Protein Science (1993), *2*, 2259–2262. Cambridge University Press. Printed in the USA.

FOR THE RECORD

Idiographic representation of conserved domain of a class II tRNA synthetase of unknown structure

LLUIS RIBAS DE POUPLANA, DOUGLAS D. BUECHTER, MATTHEW W. DAVIS, AND PAUL SCHIMMEL

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(RECEIVED August 9, 1993; ACCEPTED September 2, 1993)

We report here an idiographic representation of the secondary structure predicted for an enzyme whose three-dimensional structure has not yet been determined. In our example, we applied the recently described neural network-based prediction program PHD (Rost & Sander, 1992) to identify secondary structure elements of a major portion of the active site of *Escherichia coli* alanyl-tRNA synthetase (AlaRS). The active site domain of AlaRS encompasses approximately the N-terminal half of the primary structure (Putney et al., 1981b; Jasin et al., 1983; Regan et al., 1987; Buechter & Schimmel, 1993) and is believed to be the primordial enzyme, which, together with other aminoacyl-tRNA synthetases, arose as one of the earliest proteins. These proteins established the rules of the genetic code through the aminoacylation of RNA substrates. The alanine enzyme is of particular interest because it is one of the clearest examples of a tRNA synthetase that makes no contact with the tRNA anticodon (Hou & Schimmel, 1988; Park et al., 1989), showing that the genetic code relationship between alanine and its four trinucleotide codons is indirect. Instead, the enzyme aminoacylates RNA helical structures, including short hairpin duplexes with as few as four base pairs, provided that they contain a G:U wobble base pair at the third position in the helix (Francklyn & Schimmel, 1989; Musier-Forsyth et al., 1991; Shi et al., 1992).

The aminoacyl-tRNA synthetases are divided into two classes of 10 enzymes each (Cusack et al., 1990; Eriani et al., 1990). This classification is based on conserved sequence elements and structures that are shared by the active sites of members of the same class but are completely different from those of the other class. The overall secondary structure architecture of the core of the active-site domain is similar for members of the same class whose three-dimensional structures are known (Brick et al., 1988; Rould et al., 1989; Brunie et al., 1990; Cusack et al., 1990; Ruff et al., 1991). This structural similarity is in spite of the apparent lack of relatedness of sequences outside of the conserved elements.

Escherichia coli AlaRS is an α_4 tetramer of 875 amino acid subunits. A monomeric 461-amino acid N-terminal fragment catalyzes as efficiently as native enzyme the aminoacylation of small RNA oligonucleotides that recreate only the acceptor stem of $tRNA^{Ala}$ (Buechter & Schimmel, 1993). As a class II enzyme, AlaRS is related to the seryl- (Cusack et al., 1990) and aspartyl- (Ruff et al., 1991) tRNA synthetases whose structures have been determined. Apart from the class II sequence motifs, AlaRS has no primary sequence similarity with either AspRS, SerRS, or the other class II enzymes. The class II sequence elements, which collectively total ~23 amino acids, constitute only a minor fraction (~10%) of the domain in AspRS and SerRS that encompasses them.

We aligned the four published (*E. coli* [Putney et al., 1981a], *Bombyx mori* [Chang & Dignam, 1990], *Rhizobium leguminosarum* [Selbitschka et al., 1991], and *Rhizobium meliloti* [Selbitschka et al., 1991]) and two unpublished (*Arabidopsis thaliana* [I. Small, pers. comm.] and human [K. Shiba, pers. comm.]) sequences of the N-terminal active-site domains of AlaRS with the three sequence fragments in the C-terminal domains of yeast AspRS and *E. coli* SerRS that contain the conserved class II motifs. (The active-site domains can be in either the N- or C-terminal part of the enzyme, depending on the synthetase.) The identification of motif 3 of AlaRS was the same as previously described (Eriani et al., 1990; Cusack et al., 1991). The conserved phenylalanine of motif 2 (Phe 90 in AlaRS) aligns with the same phenylalanine in AspRS and SerRS as described previously (Cusack et al., 1991); however, the conserved arginine of motif 2, which in both SerRS and AspRS contacts bound ATP (Cavarelli et al.,

Reprint requests to: Paul Schimmel, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

1993; Cusack et al., 1993), aligns with the strictly conserved Arg 69 in AlaRS rather than the previously proposed His 74, which is not conserved in all six sequences. Motif 1 of AspRS and SerRS consistently aligned with a region, ^{34}PTLL, on the N-terminal side of motif 2 of the six alanyl-tRNA synthetases. This region contains an invariant proline (Pro 34), as is found in motif 1 of seven of the nine other class II enzymes.

We used the secondary structure prediction program PHD (Rost & Sander, 1992) to identify the major secondary structural elements of the AlaRS active site. We focused on identifying the seven β-strands and three α-helices that are believed common to the conserved domain of the class II enzymes. Additional elements of secondary structure were also identified, but they are not considered further here because these elements appear to be idiosyncratic to AlaRS and to be not part of the core active-site structure that is known to be shared by the class II synthetases.

PHD uses alignments of homologous sequences and has achieved an expected accuracy of 82% (compared to actual structures) when multiple sequences are available. (Percent accuracy is defined as the number of correctly predicted residues divided by the number of all residues × 100 [B. Rost & C. Sander, PHD program documentation].) Collectively there are 20 (2 × 10) discrete secondary structure elements that make up the conserved domains of AspRS and SerRS. In a test of the applicability of PHD to the aminoacyl-tRNA synthetases, 18 of the 20 predicted secondary structure elements for the active-site domains of AspRS and SerRS overlapped by two or more residues with those from the X-ray structure. When the known secondary structure of the conserved sequence motifs is also considered, all 20 elements can be identified.

The PHD program located two helices (H1 and H2, Fig. 1) in AlaRS at positions with respect to motifs 1 and

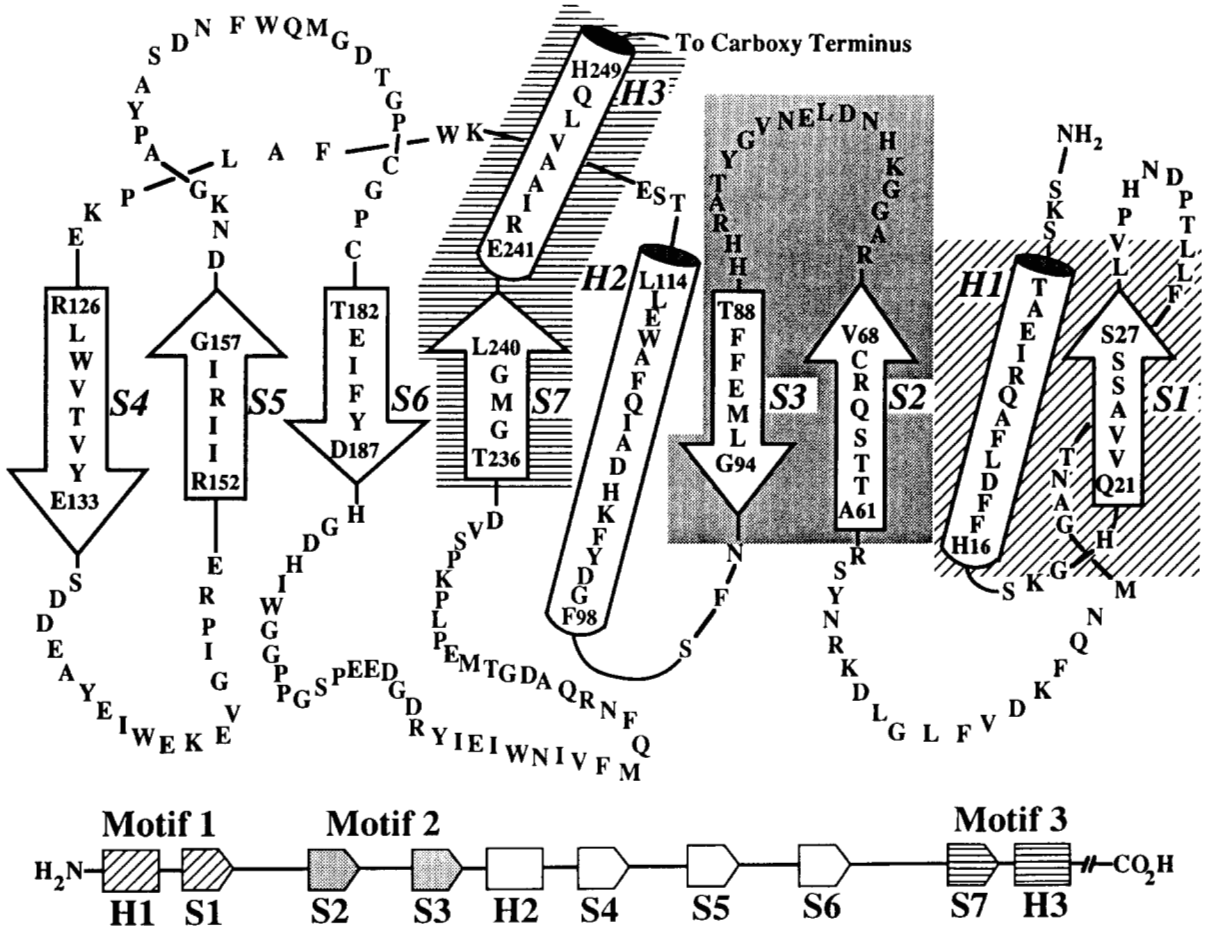

Fig. 1. Idiographic representation of the seven β-strands and three α-helices that comprise the active-site domain of *E. coli* alanyl-tRNA synthetase. A linear depiction of the active-site secondary structure is shown below the idiograph. Rectangles and cylinders denote α-helices, and arrows denote β-strands. The secondary structure elements that encompass the conserved class II-defining sequence motifs are shaded and labeled. Note that the sequences ^{30}PITD and ^{157}NDNKGAPYASGNFWRMGGTGPCD, reported by Putney et al. (1981a), have been corrected to ^{30}PHND and ^{157}GDNKGAPYASDNFWQMGDTGPCG, respectively (the corrected residues are underlined).

2 similar to those of the active site-flanking helices in AspRS and SerRS. We used the positions of these helices, along with the locations of the conserved sequence motifs from the multiple sequence alignment, as a starting point to assign the remaining structural elements of the active-site fold. The model (Fig. 1) incorporated only those secondary structure elements that had at least one residue predicted by PHD at the 82% expected accuracy level, with the exception of strand S7 of motif 3, which is not predicted by PHD for either AlaRS or AspRS. Instead, we assigned this strand on the basis of the well-conserved characteristic motif 3 sequence element (GLER) and the expected structural homology to the region of SerRS and AspRS that encompasses it. A more detailed description of each region of the active-site domain is given below.

Motif 1 in the class II synthetases comprises an α-helix, a β-strand, and a variable region that follows the conserved motif 1 proline. Previously no motif 1 for the class II tetrameric alanyl-, glycyl-, and phenylalanyl-tRNA synthetases could be located (Eriani et al., 1990; Cusack et al., 1991). Our identification of motif 1 in AlaRS suggests that the possibility of the presence of this motif in GlyRS and PheRS should be reinvestigated.

Residues in the motif 1 α-helix make contact with the corresponding helix of the other subunit in the dimeric AspRS (Ruff et al., 1991) and SerRS (Cusack et al., 1990). Additional interface contacts are made in the region following the β-strand. In our model (Fig. 1), motif 1 includes helix H1 (Thr 4 → His 16) and strand S1 (Gln 21 → Ser 27). Earlier work on *E. coli* AlaRS established that deletion of the C-terminal 176 amino acids converted the enzyme into an active monomer (Jasin et al., 1983; Regan et al., 1987). Because this deleted region does not include motif 1, it is clear that motif 1 and immediately flanking sequences are not sufficient for oligomerization, at least in AlaRS. The role of motif 1 in oligomerization of other class II enzymes needs to be investigated. In AspRS, residues in the region following the motif 1 β-strand contact the acceptor stem of bound tRNA (Cavarelli et al., 1993). Similar contacts may be made in AlaRS.

In our model, motif 2 comprises strands S2 (Ala 61 → Val 68) and S3 (Thr 88 → Gly 94) connected by a loop of 19 amino acids. The position of strand S2, although different from that previously proposed (Cusack et al., 1991), is supported by the alignment of six sequences of AlaRS with those of SerRS and AspRS and by the PHD-derived secondary structure prediction. Motif 2 in the class II synthetases contains an invariant arginine and a conserved acidic residue two amino acids beyond the arginine. Both of these amino acids make contact with bound ATP in the SerRS–ATP complex (Cusack et al., 1993). The conserved arginine in SerRS and AspRS aligns with Arg 69, which is conserved in all six sequences of AlaRS. Arg 69 is at the beginning of the loop between the two strands, similar to where it is located in the other class II enzymes. The strictly conserved Asp 76 (in the six AlaRS sequences) is the first acidic residue to follow Arg 69.

The 19-amino acid loop between the two strands of motif 2 in AlaRS is the longest of all of the class II synthetases. The 3′-end of periodate oxidized-tRNA^{Ala} cross-links to Lys 73 within the predicted loop of motif 2 of *E. coli* AlaRS (Hill & Schimmel, 1989). The corresponding loop of eight amino acids in AspRS makes sequence-specific contacts with the discriminator base and the first base pair of bound tRNA^{Asp} (Ruff et al., 1991; Cavarelli et al., 1993). The motif 2 loop in AlaRS is also likely to make acceptor stem contacts with bound tRNA^{Ala}.

The region between motifs 2 and 3 in the class II synthetases is of variable size and sequence (Eriani et al., 1990; Cusack et al., 1991). However, in yeast AspRS (Ruff et al., 1991) and *E. coli* SerRS (Cusack et al., 1990) part of this region is topologically conserved and contributes three β-strands and an α-helix to the active-site fold. Although these four elements of secondary structure in AlaRS were each predicted by PHD at the 82% expected accuracy level, their identification cannot be supported further by the presence of one of the three class-defining sequence motifs. Acknowledging this limitation, but recognizing that these four elements of secondary structure are most likely present between motifs 2 and 3, the putative helix H2 and strands S4, S5, and S6 strongly predicted by PHD were included in the model. Helix H2 corresponds to the "cross-over" helix that follows the second strand of motif 2 in both AspRS and SerRS.

The conserved structural elements in the region between motifs 2 and 3 are connected by loops/insertions that are of variable size in the class II synthetases (Eriani et al., 1990; Cusack et al., 1991). Serine, threonine, and lysine residues in the loop following the outermost strand of the active-site fold (S4 in Fig. 1) make contact with the phosphate backbone of the acceptor stem of tRNA^{Asp} in the AspRS–tRNA^{Asp} complex (Ruff et al., 1991; Cavarelli et al., 1993). Residues in these variable regions may contribute idiosyncratically to tRNA recognition by other class II enzymes.

Motif 3 contributes the last β-strand to the active-site fold. A topologically conserved α-helix follows this strand in AspRS and SerRS. The conserved arginine in the motif 3 GLER sequence contacts the ribose of ATP in the AspRS–ATP complex (Cavarelli et al., 1993). Motif 3 in AlaRS comprises strand S7 (Thr 236 → Leu 240) and helix H3 (Glu 241 → His 249). The invariant arginine (Arg 242) is at the beginning of helix H3, similar to where it is found in both AspRS and SerRS. It is likely that this motif in AlaRS makes contacts to ATP similar to those seen in AspRS.

The acceptor stem G3:U70 base pair is unique to alanine tRNAs and is conserved in prokaryote and cytoplasmic eukaryote alanine tRNAs (Sprinzl et al., 1989; Schimmel, 1991). Specific recognition of the G:U pair by

AlaRS provides a major portion of the basis for discriminating $tRNA^{Ala}$ from other tRNAs (Francklyn et al., 1992). Thus, amino acid residues within the active site of AlaRS that make G3:U70 contacts are likely to be conserved. We expect that these contacts in AlaRS are made by residues in the region following motif 1, in the loop between the two strands of motif 2, or in regions between motifs 2 and 3 that are variable in the class II enzymes. Candidates in these regions can be identified based on the potential for side-chain hydrogen bonding or electrostatic interactions (Seeman et al., 1976; Steitz, 1990) and on whether they are conserved in the known AlaRS sequences. Using these criteria, possibilities include Asn 40, Gln 45, and Lys 47 in the region following strand S1; Lys 73 (whose proximity to the 3'-end of $tRNA^{Ala}$ has been established [Hill & Schimmel, 1989]), Asn 75, Asp 76, Glu 78, Asn 79, Thr 83, and His 86 in the motif 2 loop; and, analogous to the AspRS system, Asp 135 in the region between strands S4 and S5. We are investigating these possibilities.

Acknowledgments

This work was supported by a grant from the Human Frontier Science Program and grants GM15539 and 23562 from the National Institutes of Health. D.D.B. is a Fellow of the American Cancer Society. We thank Drs. Dino Moras and Stephen Cusack for useful discussions and suggestions during the past year.

References

Brick, P., Bhat, T.N., & Blow, D.M. (1988). Structure of tyrosyl-tRNA synthetase refined at 2.3 Å resolution. Interaction of the enzyme with the tyrosyl adenylate intermediate. *J. Mol. Biol. 208*, 83–98.

Brunie, S., Zelwer, C., & Risler, J.-L. (1990). Crystallographic study at 2.5 Å resolution of the interaction of methionyl-tRNA synthetase from *Escherichia coli* with ATP. *J. Mol. Biol. 216*, 411–424.

Buechter, D.D. & Schimmel, P. (1993). Dissection of a class II tRNA synthetase: Determinants for minihelix recognition are tightly associated with domain for amino acid activation. *Biochemistry 32*, 5267–5272.

Cavarelli, J., Rees, B., Ruff, M., Thierry, J.-C., & Moras, D. (1993). Yeast $tRNA^{Asp}$ recognition by its cognate class II aminoacyl-tRNA synthetase. *Nature 362*, 181–184.

Chang, P.K. & Dignam, J.D. (1990). Primary structure of alanyl-tRNA synthetase and the regulation of its mRNA levels in *Bombyx mori*. *J. Biol. Chem. 265*, 20898–20906.

Cusack, S., Berthet-Colominas, C., Biou, V., Borel, F., Fujinaga, M., Hartlein, M., Kriklivіy, I., Nassar, N., Price, S., Tukalo, M.A., Yaremchuk, A.D., & Leberman, R. (1993). The crystal structure of seryl-tRNA synthetase and its complexes with ATP and $tRNA^{Ser}$ In *The Translation Apparatus*, in press. Springer-Verlag, Berlin.

Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N., & Leberman, R. (1990). A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.4 Å. *Nature 347*, 249–255.

Cusack, S., Härtlein, M., & Leberman, R. (1991). Sequence, structure and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. *Nucleic Acids Res. 19*, 3489–3498.

Eriani, G., Delarue, M., Poch, O., Gangloff, J., & Moras, D. (1990). Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature 347*, 203–206.

Francklyn, C. & Schimmel, P. (1989). Aminoacylation of RNA minihelices with alanine. *Nature 337*, 478–481.

Francklyn, C., Shi, J.-P., & Schimmel, P. (1992). Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices. *Science 255*, 1121–1125.

Hill, K. & Schimmel, P. (1989). Evidence that the 3'-end of a tRNA binds to a site in the adenylate synthesis domain of an aminoacyl-tRNA synthetase. *Biochemistry 28*, 2577–2586.

Hou, Y.-M. & Schimmel, P. (1988). A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature 333*, 140–145.

Jasin, M., Regan, L., & Schimmel, P. (1983). Modular arrangement of functional domains along the sequence of an aminoacyl tRNA synthetase. *Nature 306*, 441–447.

Musier-Forsyth, K., Scaringe, S., Usman, N., & Schimmel, P. (1991). Enzymatic aminoacylation of single-stranded RNA with an RNA cofactor. *Proc. Natl. Acad. Sci. USA 88*, 209–213.

Park, S.J., Hou, Y.-M., & Schimmel, P. (1989). A single base pair affects binding and catalytic parameters in the molecular recognition of a transfer RNA. *Biochemistry 28*, 2740–2746.

Putney, S.D., Royal, N.J., Neuman de Vegvar, H., Herlihy, W.C., Biemann, K., & Schimmel, P. (1981a). Primary structure of a large aminoacyl-tRNA synthetase. *Science 213*, 1497–1500.

Putney, S.D., Sauer, R.T., & Schimmel, P.R. (1981b). Purification and properties of alanine tRNA synthetase from *Escherichia coli*. *J. Biol. Chem. 256*, 198–204.

Regan, L., Bowie, J., & Schimmel, P. (1987). Polypeptide sequences essential for RNA recognition by an enzyme. *Science 235*, 1651–1653.

Rost, B. & Sander, C. (1992). Jury returns on structure prediction. *Nature 360*, 540.

Rould, M.A., Perona, J.J., Söll, D., & Steitz, T.A. (1989). Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. *Science 246*, 1135–1142.

Ruff, M., Krishnaswamy, S., Boeglin, M., Poterszman, A., Mitschler, A., Podjarny, A., Rees, B., Thierry, J.C., & Moras, D. (1991). Class II aminoacyl transfer RNA synthetases: Crystal structure of yeast aspartyl-tRNA synthetase complexed with $tRNA^{Asp}$. *Science 252*, 1682–1689.

Schimmel, P. (1991). RNA minihelices and the decoding of genetic information. *FASEB J. 5*, 2180–2187.

Seeman, N.C., Rosenberg, J.M., & Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA 73*, 804–808.

Selbitschka, W., Arnold, W., Priefer, U.B., Rottschafer, T., Schmidt, M., Simon, R., & Puhler, A. (1991). Characterization of recA genes and recA mutants of *Rhizobium meliloti* and *Rhizobium leguminosarum* biovar *viciae*. *Mol. Gen. Genet. 229*, 86–95.

Shi, J.-P., Martinis, S.A., & Schimmel, P. (1992). RNA tetraloops as minimalist substrates for aminoacylation. *Biochemistry 31*, 4931–4936.

Sprinzl, M., Hartmann, T., Weber, J., Blank, J., & Zeidles, R. (1989). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res. 17*, r1–r172.

Steitz, T.A. (1990). Structural studies of protein–nucleic acid interaction: The sources of sequence-specific binding. *Q. Rev. Biophys. 23*, 205–280.