THE AMINO ACID SEQUENCE OF TRYPTOPHANYL tRNA SYNTHETASE FROM *BACILLUS STEAROTHERMOPHILUS*

G. P. WINTER and B. S. HARTLEY

Department of Biochemistry, Imperial College of Science & Technology, London SW7 2AZ, England

Received 11 June 1977

1. Introduction

Amino acyl tRNA synthetases of different amino acid specificities have diverse quaternary structures, ranging from dimers of 2 × 37 000 daltons to monomers of 1 × 110 000 daltons [1]. Prima facie, it seems unlikely that these enzymes will have the kind of sequence homology shown by the serine proteases [2]. Peptide mapping, isolation of several tryptic peptides in greater than molar yield, and more detailed sequence studies have revealed, however, that the enzymes with large polypeptide chains (isoleucyl-, leucyl-, methionyl- and valyl-tRNA synthetases) contain extensive internal repetition of sequence [3–6]. The larger chains could therefore have arisen by duplication and fusion of common ancestral genes originally coding for chains around 40 000 daltons.

Comparisons of the primary and tertiary [7–8] structures of several of these enzymes may resolve this question. Complete amino acid sequences are obviously desirable for such comparisons, and as the first contribution we offer the sequence of tryptophanyl tRNA synthetase from *Bacillus stearothermophilus*, a dimer of 2 × 37 000 daltons.

2. Materials and methods

Tryptophanyl tRNA synthetase was purified from the early stages of a large scale purification of several amino acyl tRNA synthetases from *Bacillus stearothermophilus*. Approximately 300 mg of this enzyme could be obtained from 50 kg of wet cells [9]. An N-terminal sequence determination was made on the carboxymethylated protein by automated Edman degradation in a Beckmann sequenator: 30 residues were obtained without ambiguity. Subsequently the protein was subjected to 3 major digests: tryptic, chymotryptic and cyanogen bromide. Peptides were separated by gel-filtration, ion exchange on DEAE-cellulose and sulphonated polystyrene (Locarte amino acid analyser resin), paper chromatography and high voltage paper electrophoresis at pH values 2.1, 3.5 and 6.5. Sequences were determined manually by the dansyl-Edman technique [10], and occasionally by mass spectrometry [11]. Subdigestions of the peptides with trypsin, chymotrypsin, elastase, pepsin, thermolysin, staphyloccocal protease, cyanogen bromide or acid cleavage at aspartic acid, were used extensively to check the dansyl-Edman sequence and to help assign the amides. Although most of the amides were assigned by measuring peptide electrophoretic mobilities, a few were assigned by thin-layer chromatography of the P.T.H. derivatives after automated sequencing or directly by mass spectrometry.

3. Results

The amino acid sequence is shown in fig.1. All amides were assigned with the exception of those at residues 32, 41 and 42. However, position 32 is probably glutamic acid since it can be cleaved by staphyloccocal protease and at least one of the positions 41 and 42 is amidated since the thermolytic peptide Ile–Val–Glx–Asx–His–Ala (residues 39–44) is electrophoretically neutral at pH 6.5.

Ten fragments, Flc (residues 1–61), cF2c (residues 62–65), cF3c (residues 66–78), cF4c (residues 79–91), cF5c (residues 92–108), cF6c (residues

North-Holland Publishing Company – Amsterdam

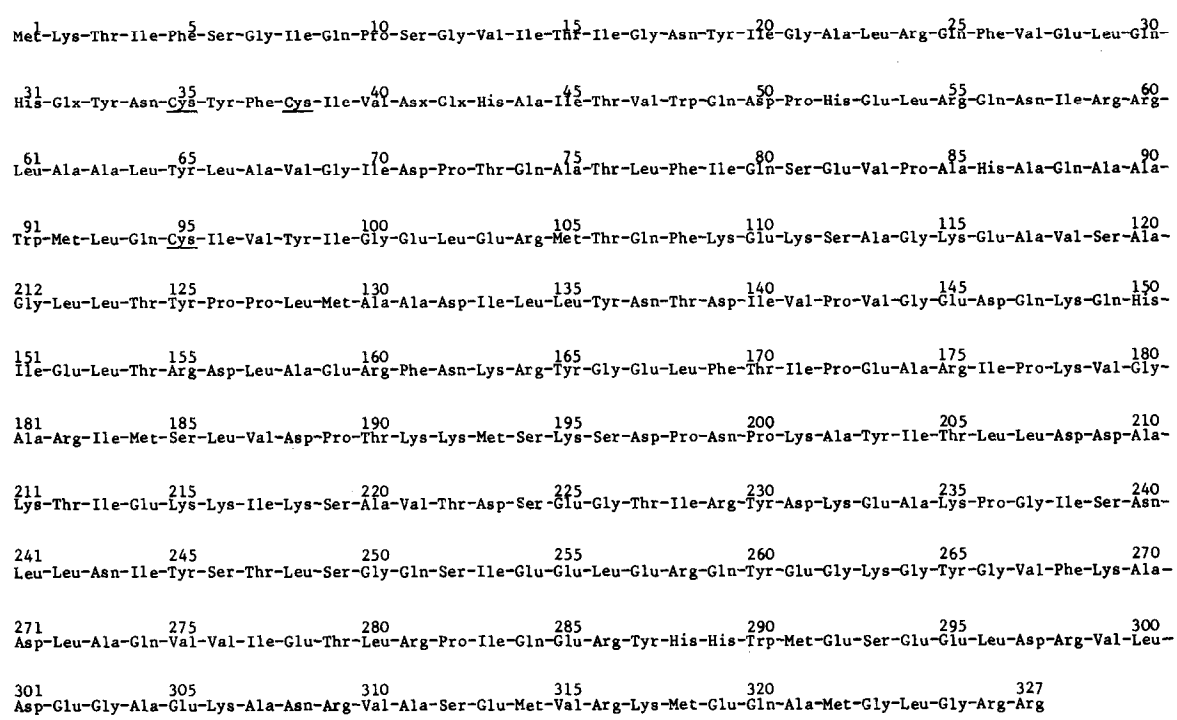

Fig.1. The amino acid sequence of tryptophanyl tRNA synthetase from *Bacillus stearothermophilus.*

109–161), tF7c (residues 161–230), tF8t (residues 230–269), cF9c (residues 269–287), and tF10 (residues 287–327) can be constructed by overlapping peptides from the tryptic and chymotryptic digests. The N-terminus and C-terminus of each fragment are denoted t or c depending on whether they were derived from a tryptic or chymotryptic split respectively, and each fragment is internally overlapped by at least two residues.

Fl was clearly the N-terminal fragment from the sequencer analysis and F10 the C-terminal fragment as demonstrated by the isolation of peptides Gly–Arg–Arg (residues 325–327) from the chymotryptic digest and Gly–Leu–Gly–Arg–Arg (residues 323–327) from the cyanogen bromide digest.

The one residue overlaps between F6, F7, F8, F9 and F10 were resolved by noting which tryptic peptides were present in the cyanogen bromide fragments corresponding to residues 130–184 and 194–314 (internal methionine or homoserine) and confirmed by amino acid analysis of these cyanogen bromide fragments. The overlap between F5 and F6 was found by isolating the peptide Thr (Gln, Phe, Lys) Glu (residues 106–110) from a staphyloccocal protease digest of a mixture of cyanogen bromide fragments. The overlaps between F2 and F3, and between F3 and F4 were found by isolating the peptides Ala–Ala–Leu–Tyr–Leu (residues 62–66) and Leu–Phe–Ile–Glx–Ser–Glx (Val, Pro, Ala) (residues 77–85) from an elastase digest of the impure hydrophobic cyanogen bromide fragment corresponding predominantly to residues 1–92, and confirmed by the isolation in low yield of a peptide Tyr–Leu–Ala–Val–Gly . . . . (residues 65–69 . . . .) from a peptic digest of the insoluble 'core' remaining at the end of the chymotryptic digest.

The chain now lay in three pieces, F1, F2–F4 and F5–F10, which assuming no outstanding sequence must abutt directly. The possibility of unknown sequence lying between these three pieces is remote: in each of the chymotryptic and tryptic digests peptides corresponding to over 95% of the sequence

were isolated and yet no peptides were ever found which could not be inserted in the proposed sequence. Furthermore, the amino acid composition and molecular weight of the enzyme tally with those predicted by the sequence.

4. Discussion

We can detect, by eye or computer [12], no homologous internal repetition of sequence and no significant homology between this sequence and any of the partial sequences made available to us from the tyrosyl-tRNA synthetase of *B. stearothermophilus* or methionyl-tRNA synthetase of *E. coli.* Partial sequences of the tryptophanyl-tRNA synthetase from *E. coli* do, however, reveal considerable homologies with the *B. stearothermophilus* enzyme. [13,14].

Acknowledgements

We would like to thank Mr F. Northrop for the N-terminal sequence determination; Dr H. Morris and Dr A. Dell for help with mass spectrometric sequencing; Dr A. McLachlan for the computer search for internal homologies; Dr G. Koch and Dr C. J. Bruton for making available partial sequences of tyrosyl tRNA synthetase *(B. stearothermophilus)* and the methionyl tRNA synthetase (*E. coli*), and for discussions on the sequence strategy. The work originated at the M. R. C. Laboratory for Molecular Biology, Cambridge, and was completed at Imperial College, London.

References

[1] Kisselev, L. L. and Favorova, O. O. (1974) Advances in Enzymology 40, 141–238.
[2] Hartley, B. S., Brown, J. R., Kauffman, D. L. and Smillie, L. B. (1965) Nature 207, 1157–1159.
[3] Kula, M. R. (1973) FEBS Lett. 35, 299–302.
[4] Koch, G. L. E., Boulanger, Y. and Hartley, B. S. (1974) Nature 249, 316–320.
[5] Waterson, R. M. and Konigsberg, W. H. (1974) Proc. Natl. Acad. Sci. USA 71, 376–380.
[6] Bruton, C. J. (1975) Biochem. J. 147, 191–192.
[7] Irwin, M. J., Nyborg, J., Reid, B. R. and Blow, D. M. (1976) J. Mol. Biol. 105, 577–586.
[8] Zelwer, C., Risler, J.-L. and Monteilhet, C. (1976) J. Mol. Biol. 102, 93–101.
[9] Atkinson, A. Process Biochem. August 1973.
[10] Hartley, B. S. (1970) Biochem. J. 119, 805–822.
[11] Morris, H. R., Williams, D. H., Midwinter, G. G. and Hartley, B. S. (1974) Biochem. J. 141, 701–713.
[12] McLachlan, A. D. (1971) J. Mol. Biol. 61, 409–424.
[13] Muench, K. H., Lipscomb, M. S. Lee, M. L. and Kuehl, G. V. (1975) Science 187, 1089–1091.
[14] Winter, G. P., Hartley, B. S., McLachlan, A. D., Muench, K. M. and Lee, M. L., Manuscript in preparation.