Aminoacyl-tRNA synthetases are the link between the worlds of protein and nucleic acid

Department of Structural Biology

■ **Mark Safro**

Naama Kessler,
Milana Frenkel-Morgenstern,
Igal Finarov, Inbal Mermershtain,
Liron Klipcan

☎ 972 8 934 3320
FAX 972 8 934 4136
@ mark.safro@weizmann.ac.il
✦ www.weizmann.ac.il/sb/faculty_pages/Safro/

The major avenue of investigations in our laboratory is the family of proteins that play pivotal role in translating the genetic code, i.e. the aminoacyl-tRNA synthetases (aaRSs). Protein structure-function-evolution, amino acid synthesis, editing, signal transduction pathways, multi-protein complexes, gene regulation, RNA splicing, RNA modification, RNA recognition, origin of life and drug discovery are all intimately related to the aaRSs. The aaRSs are a notoriously diverse family of enzymes varying considerably in primary sequence, subunit size and oligomeric organization. Their structures and functions, which have both practical and basic significance, are deserving of and have received much attention. However, it is not only the structure-function aspect of these enzymes that has captured the biologist's imagination; it is also the possibility that they could tell us about the history of the genetic code evolution.

Structural basis for discrimination of L-phenylalanine from L-tyrosine by PheRS and incorporation of non-natural amino acids within the active site of aaRSs.

Phenylalanyl-tRNA synthetase (PheRS) is known to be among the most complex and large enzymes of the aaRS family. The subunits organization of prokaryotic and eukaryotic cytoplasmic PheRS is markedly conserved during the evolution and forms functional tetramers in all known species. The 3-dimensional structure of the *Thermus thermophilus* PheRS has been determined both in apo-form and with functional substrates phenylalanine, Phe-AMP and tRNA^{Phe} and their analogs (Fishman et al., 2001; Moor et al., 2006; Mosyak et al., 1995). These, structural studies provided the basis for precisely understanding the $(\alpha\beta)_2$ subunit organization of PheRS, a dimer built of two heterodimers. The α-subunit domains create catalytic module and together with N-terminal coiled-coil domain directly involved in aminoacylation and tRNA^{Phe} binding, whereas major function of β-subunit consists in recognition and binding of tRNA^{Phe}. Structural analysis of the bacterial cytoplasmic PheRS complexed with tRNA^{Phe} further showed that one tRNA^{Phe} molecule interacts with all four subunits of the enzyme and thus, accounts for the enzyme to be a functional $(\alpha\beta)_2$ dimer (Figure 1A).

The accuracy of the genetic code translation is vitally important for living cells function, intimately associated with the proper operation of correctly assembled macromolecules. The aaRSs play a crucial role in the maintenance of the faithful translation, promoting close control over two-step aminoacylation reaction. Rates of misincorporation of non-cognate amino acids *in vivo* of approximately one out of 10^4-10^5 reactions. This brings up the question, how aaRSs can achieve this level of fidelity, and distinguish between amino acids closely resembling each other in their chemical structures? As it turned out, proofreading activity has been evolved by certain aaRSs and it is associated with a distinct active site where misaminoacylated tRNA are hydrolyzed. Even in a limited set of aaRSs possessing editing activity, each enzyme demonstrates its idiosyncratic character.

Bacterial PheRS demonstrates a significant degree of natural plasticity in the active site; amino acid binding pocket accommodates not only the cognate amino acid, phenylalanine but also the non-cognate tyrosine and unnatural derivative p-Cl-phenylalanine. Both bacterial and eukaryotic cytoplasmic PheRSs have been shown to misactivate L-tyrosine but do not attach this amino acid stably to tRNA^{Phe}. We demonstrate the capability of the active site to activate tyrosine (Figure 1B) (Kotik-Kogan et al., 2005). The editing site is localized at the β subunits, 35 Å from the active site. The residues involved in hydrolysis of tyrosine were clearly detectable in electron density map (Figure 1C). The editing site of PheRS specifically rejects Tyr only, whereas the other naturally occurring amino acids are efficiently discriminated against at the synthetic active site. The active site plasticity of PheRS and subsequent misacylation of the cognate tRNAs is a feasible way of
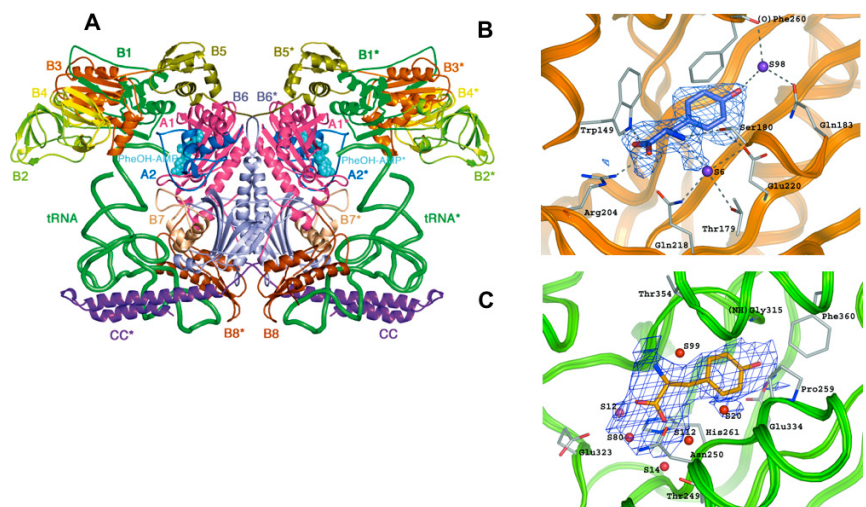
Fig. 1 The structure of T. thermophilus PheRS. (A) Overall structure of the bacterial PheRS. The structural domains associated with the symmetry-related heterodimer (colored with the same colors) are denoted with asterisks. (B) The synthetic active site of the T. thermophilus PheRS (colored orange) with bound Tyr (colored slate). (C) PheRS editing site (colored green) with bound Tyr (colored orange). The protein residues participating in direct and water-mediated (red spheres) contacts are shown.
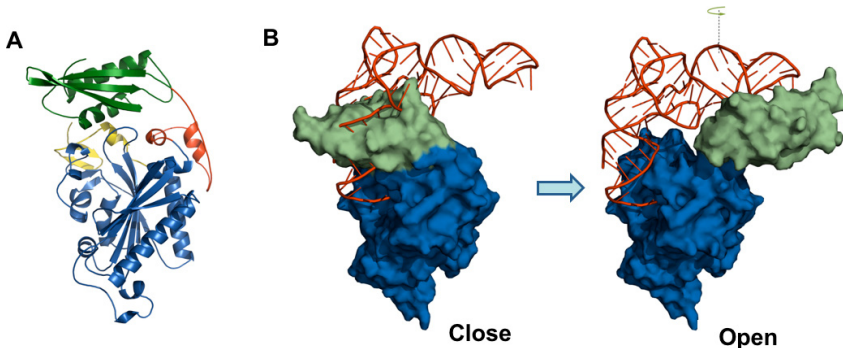
Fig. 2 *The structure of human mitPheRS. (A) Overall structure of the human mitPheRS enzyme. (B) Rotation of anticodon binding domain (ABD). The position of the tRNAPhe is obtained after superimposing the mitochondrial onto the bacterial catalytic domain (closed conformation). The ABD is colored pale-green, and the tRNA is colored red. This Figure demonstrates that a hinge-type movement (indicated by an arrow) of the mitPheRS ABD would be required to reach the anticodon of tRNAPhe (open conformation).*

in vivo incorporation of non-canonical amino acids into protein.

The structure of Human Mitochondrial Phenylalanyl-tRNA Synthetase

In eukaryotes, protein synthesis occurs not only in the cytoplasm, but also in different organelles, such as mitochondria and chloroplasts. In the majority of organisms, mitochondrial aaRSs (mit-aaRSs) are encoded in the nucleus and post-translationally transported into the organelle. The understanding of the unique molecular features of the mit-aaRSs is even more important in view of recent findings, showing that over 200 mutations associated with pathogenecity have been identified within human mitochondrial genome. More than half of these disease related mutations are located within tRNA genes, a remarkable trend given that tRNA sequences comprise only 10% of the mitochondrial genome. This can be explained by the essential role of tRNAs in the synthesis of proteins involved in energy metabolism.

We recently solved the structure, at 2.2Å resolution, of a human monomeric mitPheRS complexed with Phe-AMP (Figure 2A) (Levin et al., 2007). Human mitochondrial phenylalanyl-tRNA synthetase (mitPheRS) is the smallest known nuclear encoded synthetase exhibiting homology to bacterial PheRSs. However, while the total length of the bacterial enzyme is made up of 2270 residues and have hetero-tetramer organization $(\alpha\beta)_2$, the mature mitPheRS homolog is a single-chain enzyme consisting of 415 amino acids, and in fact is a chimera of the bacterial catalytic subunit and the anticodon binding domain (ABD) from bacterial PheRS. Thus the previously described heterodimeric structure is not a prerequisite for the phenylalanylation activity, as monomeric mitPheRS is also fully active.

Also the structure of functional mitPheRS- tRNA[Phe] yet not available, modeling studies demonstrate that the anticodon-binding domain (ABD) of mitPheRS overlaps with the acceptor stem of tRNA[Phe] when the substrate is positioned similar to that seen in the binary *T. thermophilus* complex. Thus, formation of the PheRS-tRNA[Phe] complex in human mitochondria should accompany by considerable rearrangement of the ABD and a hinge-type rotation through ~160° is the relevant functional motion of the ABD upon tRNA binding (Figure 2B).

The macromolecular tRNA-
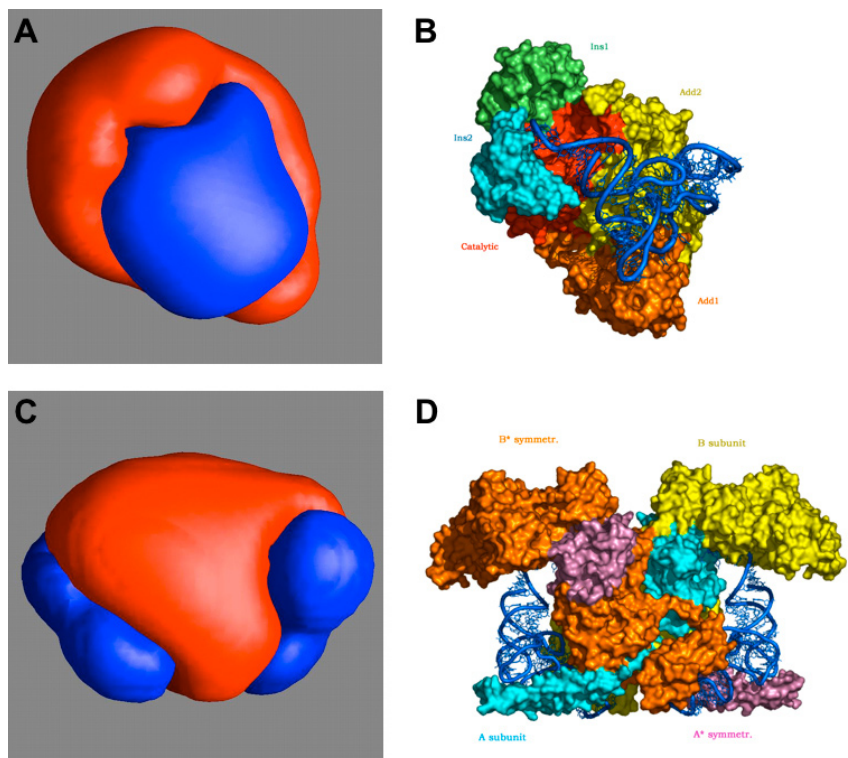
Fig. 3 *3D-isopotential surface representation of aaRSs and overall views of their complexes with cognate tRNAs. (A, B) Arginyl-tRNA synthetase. (C, D) Phenylalanyl-tRNA synthetase. Pairs of figures A and B, C and D are displayed in different scale and have equivalent orientation. Structural domains of ArgRS and subunits of PheRS are marked with different colors. The tRNA is shown as blue ribbons. All isopotential surfaces are calculated and built at ±0.01 kT/e with GRASP. Patches of positive (+) and negative (-) potentials are shown in blue and red respectively.*

protein recognition.

For the most part aaRSs are negatively charged at physiological conditions, as also are tRNA substrates. It is apparent that there are driving forces that ensure an attraction between like-charged aaRS and tRNA, and formation of the close encounters. It has been known that the bulk of contacts in the tRNA-aaRSs complexes occur between the protein and the sugar/phosphate backbone of tRNA; these contacts are non-specific, localized at widely spaced regions of tRNA and largely electrostatic. An increase in the ionic strength significantly weakens complex formation and is indicative of the importance of electrostatic interactions in aaRS-tRNA complexes.

Based on the non-linear Poisson-Boltzmann equation we evaluated the contribution of non-specific electrostatic interactions in the formation of aaRS-tRNA complexes at long distances. Our results explain how, at distances away from the molecular surfaces, monomeric, dimeric and heterotetrameric aaRSs

provide an association energy that holds macromolecules near each other, exerting control over tRNA motion towards the binding site. There are no exceptions for the observed empirical guideline: on the associated 3D-isopotential surfaces, each aaRS creates as many positive patches as there are tRNA molecules interacting with a given aaRS (Tworowski et al., 2005; Tworowski and Safro, 2003). The solutions of the Poison-Boltzmann equation and their visual representation (Figure 3) indicate that non-specific long-range electrostatic interactions are the dominant factor for general stickiness and formation of the encounter aaRSs·tRNA complexes.

## Genetic code and the aminoacyl-tRNA synthetases

The importance of aaRSs to the fidelity of the genetic code's translation implies that members of this family are probably among the earliest proteins to appear. Surprisingly, based on multiple sequences analysis and the architecture

of the catalytic sites, aaRSs were divided into two classes. Class II aaRSs are believed to have roots stretching back into antiquity further than class I enzymes, as they predominantly aminoacylate tRNA with simpler and small amino acids. Remarkably, class II related amino acids were detected in the classical Miller experiments which attempted to imitate primordial environments. Thus it would be reasonable to speculate that primordial organisms utilize predominately class II aaRSs. Searching for class II related "molecular fossils" would be next step.

Evidence suggests that methanogenic Archaea is among most ancient organisms. Remarkably methanogenic Archaea lack the canonical class I Cysteine-tRNA synthetase (CysRS), as well as other proteins, analogous to those involved in bacterial or eukaryotic cysteine biosynthesis. Instead of that methanogens use special tRNA-dependent cysteine biosynthesis pathways utilizing also class II like aaRSs called SepRS.

We examined the cysteine content among Archaea proteomes, and observed their clustering into two separate groups: first demonstrates relatively low cysteine content (~0.6 % of whole proteome) and the second contains twice as much cysteines as compare to the first one (~1.2%) (Klipcan et al., 2008). Proteomes with the low cysteine content are widespread among the different Archaea species, however all proteomes enriched in cysteine belong to Euryarchaeota, and all but one are from methanogenic organisms (Figure 4). The availability of SepRS in majority of these proteomes suggests a clear correlation between the tRNA-dependent pathway and an increase in cysteine content.

Our working hypothesis was that the enrichment in the cysteine among methanogenic Archaea is due to appearance there special cysteine clusters. We found that the occurrence of cysteine in the CX_2CX_2C-, CX_2CX_3C-, and CX_3CX_2C-type clusters in methanogenic Archaea was significantly higher than those of other Archaea. One might speculate that the higher cysteine content found
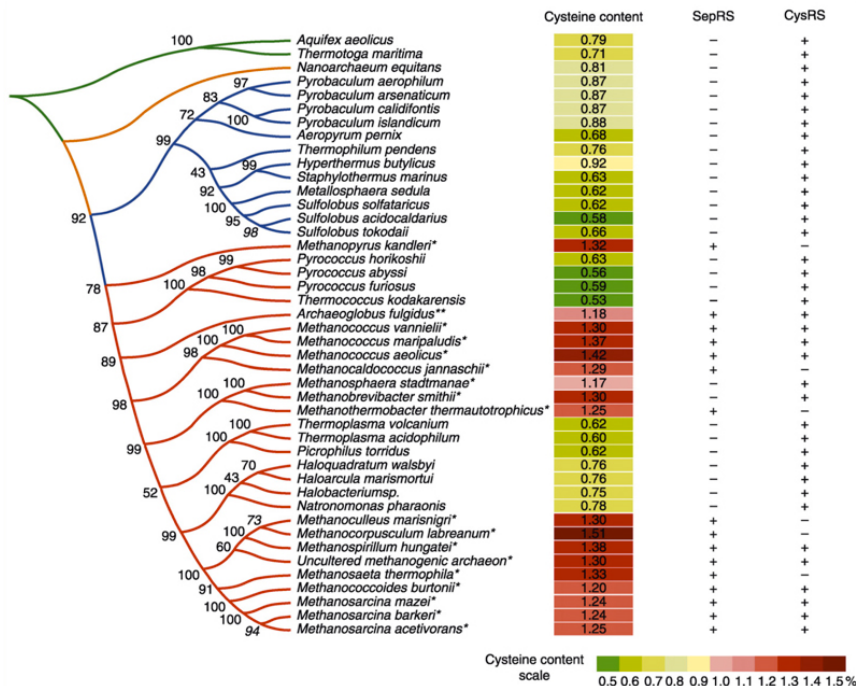
Fig. 4 *The phylogenetic tree of methanogenic Archaea and cysteine representation therein. The phylogenetic tree of methanogenic Archaea is based on a comparison of 16S rRNA sequences from the Euryarcheota (red) and Crenarcheota (blue), together with Nanoarchaeum equitans (orange) and two deep branched Bacteria (green). All cysteine-rich organisms belong to the methanogenic Archaea and utilize special tRNA-dependent pathway of cysteine biosynthesis. Methanogenic and SepRS-containing organisms are denoted by asterisks.*

in proteins might be caused by a methanogenic lifestyle. Nonetheless, analysis of Methanocaldococcus jannaschii proteins directly involved in methane biosynthesis does not indicate that they are particularly enriched in cysteine content. Thus, the overwhelming majority of $CX_{(n)}CX_{(m)}C$ clusters cannot be fully explained by the presence of methanogenic pathways. The unique cysteine clustering seen in methanogens suggests an additional, still undiscovered physiological adaptation underlying such clustering. Alternatively, the presence of these clusters might reflect remnants of the ancient metabolic cycles.

The close association between the newly discovered tRNA-dependent pathway of cysteine biosynthesis and the enrichment of methanogens with cysteine further supports the importance of cysteines in ancient metabolic processes. In light of the 'RNA world' hypothesis, such pathways probably represent remnants of an ancient translation apparatus and amino acid biosynthesis (Klipcan and Safro, 2004).

Selected publications

Fishman, R., Ankilova, V., Moor, N. and Safro, M. (2001) Structure at 2.6 A resolution of phenylalanyl-tRNA synthetase complexed with phenylalanyl-adenylate in the presence of manganese. Acta Crystallogr D Biol Crystallogr, 57, 1534-1544.

Klipcan, L., Frenkel-Morgenstern, M. and Safro, M.G. (2008) Presence of tRNA-dependent pathways correlates with high cysteine content in methanogenic Archaea. Trends Genet, 24, 59-63.

Klipcan, L. and Safro, M. (2004) Amino acid biogenesis, evolution of the genetic code and aminoacyl-tRNA synthetases. J Theor Biol, 228, 389-396.

Kotik-Kogan, O., Moor, N., Tworowski, D. and Safro, M. (2005) Structural basis for discrimination of L-phenylalanine from L-tyrosine by phenylalanyl-tRNA synthetase. Structure, 13, 1799-1807.

Levin, I., Kessler, N., Moor, N., Klipcan, L., Koc, E., Templeton, P., Spremulli, L. and Safro, M. (2007) Purification, crystallization and preliminary X-ray characterization of a human mitochondrial phenylalanyl-tRNA synthetase. Acta Crystallogr Sect F Struct Biol Cryst Commun, 63, 761-764.

Moor, N., Kotik-Kogan, O., Tworowski, D., Sukhanova, M. and Safro, M. (2006) The crystal structure of the ternary complex of phenylalanyl-tRNA synthetase with tRNA[Phe] and a phenylalanyl-adenylate analogue reveals a conformational switch of the CCA end. Biochemistry, 45, 10572-10583.

Mosyak, L., Reshetnikova, L., Goldgur, Y., Delarue, M. and Safro, M.G. (1995) Structure of phenylalanyl-tRNA synthetase from Thermus thermophilus. Nat Struct Biol, 2, 537-547.

Tworowski, D., Feldman, A.V. and Safro, M.G. (2005) Electrostatic potential of aminoacyl-tRNA synthetase navigates tRNA on its pathway to the binding site. J Mol Biol, 350, 866-882.

Tworowski, D. and Safro, M. (2003) The long-range electrostatic interactions control tRNA-aminoacyl-tRNA synthetase complex formation. Protein Sci, 12, 1247-1251.