry of the Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

Received September 3, 1998; Accepted October 1, 1998

ABSTRACT

Aminoacyl-tRNA synthetases (AARSs) are the key components of the protein biosynthesis machinery. They are responsible for maintaining the fidelity of transfer of genetic information from DNA into protein. The database is a compilation of amino acid sequences of all aminoacyl-tRNA synthetases known to date. It contains 422 primary structures of the AARSs available as separate entries or alignments of related proteins. The database is available via the World Wide Web at http://rose.man.poznan.pl/aars/index.html

INTRODUCTION

The universal genetic code is established in a single biochemical reaction—the aminoacylation of transfer ribonucleic acids (tRNAs). The reaction is carried out by aminoacyl-tRNA synthetases (AARS). Aminoacyl-tRNA synthetases are, in fact, the only components of the gene expression machinery that function on the interface between the nucleic acids and proteins. There are three particularly interesting aspects of AARSs research. First, the mechanism of amino acid recognition and activation, second, the specificity of tRNA recognition and third, their origin and evolution (1). The synthetases constitute a family of 20 enzymes that are responsible for the specific esterification of tRNAs with their cognate amino acids, and thus are essential in maintaining the fidelity of the protein biosynthesis process. In Prokaryota there are some deviations of this number which share common core structure and tRNAs have the same basic function on the interface between the nucleic acids and proteins. Although AARSs catalyse the same basic reaction and share a common catalytic function, the AARSs have long been known to differ in the size, amino acid sequences and subunit structure. The quaternary structures of synthetases include monomers (α), homodimers and tetramers (α2, α4) and heterotetramers (α2β2). The peptide size of the subunits in Escherichia coli varies from 344 aa for TrpRS to 951 for ValRS. The eukaryotic enzymes are usually larger than their prokaryotic counterparts. This is due to the presence of C- and N-terminal extensions that are dispensable for the aminoacylation, but their function is still unclear (7,8).

In spite of their common catalytic function, the AARSs have long been known to differ in the size, amino acid sequences and subunit structure. The quaternary structures of synthetases include monomers (α), homodimers and tetramers (α2, α4) and heterotetramers (α2β2). The peptide size of the subunits in Escherichia coli varies from 344 aa for TrpRS to 951 for ValRS. The eukaryotic enzymes are usually larger than their prokaryotic counterparts. This is due to the presence of C- and N-terminal extensions that are dispensable for the aminoacylation, but their function is still unclear (7,8).

Although AARSs catalyse the same basic reaction and share a common substrate (ATP and cofactor (magnesium)), they form a quite diverse group of enzymes. On the other hand, amino acids share common core structure and tRNAs have the same basic...
fold, that allows them to be recognized by other components of the protein biosynthesis pathway. The amino acids are attached to the same 3′-end of tRNA.

Structural and sequence analyses of all AARSs clearly shows, that there are two exclusive classes (class I and class II) of enzymes. This shows that two distinct structural frameworks evolved independently to perform the aminoacylation reaction. The catalytic domain of class I enzymes is formed by the so-called Rossmann fold, first recognised as a nucleotide binding element. On the other hand, class II enzymes have a novel antiparallel fold that was identified for the first time in the structure of SerRS (7,9). Apart from the different ATP-binding motifs, the two classes of synthetases differ in their mode of tRNA binding. The crystallographic studies of AARSs in free and complexed forms allow us to gain insight into the specificity of substrate recognition and the catalysis itself (7).

In contrast to prokaryotic synthetases, the eukaryotic enzymes are often found to be involved in forming supracomplexes through self-assembly or association with other protein synthesis machinery components and cellular structures (10). They are often components of multisynthetase complexes, that also include several proteins of unknown function (11). The exact structure and composition of these complexes is controversial. Different forms have been isolated, depending on the purification method (10). The understanding of the structure of these multiienzymatic complexes is important to find the functional link between the aminoacylation and other cellular processes.

In addition to the aminoacylation reaction, some AARSs have been also found to be involved in other cellular processes. Glycyl-tRNA synthetase has been shown to be responsible for the synthesis and turnover of diadenosine tetraphosphate (Ap4A) that plays an important role in a response of bacterial and eukaryotic cells to a variety of stress conditions (12,13). The diadenosine oligophosphates [Ap(n)A] including Ap4A have been recently shown to function as a new class of signalling molecules within the cell (14,15). Mitochondrial TyrRS from *Neurospora crassa* has been shown to be a key component for the splicing of group I intron of pre-rRNA, by substitution of the missing RNA domain of this otherwise self-splicing intron (16). In some instances, the AARSs are involved in autoregulation of their expression on the translation level, by binding the tRNA-like structures within the mRNA (17). Several AARSs have been found to be autoantigens for a subgroup of patients with the idiopathic inflammatory myopathies, polymyositis and dermatomyositis (18). Autoantibodies against synthetases are found almost exclusively in these cases, with patients having antibodies generally against only one synthetase. Most commonly, the antibodies are directed against HisRS, labelled ‘anti-Jo-1’ autoantibodies, but the antibodies to threonyl-, alanyl- or glycyl-tRNA synthetases or the multienzymatic complex have also been found (18).

Recently, a subset of familial and sporadic amyotrophic lateral sclerosis cases have been found to be associated with mutations in the AARS genes, suggesting a role for these enzymes in the pathogenesis of the disease (19). These findings highlight the importance of understanding the role of AARSs in cellular processes and their potential involvement in neurological disorders.
in the gene encoding Cu, Zn superoxide dismutase (SOD1), that binds lysyl-tRNA synthetase (19).

DESCRIPTION OF THE DATABASE

The AARS Database is a collection of amino acid sequences of all AARSs published to date (August 1998). The database entries are based on the EMBL/SWISS-PROT format (Table 1). 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 423 amino acid sequences of cytoplasmic and organelle synthetases from a variety of organisms. The summary of the database content, showing the numbers of primary structures for given amino acid specificity is presented in Table 2.

Most of the data in the database come from genome sequencing projects. It is interesting to note that in some cases the genes encoding some of the AARS have not been identified in the whole genomes. The best example is a genome of M.jannaschii, in which only 17 AARS genes have been identified so far. The summary of synthetases found in some of the sequenced prokaryotic chromosomes is presented in Table 3.

AVAILABILITY OF THE DATABASE

The Aminoacyl-tRNA Synthetase Database can be accessed on the World Wide Web at the URL http://rose.man.poznan.pl/aars/index.html. To make the 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.

Any comments or suggestions concerning the database are welcome. Please contact us via Email at: jbarcisz@ibch.poznan.pl (Jan Barciszewski) or mszyman@ibch.poznan.pl (Maciej Szymanski).

ACKNOWLEDGEMENT

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