Aminoacyl-tRNA synthetases database

Maciej Szymanski, Marzanna A. Deniziak and Jan Barciszewski*

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

ABSTRACT

Aminoacyl-tRNA synthetases (AARSs) are at the center of the question of the origin of life. They constitute a family of enzymes integrating the two levels of cellular organization: nucleic acids and proteins. AARSs arose early in evolution and are believed to be a group of ancient proteins. They are responsible for attaching amino acid residues to their cognate tRNA molecules, which is the first step in the protein synthesis. The role they play in a living cell is essential for the precise deciphering of the genetic code. The analysis of AARSs evolutionary history was not possible for a long time due to a lack of a sufficiently large number of their amino acid sequences. The emerging picture of synthetases' evolution is a result of recent achievements in genomics [Woese,C., Olsen,G.J., Ibba,M. and Söll,D. (2000) *Microbiol. Mol. Biol. Rev.*, 64, 202–236]. In this paper we present a short introduction to the AARSs database. The updated database contains 1047 AARS primary structures from archaebacteria, eubacteria, mitochondria, chloroplasts and eukaryotic cells. It is the compilation of amino acid sequences of all AARSs known to date, which are available as separate entries via the WWW at http://biobases.ibch.poznan.pl/aars/.

INTRODUCTION

Most living cells possess a set of 20 aminoacyl-tRNA synthetases (AARSs), specifically charging their cognate tRNAs (1,2). Although some exceptions to the '20 AARSs' rule were found in bacteria, archaea and eukaryotic organelles (2), those were assumed to be just an evolutionary anomaly. However, recent discoveries arising mainly from functional genomics studies in bacterial and archaeal systems contradict these expectations. Numerous organisms do not possess a full complement of 20 AARSs. In many cases a reduced number of synthetases is supported by a variety of novel enzymes and pathways providing the full complement of aminoacyl-tRNAs required for protein synthesis (3).

The archaebacterial cysteinylation pathway is one of the recent discoveries in that field. It is known that the genome sequences of certain *Archaea* do not code for canonical cysteinyl-tRNA synthetases (CysRSs). The first proposed candidate for Cys-tRNA^{Cys} formation was seryl-tRNA synthetase. It was assumed that serine mischarged by SerRS could be potentially converted to cysteine. However, it has been found that SerRS from *Methanobacterium thermoautotrophicum* and *Methanococcus maripaludis* do not charge tRNA^{Cys} with serine (4). Surprisingly, cysteinyl-tRNA synthesis in *Methanococcus jannaschii* is mediated by the enzyme whose amino acid sequence is homologous to ProRS (5,6). Biochemical and genetic analyses indicated that the archaeal form of ProRS synthesizes both cognate cysteinyl-tRNA^{Cys} and prolyl-tRNA^{Pro}. This AARS of double specificity contains only one set of sequence motifs characteristic for class II enzymes. The whole protein is relatively small, ~53 kDa. Cloning of the enzyme in *Escherichia coli* shows that a single subunit supports both activities. Cross inhibition between specific proline and cysteine substrates also suggests a linkage or overlap between their active sites. It seems that the new archaeal enzyme is unlikely to be a multidomain peptide, in contrast to metazoan GluProRS. However, there is a controversy concerning cysteinylation and prolylation activity levels as well as tRNA^{Cys} requirement for activation of cysteine (5,6). It was suggested that the *M.jannaschii* ProRS might recruit additional protein or RNA factors to facilitate cysteinylation (6). Nevertheless, the ability of a single AARS to synthesize two aminoacyl-tRNAs raises questions about substrate specificity in protein synthesis and may provide insights into the evolutionary origin of the charging reaction (5–7). That kind of dual activity is observed not only in *Archaea*. The eukaryote *Giardia lamblia* also contains a ProRS with CysRS activity (2).

The recognition of a specific tRNA by its cognate synthetase depends on structural features of both molecules (8,9). There are numerous examples where tRNAs undergo large conformational changes upon binding to enzymes but little is known about conformational rearrangements in the tRNA-complexed AARSs. Recently the crystal structure of the dimeric class II aspartyl-tRNA synthetase (AspRS) from yeast was solved at 2.3 Å resolution. Its free form was compared with that of the protein associated with the cognate tRNA^{Asp} (10). Within the complex the AspRS undergoes some changes in the anticodon and catalytic domains of the enzyme subunits. The most significant differences are observed in the two evolutionary conserved loops, which are critical for ligand binding (10).

Interestingly certain AARSs, namely MetRS, IleRS and LeuRS, activate *in vivo* not only cognate amino acids but also the non-protein amino acid homocysteine (Hcy), due to its structural similarity to Met, Ile and Leu. Thus, Hcy enters the first steps of protein synthesis, but is never incorporated into the protein because of the editing mechanism of AARSs converting misactivated Hcy into thiolactone. Recently it was shown that reversibly modified Hcy (*S*-nitroso-Hcy) could bypass the editing step and be translationally incorporated into

*To whom correspondence should be addressed. Tel: +48 61 852 8503; Fax: +48 61 852 0532; Email: jbarcisz@ibch.poznan.pl

the protein by *E.coli* MetRS (11). This finding has some important medical implications, because the metabolic conversion of homocysteine to thiolactone and protein homocysteinylation may play a role in vascular damage (12).

The evolutionary history of AARS genes is difficult to establish due to numerous examples of gene duplications, fusions and horizontal transfers (13–15). Prokaryotes contain a set of 18–20 AARSs, but eukaryotes also possess their organellar analogs (16–18). Like mitochondria, plastids lost most of their original genes and some of them have been transferred to the nucleus (16). There are several plant AARSs for which one gene has been replaced by another one of different origin, but carrying the same function. There are two recently described AARSs among them: *Arabidopsis thaliana* asparaginyl-tRNA synthetase (AsnRS) and CysRS (19). Both cytosolic enzymes are very similar to their mitochondrial and plastidic counterparts, resulting in closely related AARSs in all three compartments. *Arabidopsis* genome analysis revealed the presence of four genes coding for AsnRSs and two sequences corresponding to CysRSs. Moreover, phylogenetic studies suggest that all of them have organellar origin and are thought to be the best examples known to date of the plastidic AARSs captured by the cytosolic protein synthesis machinery (19).

Multicellular eukaryotes contain a macromolecular assembly of nine AARSs and three auxiliary proteins. One of them, p43, is of special interest because it is a precursor of endothelial-monocyte-activating polypeptide II (EMAP II), an inflammatory cytokine involved in apoptosis (20). Recently the EMAP II portion of p43 has been localized within the multisynthetase complex isolated from rabbit reticulocytes (21). Immunoblot analysis has clearly demonstrated a strong reaction of anti-EMAP II antiserum with p43 as well as cross-reactivity with isoleucyl-tRNA synthetase. Electron microscopy images of the immunocomplexes show two antibody-binding sites: one is near the center of the multisynthetase complex, at the intersection of the arms with the base, the second is in the base of the particle. Those data allow a refinement of the three-domain model of polypeptide distribution within the multisynthetase complex. Moreover, the central location of the p43/EMAP II suggests that this polypeptide plays a role in optimizing the normal function and in a rapid disruption of the essential cellular machinery when apoptosis is signaled (21).

A great help in understanding diverse biological activities as well as the maturation process of EMAP II came from its crystal structure solved at 1.8 Å resolution (22). It revealed the tRNA-binding fold and a domain that is structurally homologous to other chemokines. The structures similar to the EMAP II RNA-binding motif were previously observed in the anticodon-binding domain of yeast AspRS and the B2 domain of *Thermus thermophilus* PheRS. The RNA-binding pattern of EMAP II is likely to be non-specific in contrast to the AspRS of *Saccharomyces cerevisiae*.

An assembly of AARSs complex seems to be mediated by the heat shock protein 90 (Hsp 90), which binds to human glutamyl-prolyl-tRNA synthetase (GluProRS) (23). This interaction is sensitive to the Hsp 90 inhibitor, geldanamycin and ATP. It is targeted to the region of three tandem repeats linking the two catalytic domains of GluProRS. The GluProRS interacts also with IleRS. This process depends on the activity of Hsp 90, implying that the association is mediated by the chaperone. It was found that Hsp 90 binds preferentially to most of the enzymes located in the complex, rather than to those that are not found there. In addition, inactivation of Hsp 90 interfered with the *in vivo* incorporation of the nascent AARSs into the multi-AARS complex. Thus, Hsp 90 appears to mediate protein–protein interactions of mammalian tRNA synthetases associating with their subset during complex formation (23). Concerning binding of Hsp 90 to GluProRS one should mention that the recently published NMR structure of one repeated motif from the linker part of the synthetase points to its possible interactions with RNA (24). This repeat is built around an antiparallel coiled-coil with the conserved lysine and arginine residues, which form a basic path on one side of the structure. This motif can be a docking site for nucleic acids. Gel retardation and filter-binding experiments confirmed the real interaction of the studied domain with RNA. Thus, it has been suggested that the GluProRS repeated motif might represent a novel type of general RNA-binding domain appended to eukaryotic AARSs, to serve as a *cis*-acting tRNA-binding cofactor (24).

The existence of AARSs in the nucleus was implied by a study on nuclear aminoacylation of tRNA (25–27). However, there was almost no information regarding the status of active AARSs within the nuclei of eukaryotic cells. Recently at least 13 active AARSs have been found in the purified nuclei of both CHO and rabbit kidney cells, although their steady-state levels represent only a small percentage of those found in the cytoplasm (28). Most interestingly, all the nuclear aminoacyl-tRNA synthetases examined so far can be isolated as art of a multienzyme complex that is more stable, and consequently larger, than the comparable complex isolated from the cytoplasm. These data directly demonstrate the presence of active AARSs in mammalian cell nuclei. Moreover, their unexpected structural organization raises questions concerning the functional significance of those multienzyme complexes and their role in nuclear to cytoplasmic transport of tRNAs. The role played by AARSs present in the nucleus is not limited to tRNA maturation and/or export control. Human methionyl-tRNA synthetase (MetRS) was shown to be translocated into the nucleolus in proliferating cells (29). This process depends on the integrity of RNA and the activity of RNA polymerase I in the nucleolus. The treatment of MetRS with specific anti-MetRS antibodies decreased ribosomal RNA synthesis. Thus, human MetRS is thought to play a role in the biogenesis of rRNA in nucleoli, while it is catalytically involved in protein synthesis in the cytoplasm.

There are many data showing that the functional units (modules) of AARSs are arranged along their sequence in a linear fashion. This seems to be a result of a long period of evolution, because numerous examples of synthetase-like proteins have been identified in recently sequenced genomes (30). Those proteins consist mostly of just one synthetase domain. A search of the current sequence databases enabled identification of alanyl-, aspartyl-, glutamyl-, glycyl-, histidyl-, lysyl-, methionyl-, phenylalanyl-, seryl- and tyrosyl-tRNA synthetase-like proteins (30). Some of them are limited to a single kingdom of the tree of life, whereas others are more widely distributed. Their genes are probably related to AARSs through gene duplication events, although details of these processes probably vary from one system to another (30).

DESCRIPTION AND AVAILABILITY OF THE DATABASE

The AARSs database is the collection of amino acid sequences of all published AARSs. Currently it contains 1047 primary structures of cytoplasmic and organellar AARSs from various organisms. The entries are grouped according to AARS amino acid specificity. They are based on EMBL/SWISS-PROT format. Each includes the AARS amino acid sequence, its SWISS-PROT name and the accession number, a short description of the sequence, its source (organism name with taxonomic classification) and bibliographic information. For the enzymes whose sequences were determined at the nucleotide level, the appropriate EMBL/GenBank or TIGR entries are included, and for those with already known 3D structure, the cross-references to the Brookhaven Protein Data Base are indicated. The partial sequences of AARSs are also included in the database. According to the original SWISS-PROT description, some of the entries have been marked as putative or probable. The AARSs database is available on the WWW at http://biobases.ibch.poznan.pl/aars/.

REFERENCES

1. Crick,F.H.C. (1958) On protein synthesis. *Symp. Soc. Exp. Biol.*, **12**, 138–163.
2. Ibba,M., Becker,H.D., Stathopoulos,C., Tumbula,D. and Söll,D. (2000) The Adaptor hypothesis revisited. *Trends Biochem. Sci.*, **25**, 311–316.
3. Woese,C., Olsen,G.J., Ibba,M. and Söll,D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.*, **64**, 202–236.
4. Kim,H.-S., Vothknecht,U.C., Hedderich,R., Celic,I. and Söll,D. (1998) Sequence divergence of seryl-tRNA synthetases in *Archaea. J. Bacteriol.*, **180**, 6446–6449.
5. Stathopoulos,C., Li,T., Longman,R., Vothknecht,U.C., Becker,H.D., Ibba,M. and Söll,D. (2000) One polypeptide with two aminoacyl-tRNA synthetase activities. *Science*, **287**, 479–482.
6. Lipman,R.S.A., Sowers,K.R. and Hou,Y.-M. (2000) Synthesis of cysteinyl-tRNA^Cys by a genome that lacks the normal cysteine-tRNA synthetase. *Biochemistry*, **39**, 7792–7798.
7. Yarus,M. (2000) Unraveling the riddle of ProCys-tRNA synthetase. *Science*, **287**, 440–441.
8. Beuning,P.J. and Musier-Forsyth,K. (1999) Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers*, **52**, 1–28.
9. Sherlin,L.D., Bullock,T.L., Newberry,K.J., Lipman,R.S.A., Hou,Y.-M., Beijer,B., Sproat,B.S. and Perona,J. (2000) Influence of transfer RNA tertiary structure on aminoacylation efficiency by glutaminyl and cysteinyl-tRNA synthetases. *J. Mol. Biol.*, **299**, 431–446.
10. Sauter,C., Lorber,B., Cavarelli,J., Moras,D. and Giegé,R. (2000) The free yeast aspartyl-tRNA synthetase differs from the tRNA^Asp-complexed enzyme by structural changes in the catalytic site, hinge region, and anticodon-binding domain. *J. Mol. Biol.*, **299**, 1313–1324.
11. Jakubowski,H. (2000) Translational incorporation of S-nitrosohomocysteine into protein. *J. Biol. Chem.*, **275**, 21813–21816.
12. Jakubowski,H., Zhang,L., Bardeguez,A. and Aviv,A. (2000) Homocysteine thiolactone and protein homocysteinylation in human endothelial cells. *Circ. Res.*, **87**, 45–51.
13. Brown,J.R. and Doolittle,W.F. (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc. Natl Acad. Sci. USA*, **92**, 2441–2445.
14. Diaz-Lazcoz,Y., Aude,J.-C., Nitschké,P., Chiapello,H., Landès-Devauchelle,C. and Risler,J.-R. (1998) Evolution of genes, evolution of species: the case of aminoacyl-tRNA synthetases. *Mol. Biol. Evol.*, **15**, 1548–1561.
15. Wolf,Y.I., Aravind,L., Grishin,N.V. and Koonin,E.V. (1999) Evolution of aminoacyl-tRNA synthetases – analyses of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res.*, **9**, 689–710.
16. Mireau,H., Lancelin,D. and Small,I.D. (1996) The same *Arabidopsis* gene encodes both cytosolic and mitochondrial alanyl-tRNA synthetases. *Plant Cell*, **8**, 1027–1039.
17. Akashi,K., Grandjean,O. and Small,I. (1998) Potential dual targeting of an *Arabidopsis* archaebacterial-like histidyl-tRNA synthetase to mitochondria and chloroplasts. *FEBS Lett.*, **431**, 39–44.
18. Menand,B., Marechal-Drouard,L., Sakamoto,W., Dietrich,A. and Wintz,H. (1998) A single gene of chloroplast origin codes for mitochondrial and chloroplastic methionyl-tRNA synthetase in *Arabidopsis thaliana. Proc. Natl Acad. Sci. USA*, **95**, 11014–11019.
19. Peeters,N.M., Chapron,A., Giritch,A., Grandjean,O., Lancelin,D., Lhomme,T., Vivrel,A. and Small,I. (2000) Duplication and quadruplication of *Arabidopsis thaliana* cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin. *J. Mol. Evol.*, **50**, 413–423.
20. Quevillon,S., Agou,F., Robinson,J.-C. and Mirande,M. (1997) The p43 component of the mammalian multi-synthetase complex is likely to be the precursor of the endothelial-monocyte-activating polypeptide II cytokine. *J. Biol. Chem.*, **272**, 32573–32579.
21. Norcum,M.T. and Warrington,A. (2000) The cytokine portion of p43 occupies a central position within the eykaryotic multisynthetase complex. *J. Biol. Chem.*, **275**, 17921–17924.
22. Kim,Y., Shin,J., Li,R., Cheong,C., Kim,K. and Kim,S. (2000) A novel anti-tumor cytokine contains an RNA binding motif present in aminoacyl-tRNA synthetases. *J. Biol. Chem.*, **275**, 27062–27068.
23. Kang,J., Kim,T., Ko,Y.-G., Rho,S.B., Park,S.G., Kim,M.J., Kwon,H.J. and Kim,S. (2000) Heat shock protein 90 mediates protein-protein interactions between human aminoacyl-tRNA synthetases. *J. Biol. Chem.*, **275**, 31682–31688.
24. Cahuzac,B., Berthonneau,E., Birlirakis,N., Guittet,E. and Mirande,M. (2000) A recurrent RNA-binding domain is appended to eukaryotic aminoacyl-tRNA synthetases. *EMBO J.*, **19**, 445–452.
25. Hooper,A.K. (1998) Nuclear function charge ahead. *Science*, **282**, 2003–2004.
26. Lund,E. and Dahlberg,J.E. (1998) Proofreading and aminoacylation of tRNAs before export from the nucleus. *Science*, **282**, 2082–2085.
27. Sarkar,S., Azad,A.K. and Hooper,A.K. (1999) Nuclear tRNA aminoacylation and its role in nuclear export of endogenous tRNAs in *Saccharomyces cerevisiae. Proc. Natl Acad. Sci. USA*, **96**, 14366–14371.
28. Nathanson,L. and Deutscher,M.P. (2000) Active aminoacyl-tRNA synthetases are present in nuclei as high molecular weight multienzyme complex. *J. Biol. Chem.*, **275**, 31559–31562.
29. Ko,Y.-G., Kang,Y.-S., Kim,E.-K., Park,S.G. and Kim,S. (2000) Nucleolar localization of human methionyl-tRNA synthetase and its role in ribosomal RNA synthesis. *J. Cell Biol.*, **149**, 567–574.
30. Schimmel,P. and de Pouplana,L.R. (2000) Footprints of aminoacyl-tRNA synthetases are everywhere. *Trends Biochem. Sci.*, **25**, 207–209.