SEQUENCE HOMOLOGIES BETWEEN THE TRYPTOPHANYL tRNA SYNTHETASES OF *BACILLUS STEAROTHERMOPHILUS* AND *ESCHERICHIA COLI*

G. P. WINTER and B. S. HARTLEY,
Department of Biochemistry, Imperial College of Science & Technology, London SW7 2AZ

and

A. D. McLACHLAN*
MRC Laboratory of Molecular Biology, University Postgraduate Medical School, Hills Road, Cambridge CB2 2QH, England

and

M. LEE* and K. H. MUENCH
Division of Genetic Medicine, Departments of Medicine and Biochemistry, University of Miami School of Medicine, Miami, Florida 33152, USA

Received 22 August 1977

1. Introduction

The tryptophanyl tRNA synthetases of *Escherichia coli* and *Bacillus stearothermophilus* are dimers of identical mol. wt (2 × 37 000) and have similar amino acid compositions [1,2]. One would expect the amino acid sequences of the same enzyme from closely related organisms to show considerable homology and this homology to be greatest in the critical regions of the molecule, such as those involved in substrate binding and catalysis [3]. We have therefore made systematic alignments of the partial sequences of the *E. coli* enzyme [4] with the *B. stearothermophilus* sequence [2] in an attempt to identify these regions.

2. Methods

By computer methods [5] every *E. coli* peptide (fig.1) was first scored systematically for homology with every part of the *B. stearothermophilus* sequence. The results were expressed in a two-dimensional matrix in which contour symbols were used to show whether the weighted score around any residue exceeded a series

AABZXLLGP.....	LACGIBZPK
AITVR	LVPVGBZZK
ASAH.....	LR
ATLBLY	MGALR
AVTBSBZPPVR	MSK
AVYZAXGFV....	MTZFK
AYZBIG.....	SAR
BBXXGLLZBP....	SBBBR
BGAZK	SGAR
BIAZR	SIPZLZZK
BK	SVVK
CIVBZHAITVR	TIP.....
FBALYGZIFK	TKPIVFXGAEPXGEL
FIFGXLTK	XXGXYXGAL.....
GZBZK	TLK
IK	VAWPK
IPK	VPZP....
	WALBCYTYFGELSR
	YBZVK

Fig.1. *E. coli* peptides derived mainly from a tryptic digest [4] which have been tested for homology with the *B. stearothermophilus* sequence.

of threshold values. This allowed the identification of plausible homologous alignments. The unweighted score was calculated for each plausible alignment and the maximum score, for an identity, obtained by comparing each *E. coli* peptide with itself. The *B. stearothermophilus* enzyme was taken as the standard for composition, and the single matching probability calculated for each alignment. In a comparison of two proteins each of about 300 amino acids there are approx. 10^5 pairs of peptides of a given small length which can be matched. Thus a matching probability of 10^{-5} or more may easily arise by chance and need not be significant. Conversely, a matching probability below 10^{-8} is highly significant. Finally the alignments of fig.2 were made, in which conflicts were resolved by giving priority to the peptide with the lower single matching probability. For example, residues 140–148 of the *B. stearothermophilus* sequence (IVPVGEDQK) match up with both the *E. coli* peptides LVPVGBZZK (p 4.1 × 10^{-7}) and GZBZK (p 9.0 × 10^{-6}). The former alignment was therefore given priority.

Sequence in one letter code	Residue numbers
MKTI-FSGIQPSGVITIGNYIGALRQ	B.S. 1 - 25
TKPIVFXGAEPXGELXXGXYXGALR	E.C.
FVELQHZYNCYFCIVBZHAITVWQD	B.S. 26 - 50
CIVBZHAITVR	E.C.
PHELRQNIRRLAALYLAVGIDPTQA	B.S. 51 - 75
LACGIBZPK	E.C.
TLFIQSEVPAHAQAAWMLQCIVYIG	B.S. 76 - 100
WALBCYTYFG	E.C.
ELERMTQFKEKSAGKEAVSAGLLTY	B.S. 101- 125
ELSRMTZFK	E.C.
PPLMAADILLYNTDIVPVGEDQKQH	B.S. 126- 150
LVPVGBZZK	E.C.
IELTRDLAERFDKRYGELFTIPEAR	B.S. 151- 175
BIAZRFBALYGZIFK	E.C.
IPKVGARIMSLVDPTKKMSKSDPNP	B.S. 176- 200
IPKSGAR MSK	E.C.
KAYITLLDDAKTIEKKIKSAVTDS-E	B.S. 201- 225
IK AVTBSBZ	E.C.
GTIRYDKEAKPGISNLLNIYSTLSG	B.S. 226- 250
PPVR	E.C.
QSIEELERQYEGKGYGVFKADLAQV	B.S. 251- 275
SIPZLZZK	E.C.
VIETLRPIQERYHHWMESEELDRVL	B.S. 276- 300
	E.C.
DEGAEKANRVASEMVRKMEQAMGLG	B.S. 301- 325
BGAZK	E.C.
RR.	B.S. 326- 327
	E.C.

Fig.2. Alignment of *E. coli peptides* with *B. stearothermophilus* sequence. The single matching probability (p) is marked below each alignment as follows: (a) ———— $10^{-8} > p$ (b) — —— $10^{-5} > p > 10^{-8}$ (c) . . . . . $10^{-4} > p > 10^{-5}$ (d) $p > 10^{-4}$ Only small peptides, identical to or very similar to the *B. stearothermophilus* sequence, have been inserted with single matching probabilities of $> 10^{-4}$. Residues totalling over two-thirds of the sequence have been compared and tentative alignments corresponding to one-third of the sequence have been made.

3. Results and discussion

The striking identity at positions 38–47 suggests that this portion of the sequence may be involved in highly critical interactions. A critical portion of sequence, essential for pyrophosphate exchange and tRNA loading reactions, has also been identified near the N-terminus of the proteolytically trimmed tryptophanyl tRNA synthetase from beef pancreas [6].

The overall conservation of sequence is markedly greater in the N-terminal third than in the C-terminal two-thirds of the molecule. This may be related to a double-domain structure of the *B. stearothermophilus* enzyme which is detectable by mild proteolysis: the native subunit can be split by a range of proteases into a 24 000 dalton fragment corresponding to the C-terminal two thirds of the molecule [7]. Since a similar fragment can be derived from proteolysis of the tryptophanyl tRNA synthetase from beef pancreas [8], this may well reflect the division of these enzymes into functional domains [6,9].

Of the three cysteine residues identified so far in the *E. coli* enzyme [10,11], only two, Cys 38 and Cys 95, are conserved. The third cysteine seems to be replaced by Val 68 in the *B. stearothermophilus* sequence, although it is apparently conserved in the human placental enzyme [12]. If we assume that this is indeed the required cysteine of the *E. coli* enzyme, as has been tentatively suggested [11], then this residue can have no catalytic role.

Acknowledgements

We are grateful to Dr C. J. Bruton for discussions and for his comments on the manuscript.

References

[1] Joseph, D. R. and Muench, K. H. (1971) J. Biol. Chem. 246, 7602–7609.
[2] Winter, G. P. and Hartley, B. S. (1977) FEBS Lett. 80, 340–342.
[3] Dayhoff, M. O., Barker, W. C. and Hardman, J. K. (1972) in: Atlas of Protein Sequence and Structure (Dayhoff, M. O. ed) Vol. 5, pp. 53–66, National Biomedical Research Foundation, Washington D.C.
[4] Lee, M., Safille, P. A., del Valle, U., Kuehl, G. V. and Muench, K. H. (1976) 10th Int. Cong. Biochem., Hamburg.
[5] McLachlan, A. D. (1971) J. Mol. Biol. 61, 409–424.
[6] Epely, S., Gros, C., Labouesse, J. and Lemaire, G. (1976) Eur. J. Biochem. 61, 139–146.
[7] Winter, G. P. unpublished.
[8] Prasolov, V. S., Favorova, O., Margulis, G. V. and Kisselev, L. L. (1975) Biochim. Biophys. Acta 378, 92–106.
[9] Winter, G. P. (1976) Ph. D. thesis Cambridge University, November. 1976.
[10] Muench, K. H., Lipscomb, M. S., Lee, M. and Kuehl, G. V. (1975) Science 187, 1089–1091.
[11] Kuehl, G. V., Lee, M. and Muench, K. H. (1976) J. Biol. Chem. 251, 3254–3260.
[12] Lipscomb, M. S., Lee, M. and Muench, K. H. (1976) J. Biol. Chem. 251, 3261–3268.