Minireview

Structural Relationships and the Classification of Aminoacyl-tRNA Synthetases*

Jonathan J. Burbaum and Paul Schimmel

The aminocyl-tRNA synthetases are a family of 20 enzymes that catalyze the same two-step chemical reaction,

\[ \text{AA} + \text{ATP} \rightarrow \text{AA-AMP} + \text{PP}, \quad \text{tRNA}^{\text{AA}} \rightarrow \text{AA-tRNA}^{\text{AA}} + \text{AMP} \]

where AA represents one of the 20 naturally occurring amino acids and tRNA^{aa} represents a transfer RNA specific for that amino acid. In all, there are 20 such enzymes (one for each amino acid) that share the role of assigning amino acids to triplet codons in genetic translation. The synthetases are among the oldest proteins, and in contrast to more recently evolved and functionally related proteins, such as hemoglobin and myoglobin, their functional similarity is not reflected in a uniform structural framework. This relationship also contrasts with that of enzymes that share structural homology but are mechanistically distinct, e.g., mandelate racemase and muconate lactonizing enzyme (11). The diversity of synthetase structures has been a problem of long standing, whose solution may lead to a more basic understanding of protein structure/function and evolutionary relationships.

The diversity of synthetases is illustrated in Fig. 1, where the relative sizes and subunit compositions of the Class I and Class II synthetases from Escherichia coli are summarized. The enzymes range in quaternary structure from monomers (e.g., IleRS) to tetramers (AlaRS), in primary structure from 334 amino acids (TrpRS) to 1112 amino acids (PheRS), and in native molecular mass from 51,000 (CysRS) to 384,000 (AlaRS). The Class I enzymes are chiefly monomeric (only the dimeric Tyr and Trp enzymes cannot function as monomers (4, 5); the Met enzyme is converted into a monomer by limited proteolysis (6)), while the Class II enzymes are entirely oligomeric.

Based on a limited number of sequences, structural information (7, 8) and structural modeling, a group of related synthetases that initially included isoleucine, methionine, tyrosine, and glutamine were recognized (9, 10). This group was later expanded to include the Arg, Glu, Leu, Trp, and Val-tRNA synthetases (11-15). Each member of the group contains a “signature sequence” (9, 10), an 11-mer peptide that ends in a characteristic and mostly conserved tetrapeptide, HIGH (13). Additional sequence comparisons identified a second similarity, the pentapeptide KMSKS, in the same tetrapeptides in these three structures have high three-dimensional similarity, and together they form part of the ATP binding site (43). With the determination of the seryl-tRNA synthetase crystal structure and a complete set of sequences for the E. coli synthetases, the diversity of the synthetases was extended to their three-dimensional frameworks. The structure of SerRS (37) is unlike that of the Met, Tyr, or Gln synthetases. The dominant structural motif of SerRS is an eight-stranded antiparallel β-sheet that somewhat resembles the NAD-binding motif of the enterotoxins (44), which contain a seven-stranded sheet, but bears no relationship to the Rossmann folds of the earlier structures. Moreover, in those synthetases that lack the sequence motifs mentioned above, three unique regions of degenerate sequence similarity (motifs 1-3; Ref. 2) were identified. The lengths and the sequences of the central parts of these motifs are as follows: motif 1 (18 amino acids), (FRNE)_{2}; motif 2 (25-31 amino acids), (GRRK)(ERF); motif 3 (29-34 amino acids), (GQQGQ)(GERFF). As with the signature sequence 11-mer, a mostly conserved tetrapeptide can be used to define two of the three motifs; for motif 2, this tetrapeptide is FRNE, and for motif 3, it is GLER. These motifs (and the signature sequence similarities mentioned above) are the basis for dividing the 20 synthetases into two groups of 10 each (Fig. 1).

Further support for this grouping has come from a partial sequence of the S. cerevisiae aspartyl-tRNA synthetase: tRNA^{Asp} complex (41). This Class II enzyme contains a segment that is similar to the antiparallel β domain of SerRS. Based on the two structures, the active site of the Class II enzymes has been tentatively identified as being formed in part from amino acids that make up motifs 2 and/or 3. Motif 1 in the Class II alignment forms a portion of a subunit interface in both the SerRS and AspRS structures. This motif may be present because the Class II synthetases are predominantly dimers (Fig. 1). As an exception is the tetrameric E. coli alanine enzyme, which can function as a monomer (45-47) and is naturally monomeric in higher organisms (for example, Bombyx mori; Ref. 48).

Subgroups within Class I Enzymes—Within the Class I synthetases, additional diversity is apparent. Each of the three crystal structures has a unique variant of the nucleotide-binding fold framework. In Fig. 2, the different nucleotide-binding fold topologies of the three known structures are aligned according to the position of the signature sequence and KMSKS motifs. In the enzymes for which no structures are yet known, we assumed that the three-dimensional dispositions of these two motifs are fixed

---

*This work was supported by Grants GM15539 and GM23062 from the National Institutes of Health.

1 The classes are defined on the basis of sequence comparisons (2), rather than on their occurrence in monooenzyme complexes (5).

2 The abbreviations used are: RS, tRNA synthetase(s); D, dihydrouridine.

3 For quaternary structure references, see Schimmel (11) and Miranda (3).

4 Sequences: AlaRS (16); ArgRS (20); CysRS (22, 23); GlnRS (24); GluRS (27); HisRS (26); IleRS (9); LeuRS (27); LysRS (2); MetRS (30); PheRS (31); ProRS (2); SerRS (32); ThrRS (33); TrpRS (34); TyrRS (35); ValRS (36).

5 The codes for similar amino acids are as follows: Φ, hydrophobic amino acid; Ψ, positively charged amino acid; Ω, negatively charged amino acid. Gaps in similarity of fixed length are denoted by strings of the letter X, and variable length gaps are denoted by ( ).

16965
relative to one another and to other secondary structure elements of the fold. Based on this assumption, the seven remaining Class I structures can be tentatively divided into three groups (Fig. 2).

In all three families, the first half of the nucleotide-binding fold contains the signature sequence, and in all but the Arg enzyme, this sequence occurs within the first 70 amino acids. The signature sequence occurs between an inner helix and β-strand within the N-terminal half of the nucleotide-binding fold. The catalytic residues that form the active site include those in a loop after strand E, which contains the KMSKS region. For TyrRS, whose signature sequence is found at residues 40–50, the N-terminal sequence forms the outermost antiparallel strand of the second half of the fold. Because the signature sequence of the Arg enzyme is also displaced (by 122 amino acids) from the N terminus, it is tentatively classified with Tyr rather than with Gln and Glu. Furthermore, the Trp enzyme is not classified with Tyr because its signature sequence is found only 10 residues from the N terminus, not far enough to allow for the extra strand.

Additional differences between the families are found in the connectivity within the nucleotide-binding fold. The largest family corresponds to the methionyl-tRNA synthetase structure and contains synthetases which activate the hydrophobic amino acids Ile, Leu, Met, and Val, as well as Cys (which can also be classified as hydrophobic; Ref. 49). This family has both the largest and the smallest Class I enzymes, with the differences in size concentrated in the insertions that are designated connective polypeptide domains CP1 and CP2 (10, 22). The connectivity differences between families are mainly in the second half of the fold, where the Met family contains a connective polypeptide (CP2) between the first and second strands (D and E), the Gln family has CP2 between the second and third strands (E and F), and the Tyr family has no identifiable CP2. The alignment and subgroup classification is assisted by a short stretch of similarity that defines the first strand of the second half of the fold (strand D; Ref. 27), which allows the end of CP1 and the start of CP2 to be located.

These specific differences between Class I enzymes should not obscure their close relationship. A recent comparison of the Gln and Met enzyme structures (50) shows that the two enzymes, although in different families, share a rare "left-handed crossover" between strands E and F (where the corresponding strands of TyrRS are not directly connected). This topology places the signature sequence and KMSKS motifs on the same side of the β-sheet. In the GlnRS:RNA^Glu^coscrystal, strand F fits into the "armpit" region of the L-shaped tRNA molecule, between the minihelix domain and the D-anticodon domain (Fig. 3). The left-handed crossover, together with the unusually short strand "D," provides part of the docking interaction for the acceptor stem of tRNA.

**Motifs for Interaction of Class I Enzymes with tRNA**

In the Class I Gln- and Met-tRNA synthetases, two separate domains are used for interactions with the two domains of the L-shaped tRNA structure (51–53; Fig. 3). The acceptor-TYC minihelix domain of tRNA^Glu^interacts with the insertions into the nucleotide-binding fold framework of GlnRS (39). This nucleic acid domain (M_r ~ 12,000) must be placed in position to accept an activated amino acid from a labile adenylate intermediate (M_r ~ 450) at the catalytic site of the enzyme. The CP1 insertion occurs at a natural break in the nucleotide-binding fold topology of the catalytic domain, as a connector which joins the two β102 halves of the fold (54). In GlnRS, the acceptor helix-binding CP1 structure consists of a five-stranded antiparallel β-sheet flanked...
AspRS cocrystal with tRNA (41) demonstrates that contacts domain, followed by a small domain (containing the motif anti-parallel sequence) that is involved with oligomerization. This in turn is noted above, the topology of the Class I1 SerRS resembles AspRS, occurring after the first β-strand of the antiparallel β-helix.

A separate C-terminal domain interacts with the distal domain of the tRNA, which is comprised of the dihydrouridine (D) stem-loop and the anticodon stem-loop (Fig. 3). The anticodon stem-loop comprises of a continuous helix spanning the acceptor stem and TΨC stem-loop. The D-anticodon stem-loop domain is comprised of the D and anticodon stem-loops.

Assembly of Class I and Class II Enzymes

The assembly of Class I and Class II structures from individual domains is depicted in Fig. 4. The modular arrangement of functional domains common to all classes of synthetase was first suggested by deletion analysis of the Class I B. stearothermophilus tyrosyl-tRNA synthetase (57) and of the Class II E. coli alanyl-tRNA synthetase (45-47), where domains could individually be isolated and investigated. From genetic and biochemical studies, the Class II alanyl-tRNA synthetase appears to have a domain organization that resembles the Class I enzymes (11). It has an N-terminal amino acid activation domain, with insertions of sequences that contact the acceptor stem of bound tRNA (58, 59). This domain is followed by another motif that contacts tRNA (60), possibly base in the D stem-loop and anticodon stem (61), followed by an oligomerization domain (44-46).

Unlike the Class I enzymes and alanyl-tRNA synthetase, the AspRS cocrystal with tRNA (41) demonstrates that contacts outside of the acceptor stem are mediated by an N-terminal domain, followed by a small domain (containing the motif anti-parallel sequence) that is involved with oligomerization. This in turn is followed by an antiparallel β-domain that contains an insertion (or insertions) to contact the acceptor stem of the tRNA. As noted above, the topology of the Class II SerRS resembles AspRS, although the oligomerization domain is formally an insertion occurring after the first β-strand of the antiparallel β-domain.
significant sequence similarity to the ammonia-dependent aspartate synthase from E. coli (83). One region of similarity includes motif 3. Because both enzymes proceed via an adenylate intermediate, this similarity supports the assignment of motif 3 to the active site.

The establishment of two broad classes of synthetases may have occurred early. There is no example of a "class switch" of a synthetase in evolution. The subunit (B. stearothermophilus, Ref. 84; Bacillus caldolyticus, Ref. 85; E. coli, Ref. 35) and eukaryotic (S. cerevisiae; Ref. 86) nucleotide-acceptor synthetase sequences established them to be evolutionarily related Class I enzymes. Similarly, the recently sequenced methionyl-tRNA synthetase from the thermophilic bacterium Thermus thermophilus (89) shows that it is a Class I enzyme, as are the methionine-tRNA synthetases from a mesophilic bacterium (E. coli; Ref. 36) and an eluates (S. cerevisiae; Ref. 88). The isoleucyl-tRNA synthetase from an archaebacterium, Methanobacterium thermoautotrophicum (89) shows to it be a Class I synthetase, as are its subunit (E. coli, Ref. 9) and eukaryotic (S. cerevisiae; Ref. 90) counterparts. Additional examples of conservation of the classification are evident from comparisons with four human sequences (91-94). Before a generalization is possible, however, more examples are needed.

Acknowledgments—We thank Prof. Dino Maras (Strasbourg) for providing us with results prior to publication and Drs. D. Morris and S. Cusk for helpful comments. We also thank: R. Biemann, K., Eisenheis, S. J., Eriani, S. E., Gangloff, J. S., Goff, S. P., Lim, W. A., Mayaux, J. F., Fayat, G., Fromant, J., Springer, M. Grunherg-Manago, W. A., Rossmanni, M., Tsui, F. W. L., and J. Alan. Personal communication.

REFERENCES

1. Nejhat, D. J., Kenyon, G. L., Gerlt, J. A. & Petsko, G. A. (1990) Nature 347, 692-694
2. Enari, G., Delarie, P., Och, G., and G. G. (1990) Nature 347, 692-694
3. Mirande, M. (1991) Prog. Nucleic Acid Res. Mol. Biol. 40, 90-142
4. Jones, D. H., McMillan, A. J., Freisheit, A. R. & Winter, G. (1985) Biochemistry 24, 5029-5036
5. Ibarra, F., Dorizzi, M. and Laboissiere, J. (1973) Eur. J. Biochem. 29, 275-282
6. Cossio, D. & Waller, J. P. (1971) Eur. J. Biochem. 20, 283-300
7. Rulier, J.-L., Zewer, C. & Brunei, S. (1981) Nature 292, 384-386
8. Stotz, T. N., Blow, D. M., Brick, F., Nyborg, P. & Nyborg, J. (1992) J. Mol. Biol. 198, 159-209
9. Webster, T. S., Tsai, H., Kuma, M., Mackie, G. A. & Schimmel, P. (1994) Science 266, 1310-1317
10. Starzynski, R. M., Webster, T. A. & Schimmel, P. (1987) Science 237, 1614-1617
11. Schimmel, P. (1987) Annu. Rev. Biochem. 56, 125-158
12. Enari, G., Dirheimer, G. & Gondolf, J. (1989) Nucleic Acids Res. 17, 2755-2761
13. Burburna, J. J., Starzynski, R. M. & Schimmel, P. (1990) Proteins: Struktur. Funktion. Genet. 7, 93-111
14. Hardesty, B., Klug, A. G. W. & White, R. (1988) J. Biol. Chem. 263, 868-877
15. Breeton, R., Sasafoha, H., Papaya, et Nousou, I., Biehn, K. & Lapointe, J. (1986) J. Biol. Chem. 261, 1001-1007
16. Hoang, D. T., Do, P. M. & White, R. (1986) Biochimie 68, 1071-1078
17. Griepken, M., Makashov, A. D. & Eisen, B. (1987) Proc. Nat Acad. Sci. U. S. A. 84, 1055-1059
18. Wehber, T. A., Latr, R. R. & Schimmel, P. (1987) Biochemistry 26, 8990-8997
19. Putnam, N. D., Royal, N. D., De Veyra, H. N., Hikita, W., Biemnnk, I. & Schimmel, P. (1981) Science 215, 1497-1500
20. Asselme, J. & Hartleien, M. (1988) Gene (Amst.) 84, 451-485
21. Enari, G., Dirheimer, G. & Gondolf, J. (1990) Nucleic Acids Res. 18, 1109-1116
22. Hoa, V. M., Shiba, K., Motten, C. & Schimmel, P. (1991) Proc. Nat Acad. Sci. U. S. A. 88, 976-980
23. Enari, G., Dirheimer, G. & Gondolf, J. (1991) Nucleic Acids Res. 19, 365-370
24. Hoben, P., Royal, N., Cheung, A., Yamato, F., Biemann, I. & Schimmel, P. (1982) J. Biol. Chem. 257, 1644-1650
25. Wood, T., Riordon, J. B., Kung, T., Biemann, I. & Schimmel, P. (1983) J. Biol. Chem. 258, 10637-10641
26. Freseman, K., Gibson, B., Donovan, D., Biemann, K., Eisenhieis, H. S., Hartleion, M. & Parker, R. A. (1985) J. Biol. Chem. 260, 10053-10068
27. Hartlein, M. & Morden, D. (1987) Nucleic Acids Res. 15, 10190-10210
28. Kawai, K., Johnson, H. V., Bishoyi, G. R., Iieda, H. & Nakamura, Y. (1988) Proc. Nat Acad. Sci. U. S. A. 85, 5650-5654
29. Levets, F., Plateau, D., Pessen, Y. & Blanquet, G. (1990) Nucleic Acids Res. 19, 1565-1572
30. Dardel, F., Faye, G. & Blanquet, S. (1984) J. Biol. Chem. 263, 1115-1122
31. Macchioni, M., Faye, G. & Blanquet, S. (1984) J. Biol. Chem. 263, 787-791
32. Mocquard, M., Lepoivre, L. & Aymard, J. (1988) Nucleic Acids Res. 16, 1005-1017
33. Rossmanni, M., Tsui, F. W. L., and J. Alan. Personal communication.

*J. Hill and A. Tsagoloff, personal communication.