Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 4561–4565, June 1990
Biochemistry

Sequence and structural similarities between the leucine-specific binding protein and leucyl-tRNA synthetase of *Escherichia coli*

(evolution/protein structure)

RUFUS M. WILLIAMSON AND DALE L. OXENDER*

Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI 48109-0606

Communicated by M. J. Coon, March 12, 1990 (received for review December 21, 1989)

ABSTRACT **A role for the leucyl-tRNA synthetase (EC 6.1.1.4) has been established for regulating the transport of leucine across the inner membrane of *Escherichia coli* by the leucine, isoleucine, valine (LIV-I) transport system. This transport system is mediated by interactions of periplasmic binding proteins with a complex of membrane-associated proteins, and transcription of the high-affinity branched-chain amino acid transport system genes is repressed by growth of *E. coli* on high levels of leucine. We now report results from sequence comparisons and structural modeling studies, which indicate that the leucine-specific binding protein, one of the periplasmic components of the LIV-I transport system, contains a 121-residue stretch, representing 36% of the mature protein, which displays both sequence and structural similarities to a region within the putative nucleotide-binding domain of leucyl-tRNA synthetase. Early fusion events between ancestral genes for the leucine-specific binding protein and leucyl-tRNA synthetase could account for the similarity and suggest that processes of aminoacylation and transport for leucine in *E. coli* may be performed by evolutionarily interrelated proteins.**

In our laboratory, we have been studying the regulation of high-affinity transport of leucine in *Escherichia coli*. This transport system is mediated by interaction of two periplasmic branched-chain amino acid-binding proteins, one of which is specific for leucine, with a common set of membrane-bound proteins (1). Genes for the periplasmic and membrane proteins lie in a cluster at 76 minutes on the *E. coli* chromosome map (2). Translation products of *livH*, *livM*, *livG*, and *livF* form a complex of membrane-associated proteins. The leucine-specific and general branched-chain amino acid-binding proteins are the products of *livK* and *livJ*, respectively (3–6).

Relationships between transport of leucine and aminoacylation by leucyl-tRNA synthetase (EC 6.1.1.4) have been previously established from cells of organisms as diverse as bacteria and mammals. Regulation of leucine transport in *E. coli* involves interaction of leucine with its aminoacyl-tRNA synthetase and cognate leucyl-tRNA species. Strains harboring a gene for a temperature-sensitive leucyl-tRNA synthetase display a 5-fold increase in leucine transport when the enzyme is inactivated by growth at a restrictive temperature (7). Expression of leucine transport proteins in *E. coli* may also be involved in regulating leucyl-tRNA synthetase because a mutant with an apparent deficiency in leucine transport displays a 3-fold elevation in enzyme level (8). Several lines of evidence also suggest a role for leucyl-tRNA synthetase in regulating leucine transport in Chinese hamster ovary cells (9–13). We report here that sequences in the potential nucleotide-binding domain of leucyl-tRNA synthetase display a high degree of sequence and structural similarity to the carboxyl-terminal domain of the leucine-specific binding protein.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

RESULTS AND DISCUSSION

Sequence Similarity Between Leucyl-tRNA Synthetase and Leucine-Specific Binding Protein. The leucine-specific binding protein is synthesized as a 369-amino acid precursor (14). The precursor undergoes processing to a mature form containing 346 amino acids by cleavage of a signal peptide at the amino terminus (15). Leucyl-tRNA synthetase is a monomer of 860 amino acids (16). A region of the leucine-specific binding protein, 121 residues in length and comprising 36% of the mature binding protein, was found to display similarity with two regions of leucyl-tRNA synthetase that are separated by 14 amino acids. Identical amino acids were aligned, and gaps were introduced to improve similarity according to the algorithm of Needleman and Wunsch (17). Sequences were also compared by application of the logarithm-of-odds scoring matrix of Dayhoff *et al.* (18), which rates the probability for conservative replacements of amino acids based on observed changes in many closely related sequences. The scores obtained are described as relatedness odds for an evolutionary distance of 250 accepted point mutations (PAM), which express the frequency at which similarities seen between two sequences of aligned residues occur among related proteins in excess of chance expectations. In addition to identities and conservative replacements, we also considered cases in which amino acids could be derived from one another by a single base change in the corresponding codons from inspection of the nucleotide sequences (19). When a residue match resulted from a single base change or an accepted replacement, it was scored only once. The significance of the alignments was estimated from Z values obtained by comparisons for similarity of one sequence with 20 randomized permutations of the second sequence. Each of the randomized permuted sequences had the same length and amino acid composition as the native sequence.

Fig. 1 shows that residues 149–236 of the leucine-specific binding protein can be aligned with residues 507–589 of leucyl-tRNA synthetase (region I) to achieve 41.1% similarity (22 identities and 15 conservative matches of 90 positions). Region II (residues 237–269 of the leucine-specific binding protein and residues 604–636 of leucyl-tRNA synthetase) was found to display 44.1% similarity with 10 identities and 5 conservative matches of 34 positions. An interruption occurs in the similarity from residues 590–603 in the leucyl-tRNA synthetase sequence. Regions I and II contain, respectively, 10 and 4 amino acids which, from inspection of the nucleotide sequences, could be derived from one another by a single base change in the corresponding codons. These represent ≈13% (region I) and 17% (region II) of the differences between the two polypeptides and would increase the overall similarities seen between leucyl-tRNA synthetase

*To whom repint requests should be addressed.

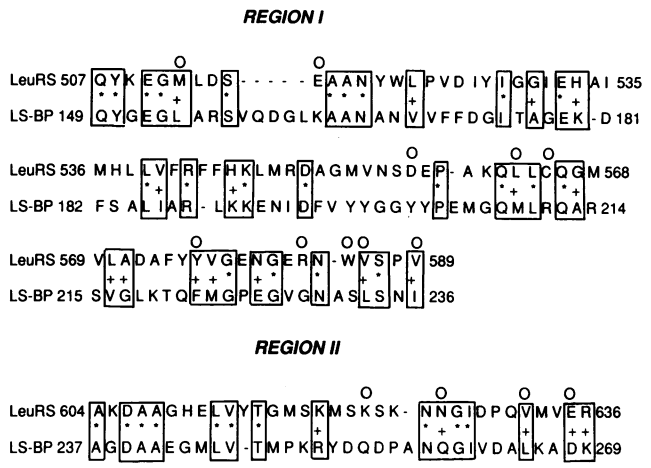

FIG. 1. Comparative sequence analysis of the *E. coli* leucine-specific binding protein (LS-BP) residues 149–269 and leucyl-tRNA synthetase (LeuRS) residues 507–636. Identical residues (*) and conservative replacements (+) are boxed. Amino acids that could be derived from one another by a single base change in the corresponding codons from inspection of the nucleotide sequences are indicated by (○). The conservative amino acid (in one-letter code) replacement groups used are: A,G; D,E,N,Q; I,L,V,M; F,W,Y; H,R,K; S,T; C; P.

and leucine-specific binding proteins to ≈47% for both regions I and II. Many of the single base codon changes result in conservative substitutions. Relatively few gaps were required in the regional alignments to achieve the observed level of matching residues (nine gaps in region I, five of which are contiguous, and two gaps in region II). From relatedness odds established by Dayhoff *et al.* (18), the similarities among aligned amino acids are observed 1.2×10^3 (region I) and 251.1 (region II) times more frequently among related proteins than random chance would predict. *Z* values obtained from randomized comparisons for similarity were 20.15 for region I and 9.21 for region II, suggesting probable significance. The leucine-specific binding protein is ≈60% smaller in size than leucyl-tRNA synthetase. No extensive sequence similarities were found between the binding protein and the first 506 residues and last 164 residues of the synthetase.

Structural Similarities. The leucine-specific binding protein has been crystallized, and its structure has been determined by Quiocho and coworkers using x-ray crystallographic analysis (20). This structural analysis of the tertiary structure indicates that the protein has a bilobate structure consisting of two globular domains (20). Each domain is composed of a central β-pleated sheet flanked on either side by, at least, two α-helices. From a correlation of regions of sequence similarity between the leucine-specific binding protein and leucyl-tRNA synthetase with the corresponding secondary structural elements of the binding protein tertiary structure, regions of sequence similarity extend throughout most of the carboxyl-terminal domain of the binding protein and extends to a hinge region that separates each globular domain. Most of residues 148–269 involved in the similarity are clustered on the surface of the carboxyl-terminal domain from inspection of the three-dimensional structure. The secondary structural element boundaries for residues 148–269, indicated in Fig. 2, are from the most recent structural analyses of the leucine-specific binding protein (20). Letter and numerical designations have been assigned to the individual elements of β-structure and α-helix, respectively. The algorithms of Kyte and Doolittle (21) and Garnier *et al.* (22) were used to compare the secondary structure predictions and hydrophobicities for the sequence of leucyl-tRNA synthetase with the known structure for the leucine-specific binding protein. Similarities exist in both the distributions of α-helical and

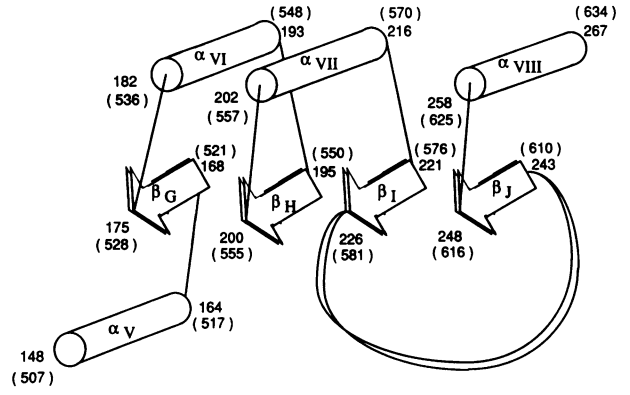

FIG. 2. Schematic representation of the three-dimensional structure for residues 149–269 of the *E. coli* leucine-specific binding protein carboxyl-terminal domain with sequence similarity to leucyl-tRNA synthetase. Boundaries for the α-helices (cylinders) and β-strands (arrows) are indicated. Corresponding amino acids in leucyl-tRNA synthetase are numbered in parentheses.

β-sheet structures and in the distribution of hydrophobic and hydrophilic regions (Fig. 3). Predictably, primary sequence similarities between the leucine-specific binding protein and leucyl-tRNA synthetase could extend to structural similarities between the two polypeptides.

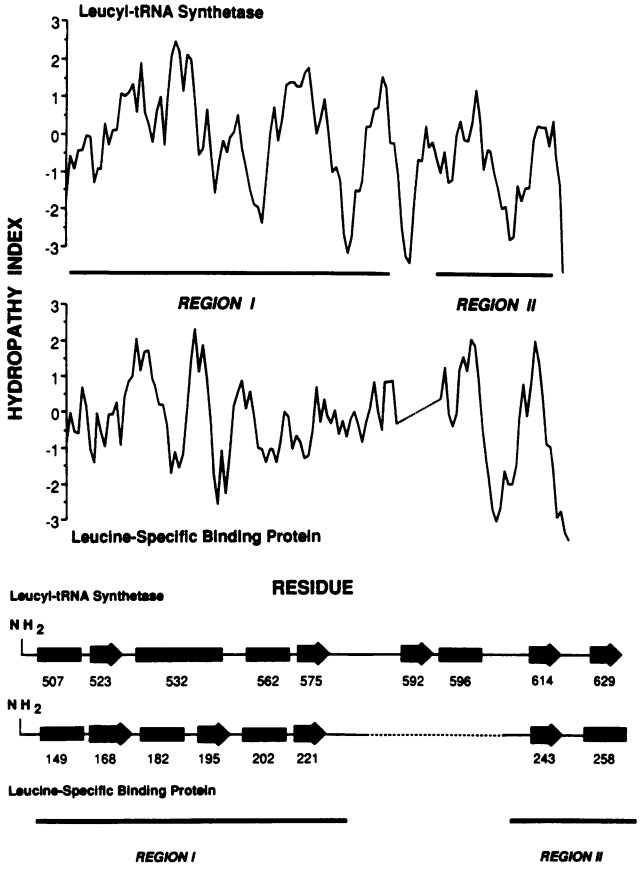

FIG. 3. Structural comparison of the leucine-specific binding protein residues 149–269 and leucyl-tRNA synthetase residues 507–636. The hydropathy index was determined from implementation of the Kyte–Doolittle algorithm (22) for consecutive sequence spans six residues in length. Predictions of secondary structure were determined by the algorithm of Garnier *et al.* (22). Secondary structural elements are indicated by rectangle (α-helices) or arrows (β-strands) along the primary sequence with the start position of each boundary given below each element. ---, sequence interruption between regions I and II from sequence alignment in Fig. 1.

Similarity Associated with the Putative Nucleotide-Binding Domain of Leucyl-tRNA Synthetase. The 3-dimensional structure of leucyl-tRNA synthetase is unknown. However, as shown in Fig. 4, a nucleotide-binding fold is highly likely to be present in leucyl-tRNA synthetase based on the conservation of amino acids to the known structure of the nucleotide-binding fold of the closely related methionyl-tRNA synthetase (23, 24) and the functional relatedness of the two enzymes. The stretch of residues from leucyl-tRNA synthetase that display similarities to the leucine-specific binding protein appear to lie within the putative nucleotide-binding fold of leucyl-tRNA synthetase (Fig. 4) and are bound by two highly conserved sequences common to leucyl-, isoleucyl, valyl-, and methionyl-tRNA synthetases. Leucyl-, isoleucyl, valyl-, and methionyl-tRNA synthetases form a functionally and evolutionarily related family of enzymes having diverged from a common ancestral gene (16, 25–27). Near the β_D secondary structural element in the established nucleotide-binding fold of the methionine enzyme are residues that share the consensus sequence DWCIRS (in one-letter code) to which the related synthetases conform (16, 25, 26). A connective polypeptide segment that interrupts the nucleotide-binding fold begins near the DWCIRS consensus sequence at position 230 in methionyl-tRNA synthetase (24). Located between β_E and α_F is the conserved consensus KMSKS sequence thought to crosslink with the 3′ end of tRNA (28). Region II shown in Fig. 1 spans the consensus KMSKS sequence of leucyl-tRNA synthetase. Although sequence similarity is found with the leucine-specific binding protein surrounding the conserved KMSKS sequence, there is a demonstrable absence of matching identities with the conserved sequence. It is interesting to note that if the nucleotide-binding domains of leucyl- and methionyl-tRNA synthetases are highly conserved with positional conservation of connective polypeptide segments, most similarity seen between the leucine synthetase enzyme and binding protein would likely correspond to the connective polypeptide interruption and immediate adjoining sequence.

Similarity Between Leucine-Binding Protein and Leucyl-tRNA Synthetase Appears to be an Addition to Leucyl-tRNA Synthetase That Is Absent in Related Synthetases. We compared the leucine-specific binding protein with corresponding regions bounded by the DWCIRS and KMSKS conserved residues in the related isoleucyl-, valyl-, and methionyl-tRNA synthetases by using group alignments reported by Hartlein and Madern (16). In this region of conserved sequences, similarity between the four synthetases appear to partition into two groups where closer relationships occur between valyl- and methionyl-tRNA synthetases and between leucyl- and isoleucyl-tRNA synthetases. We found that a significant part of the region of leucyl-tRNA synthetase (residues 512–568) that is similar to the leucine-specific binding protein corresponds to a large gap ≈73 residues in length in the isoleucyl-, valyl-, and methionyl-tRNA synthetase alignments (Fig. 5). Furthermore, the length between DWCIRS and KMSKS conserved residues in leucyl-tRNA synthetase is ≈40% larger than that seen in isoleucyl-, valyl-, and methionyl-tRNA synthetases having a length of 194 residues. The sizes of these segments for isoleucyl-, valyl-, and methionyl-tRNA synthetases have lengths of 135, 126, and 98 residues, respectively. We postulate that a large polypeptide insertion with sequence and structural similarities to the leucine-specific binding protein could have been incorporated into the leucyl-tRNA synthetase structure after divergence from the other related enzymes, and this insertion could account for the large differences in length.

Evolutionary Implications. Aminoacyl-tRNA synthetases represent an ancient family of enzymes that are widely diverse in length and quaternary structure but contain similar arrangements of functional units along their respective amino acid sequences (29). Size diversity among the aminoacyl-tRNA synthetases has been attributed to insertions of additional polypeptides to the basic structure of the enzymes, the insertions being dispensible for enzyme catalysis (30). These extra polypeptide domains are thought to endow synthetases with other biological functions in addition to their role in

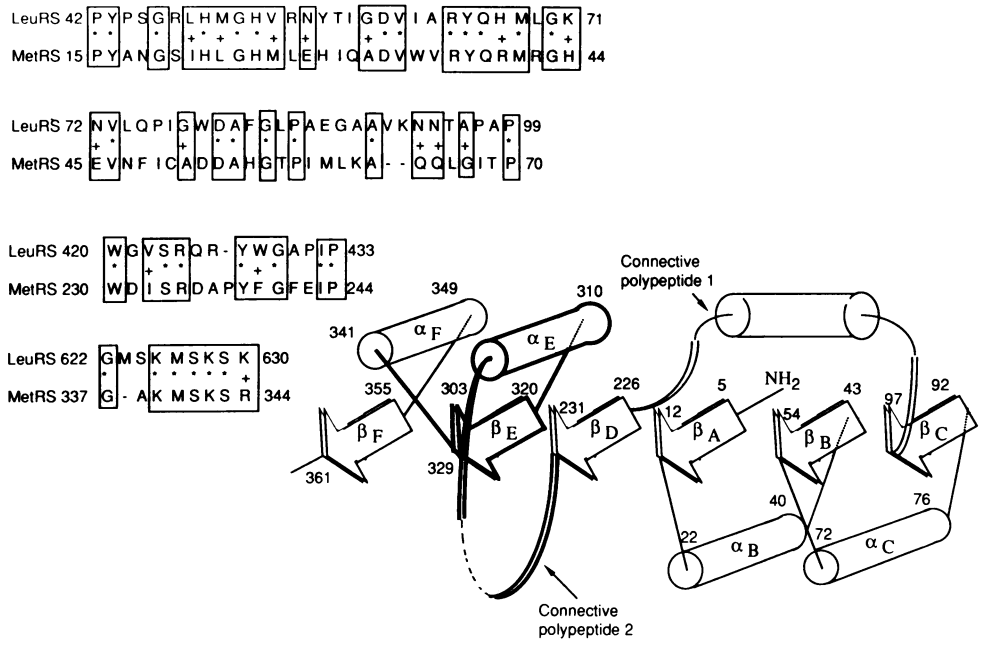

FIG. 4. Comparative sequence analysis of *E. coli* leucyl- (LeuRS) and methionyl-tRNA synthetases (MetRS), with schematic representation of the MetRS nucleotide-binding fold. Identical residues (*) and conservative replacements (+) are boxed. The conservative replacement groups are as described in Fig. 1. Boundaries for the α-helices (cylinders) and β-strands (arrows) of the MetRS nucleotide-binding fold are numbered with boundaries indicated. Connective polypeptides that interrupt the nucleotide-binding fold are indicated. Boldface structural elements indicate the location within the potential nucleotide-binding domain of LeuRS which displays similarity to the LS-BP. (This schematic is a modification of figure 7*b*, ref. 23, and figure 1, ref. 24).

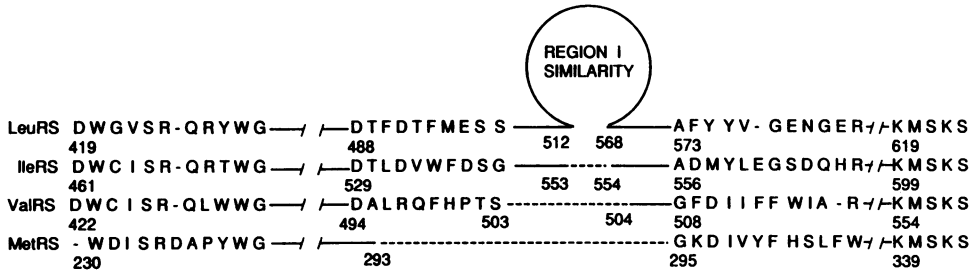

FIG. 5. Abbreviated alignment of leucyl- (LeuRS), isoleucyl- (IleRS), valyl- (ValRS), and methionyl-tRNA (MetRS) synthetases between the DWCIRS and KMSKS conserved sequences. ---, Positions where a large gap occurs in the sequence. The start position of conserved regions and boundaries surrounding a central gap in the alignment are numbered. Sequence alignments reflect the closer similarity of the valyl- with methionyl-tRNA synthetases and the leucyl- with isoleucyl-tRNA synthetases. Detailed alignments of the four synthetase sequences in this region have been reported by Hartlein and Madern (16).

initiating protein synthesis. However, the origins and evolution of the additional polypeptides joined to the basic structural format of aminoacyl-tRNA synthetases are largely unknown. An ancestral gene for the current leucine-specific binding protein could have had genetic interactions with an ancestral leucyl-tRNA synthetase gene. A gene-fusion event could then have increased the overall size of the synthetase but may not have been detrimental for enzyme catalysis. Such a gene-fusion event could be viewed as an insertion of sequences into the nucleotide-binding domain of the synthetase, such that catalysis still occurs (30). Deletion studies on isoleucyl-tRNA synthetase provide evidence that the nucleotide-binding fold might be able to accomodate sequence insertions without affecting enzyme catalysis (24). It is also interesting to note that artificial insertions have been made into a connective polypeptide segment in the nucleotide-binding fold of methionyl-tRNA synthetase (31). Many proteins with insertions were found to be stable and have activity (31), which further shows that the nucleotide-binding fold of a synthetase can accomodate the insertion of foreign sequences. Furthermore, genetic interactions between limited regions of ancestral genes for leucyl-tRNA synthetase and leucine-specific binding protein could also account for the lack of extensive sequence similarity outside the regions of high sequence conservation or identity. Potentially, ancestral gene-fusion events may also have been part of the history of other aminoacyl-tRNA synthetases, which would account for some of the reported size diversity among these enzymes. These findings illustrate the accumulating evidence that suggests aminoacyl-tRNA synthetases or certain domains of these proteins are related to proteins involved in other cellular processes. Mitochondrial tryosyl- and leucyl-tRNA synthetases have been shown to be involved in mRNA splicing in yeast (32, 33). Recently, similarities have been reported between tyrosyl-tRNA synthetase and the human estrogen receptor (34), and a common stretch of amino acids has been reported to occur between the sulfate-binding protein and glycyl-tRNA synthetase (35). Inserted domains within aminoacyl-tRNA synthetases could also be the source for evolution of additional or alternative metabolic and regulatory functions.

This work was supported by Grant GM11024 to D.L.O. from the National Institutes of Health. R.M.W. is a Postdoctoral Fellow of the Ford Foundation and recipient of an Arthur F. Thurnau Postdoctoral Fellowship from the University of Michigan.

1. Landick, R., Ames, G. F.-L. & Oxender, D. L. (1985) in *The Enzymes of Biological Membranes*, ed. Martonosi, A. (Plenum, New York), 2nd Ed., Vol. 3, pp. 577–615.
2. Anderson, J. J. & Oxender, D. L. (1978) *J. Bacteriol.* **136,** 168–174.
3. Anderson, J. J. & Oxender, D. L. (1977) *J. Bacteriol.* **136,** 384–392.
4. Oxender, D. L., Anderson, J. J., Daniels, C. J., Landick, R., Gunsalus, R. P., Zurawski, G., Selker, E. & Yanofsky, C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 1412–1416.
5. Landick, R., Anderson, J. J., Mayo, M. M., Gunsalus, R. P., Mavromara, P., Daniels, C. J. & Oxender, D. L. (1980) *J. Supramol. Struct.* **14,** 527–537.
6. Nazos, P. M. (1984) Dissertation (The University of Michigan, Ann Arbor).
7. Quay, S. C., Kline, E. L. & Oxender, D. L. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 3921–3924.
8. Jackson, J. H., Williams, L. S. & Umbarger, H. E. (1974) *J. Bacteriol.* **120,** 1380–1386.
9. Moore, P. A., Jayme, D. W. & Oxender, D. L. (1977) *J. Biol. Chem.* **252,** 7427–7430.
10. Lobaton, C. D., Moreno, A. & Oxender, D. L. (1984) *Mol. Cell. Biol.* **4,** 475–483.
11. Shotwell, M. A., Mattes, P. M., Jayme, D. W. & Oxender, D. L. (1982) *J. Biol. Chem.* **257,** 2974–2980.
12. Shotwell, M. A., Collarini, E. J., Mansukhani, A., Hampel, A. E. & Oxender, D. L. (1983) *J. Biol. Chem.* **258,** 8183–8187.
13. Collarini, E. J., Campbell, G. S. & Oxender, D. L. (1989) *J. Biol. Chem.* **264,** 15856–15862.
14. Landick, R. & Oxender, D. L. (1985) *J. Biol. Chem.* **260,** 8257–8261.
15. Oxender, D. L., Anderson, J. J., Daniels, C. J., Landick, R., Gunsalus, R. P., Zurawski, G. & Yanofsky, C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 2005–2009.
16. Hartlein, M. & Madern, D. (1987) *Nucleic Acids Res.* **24,** 10199–10209.
17. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48,** 443–453.
18. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), Vol. 5, pp. 345–358.
19. Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), Suppl. 3, Vol. 5.
20. Sack, J. S., Trakhanov, S. D., Tsigannik, I. H. & Quiocho, F. A. (1989) *J. Mol. Biol.* **206,** 193–207.
21. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
22. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120.
23. Zelwer, C., Risler, J.-L. & Brunie, S. (1982) *J. Mol. Biol.* **155,** 63–81.
24. Starzyk, R. M., Webster, T. A. & Schimmel, P. (1987) *Science* **237,** 1614–1618.
25. Webster, T., Kula, T. H., Mackie, M. & Schimmel, P. (1984) *Science* **226,** 1315–1317.
26. Heck, J. D. & Hatfield, G. W. (1988) *J. Biol. Chem.* **263,** 868–877.
27. Wetzel, R. (1978) *Origins Life* **9,** 39–50.
28. Siebenlist, U., Simpson, R. B. & Gilbert, W. (1980) *Cell* **20,** 269–281.
29. Schimmel, P. R. & Soll, D. (1979) *Annu. Rev. Biochem.* **48,** 601–648.

30. Schimmel, P. R. (1987) *Annu. Rev. Biochem.* **56**, 125–158.
31. Starzyk, R. M., Burbaum, J. J. & Schimmel, P. (1989) *Biochemistry* **28**, 8479–8484.
32. Akins, R. A. & Lambowitz, A. M. (1987) *Cell* **50**, 331–345.
33. Labouesse, M., Herbert, C. J., Dujardin, G. & Slonimski, P. P. (1987) *EMBO J.* **6**, 713–721.
34. Baker, M. E. (1989) *FASEB J.* **3**, 2086–2088.
35. McCaldon, P. & Argos, P. (1988) *Proteins* **4**, 99–122.