The tyrosyl-tRNA synthetase from *Escherichia coli*

Complete nucleotide sequence of the structural gene

D.G. Barker[+*], C.J. Bruton[•] and G. Winter[†]

[+]*Laboratoire de Biochimie, Institut de Biologie Moléculaire et Cellulaire du Centre National de la Recherche Scientifique, Strasbourg, France,* [•]*Department of Biochemistry, Imperial College of Science and Technology, London* and [†]*Medical Research Council Laboratory of Molecular Biology, Hills Rd., Cambridge CB2 2QH, England*

Received 5 November 1982

The structural component of the *tyrS* gene of *Escherichia coli*, comprising 1269 base pairs, has been fully sequenced by the combined M13/dideoxychain termination approach. The gene has a codon usage pattern which is typical of highly expressed proteins and similar to other *Escherichia coli* aminoacyl-tRNA synthetase genes. Peptide purification and sequencing has been used to locate the N-terminus and to provide confirmation of 95% of the translated protein sequence. This latter yields on M_r of 47403 for the *Escherichia coli* tyrosyl-tRNA synthetase, and reveals considerable homology with the primary structure of the analogous enzyme isolated from *Bacillus staerothermophilus*.

Aminoacyl-tRNA synthetase *M13 cloning* *DNA sequencing*

1. INTRODUCTION

One of the prime interests in studying the structures of the aminoacyl-tRNA synthetases concerns the nature and specificity of the interaction between the enzyme and its substrates: ATP, amino acid and tRNA [1]. The crystallographic structures and primary sequences for both the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*) [2] and methionyl-tRNA synthetase (*Escherichia coli*) [3,4] are now available, and although there is significant structural homology between these two enzymes, only a short stretch of amino acid sequence is conserved [5].

Binding sites for tyrosine and ATP have been located within the tyrosyl-tRNA synthetase crystallographic structure [6,7] and several lysine residues identified which may interact with the tRNA [8]. In an attempt to further delineate the essential regions of the *B. stearothermophilus* tyrosyl-tRNA synthetase structure we undertook the task of determining the primary sequence of the homologous enzyme from *E. coli*. To this end, we present below the full nucleotide sequence of the *E. coli* tyrosyl-tRNA synthetase, supported by corroborative protein sequencing data.

* Address for correspondence: Division of Microbiology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, England

2. MATERIALS AND METHODS

2.1. *DNA sequencing*

The *E. coli* K12 tyrosyl-tRNA synthetase gene, *tyrS* is located within the 2600 basepair insert of the plasmid pBR322/EcoTyrTS [9], and subcloning experiments had shown that the structural component of the gene lies between the two outer *Bam*H1 restriction sites (fig. 1). The sequencing strategy involved first purifying the 1.1×10^3 basepair and 3×10^3 basepair *Bam*H1 fragments by agarose gel electrophoresis [10], and then cloning into the *Bam*H1 site of M13mp7 [11], both

Published by Elsevier Biomedical Press
00145793/82/0000—0000/$2.75

before and after subdigestion with endonuclease *Sau*3A. These fragments were overlapped by cloning random sonicated fragments of the entire plasmid into the *Hin*dII site of M13mp7, after converting to flush ends with *E. coli* DNA polymerase I (Klenow fragment) as in [12]. The cloned inserts were sequenced in single-stranded form by the dideoxy-chain termination procedure in [13]. By 'turning around' several of the M13 clones to sequence the complementary strand [14], we ensured that >98% of the entire sequence was determined on both strands.

2.2. *Protein sequencing*

E. coli tyrosyl-tRNA synthetase was purified to homogeneity from *E. coli* K12 strain EM20031 as in [15]. The protein, after carboxymethylation, was subjected to 3 separate proteolytic digests: trypsin; chymotrypsin; and cyanogen bromide. Peptides were separated by gel-filtration, ion-exchange on DEAE-cellulose and sulphonated polystyrene, high-pressure liquid chromatography, paper-electrophoresis and paper chromatography. Sequencing was carried out using the manual dansyl-Edman method [16], and amides were assigned either by electrophoretic peptide mobilities or mass spectrometric analysis [17].

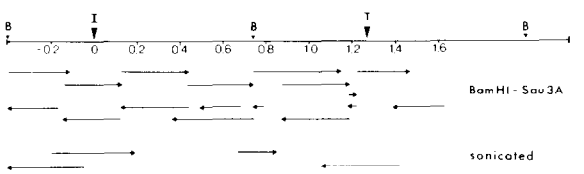

Fig. 1. DNA sequence evidence for part of the 2600 basepair insert of pBR322/EcoTyrTS. The *Bam*H1 restriction sites (B) and translation initiation (I) and termination (T) positions are indicated on the upper line, with distances marked off in 200-base intervals. Horizontal arrows represent the lengths and strand-sense of sequenced DNA fragments, with the relevant cloning strategy indicated on the right hand side.

3. RESULTS AND DISCUSSION

A 1.27×10^3 base open reading frame was identified within the DNA sequence of the insert of pBR322/EcoTyrTS, and is presented in fig. 2 with the amino acid translation. Three lines of protein chemical evidence establish the fact that this sequence does indeed represent the *E. coli* tyrosyl-tRNA synthetase structural gene:

(i) Location of all tyrosyl-tRNA synthetase chymotryptic and tryptic peptides, accounting for 95% of the 423 amino acids. There are no discrepancies between nucleotide and amino acid sequences, despite the fact that gene and protein were isolated from different K12 strains [9].

Table 1

TTT	Phe	F	7	TCT	Ser	S	5	TAT	Tyr	Y	4	TGT	Cys	C	3
TTC	Phe	F	18	TCC	Ser	S	5	TAC	Tyr	Y	7	TGC	Cys	C	4
TTA	Leu	L	4	TCA	Ser	S	0	TAA	Term	O	0	TGA	Term	O	0
TTG	Leu	L	8	TCG	Ser	S	1	TAG	Term	O	0	TGG	Trp	W	6
CTT	Leu	L	1	CCT	Pro	P	3	CAT	His	H	3	CGT	Arg	R	13
CTC	Leu	L	3	CCC	Pro	P	0	CAC	His	H	3	CGC	Arg	R	5
CTA	Leu	L	0	CCA	Pro	P	1	CAA	Gln	Q	5	CGA	Arg	R	1
CTG	Leu	L	28	CCG	Pro	P	9	CAG	Gln	Q	22	CGG	Arg	R	1
ATT	Ile	I	9	ACT	Thr	T	8	AAT	Asn	N	5	AGT	Ser	S	3
ATC	Ile	I	11	ACC	Thr	T	11	AAC	Asn	N	15	AGC	Ser	S	7
ATA	Ile	I	0	ACA	Thr	T	0	AAA	Lys	K	20	AGA	Arg	R	0
ATG	Met	M	6	ACG	Thr	T	2	AAG	Lys	K	8	AGG	Arg	R	0
GTT	Val	V	10	GCT	Ala	A	3	GAT	Asp	D	11	GGT	Gly	G	16
GTC	Val	V	2	GCC	Ala	A	9	GAC	Asp	D	13	GGC	Gly	G	14
GTA	Val	V	4	GCA	Ala	A	9	GAA	Glu	E	22	GGA	Gly	G	1
GTG	Val	V	7	GCG	Ala	A	15	GAG	Glu	E	9	GGG	Gly	G	3

Codon usage for the *Escherichia coli* tyrosyl-tRNA synthetase gene

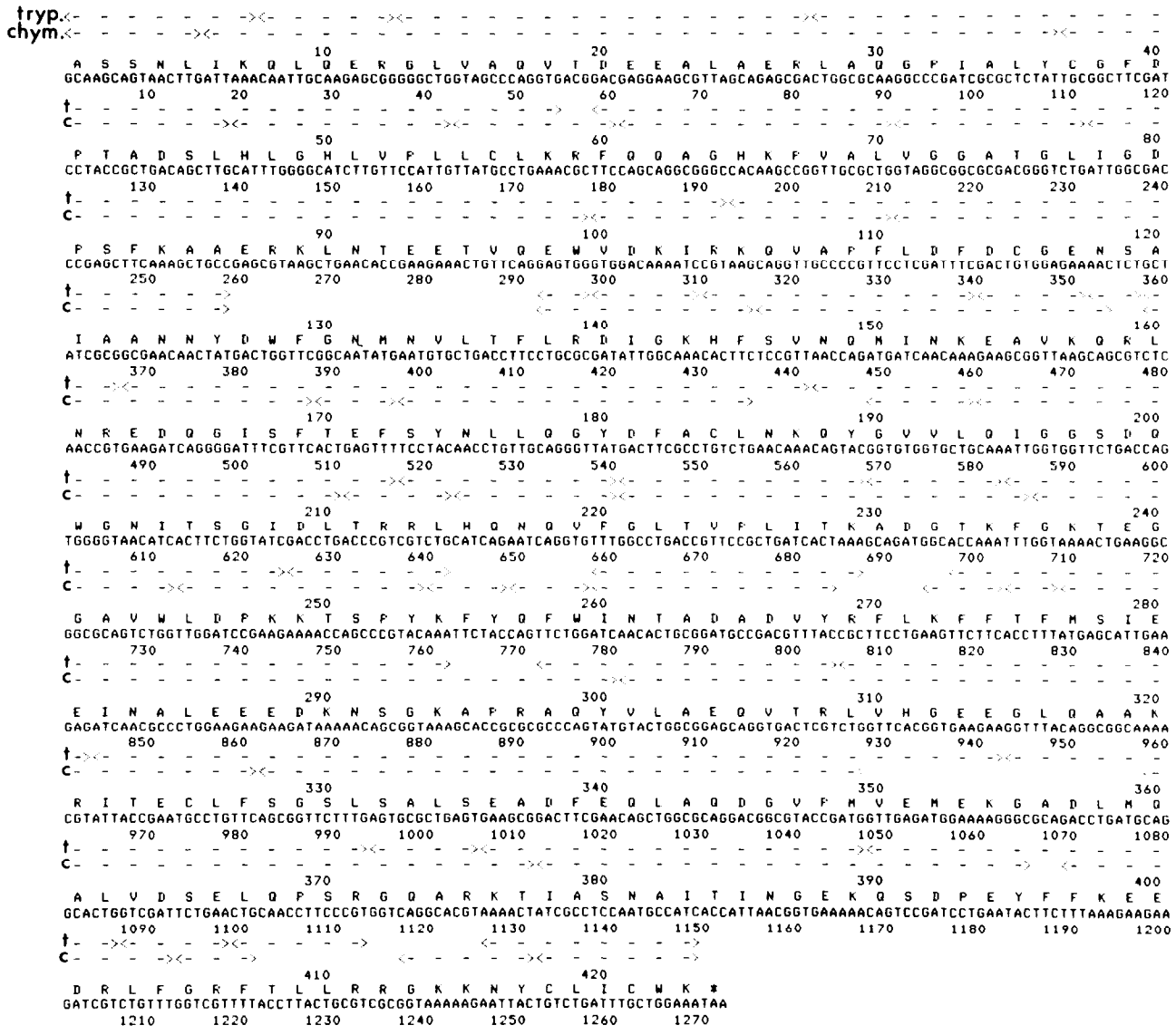

Fig. 2. Complete DNA sequence with amino acid translation (single-letter code, see table 1) for the structural component of the *E. coli* tyrosyl-tRNA synthetase gene. (The ATG initiation codon is not included). Upper (tryptic) and lower (chymotryptic) dashed lines respresent the aligned sequenced peptides, and amino acid and nucleotide distances are indicated above and below the sequence, respectively.

(ii) The N-terminal cyanogen bromide fragment (residues 1–132) has been purified and partially sequenced, allowing us to assign Ala Ser Ser Asn Leu ... as the N-terminal sequence of the protein.

(iii) The M_r of 47403, calculated on the basis of the translated DNA sequence agrees well with estimates of 47500 [18] and 48000 [19].

The pattern of codon usage of the *E. coli* tyrosyl-tRNA synthetase (table 1) shows a strong

preference for ATPy (Ile), 100%, CGPy(Arg), 90%; ACPy (Thr), 90%; GGPy(Gly), 88% and a weaker preference for TTC (Phe), 72%; GAA(Glu), 71% and CCG(Pro), 69%. This broadly resembles the distribution found for all of the *E. coli* aminoacyl-tRNA synthetase genes reported [20–22], and is typical of the category of moderately to highly expressed genes [23]. Although this may explain why the majority of aminoacyl-tRNA synthetase genes can be expressed at very high levels in vivo when cloned in multicopy plasmids [4,9,24], the purpose of such a codon distribution for enzymes which are normally expressed to only a few copies/cell [25] is far from clear. Optimising the translation efficiency of a particular mRNA may also result in a minimization of the error frequency [26]. Since the aminoacyl-tRNA synthetases themselves play a crucial role in the fidelity of protein synthesis, missense or frameshift errors introduced during the translation of their mRNAs could initiate the type of error-catastrophe originally envisaged in [27].

A sequence comparison of the *E. coli* tyrosyl-tRNA synthetase with the *B. stearothermophilus* enzyme [28] reveals considerable homology, similar to that already reported for the sequences of the analagous two tryptophanyl-tRNA synthetases [21,29]. In [5] we discussed the significance of a constellation of cysteine and histidine residues which forms part of the nucleotide binding domain of the *B. stearothermophilus* tyrosyl-tRNA synthetase and is apparently conserved in the *E. coli* methionyl-tRNA synthetase. It is surely no coincidence that these residues are contained within a 17 amino acid long stretch of sequence which is 95% homologous between the two tyrosyl-tRNA synthetases (*E. coli*, residues 35–51; *B. stearothermophilus*, residues 33–49). However, a detailed comparison of the 2 primary structures is best reserved for our forthcoming paper on the complete sequence of the tyrosyl-tRNA synthetase from *B. Stearothermophilus*.

ACKNOWLEDGEMENTS

We wish to thank Professors Brian Hartley and Jean-Pierre Ebel for their encouragement throughout this project. Our gratitude also goes to Dr F. Sanger for allowing D.G.B. to spend 2 months in his laboratory in order to learn the techniques of DNA sequencing and to Dr Ann Dell for the mass spectrometric analysis of various tryptic and chymotryptic peptides. This work was supported by SRC grant number GR/A33354 and an EMBO long-term fellowship held by D.G.B.

REFERENCES

[1] Schimmel, P.R. and Söll, D. (1979) Annu. Rev. Biochem. 48, 601–648.
[2] Bhat, T.N., Blow, D.M., Brick, P. and Nyborg, J. (1982) J. Mol. Biol. 158, 699–709.
[3] Zelwer, C., Risler, J.L. and Brunie, S. (1982). J. Mol. Biol. 155, 63–81.
[4] Barker, D.G., Ebel, J.-P., Jakes, R. and Bruton, C.J. (1982) Eur. J. Biochem. 127, 449–457.
[5] Barker, D.G., and Winter, G. (1982) FEBS Lett. 145, 191–193.
[6] Monteilhet, C. and Blow, D.M. (1978) J. Mol. Biol. 122, 407–417.
[7] Rubin, J. and Blow, D.M. (1981) J. Mol. Biol. 145, 489–500.
[8] Bosshard, H.R., Koch, G.L.E. and Hartley, B.S. (1975) Eur. J. Biochem. 53, 493–498.
[9] Barker, D.G. (1982) Eur. J. Biochem. 125, 357–360.
[10] Polsky, F.I., Edgell, M., Seidman, J.G. and Leder, P. (1978) Anal. Biochem. 87, 397–410.
[11] Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309–321.
[12] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 14, 161–178.
[13] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[14] Winter, G.P. and Fields, S. (1980) Nucleic Acids Res. 8, 1965–1974.
[15] Bruton, C.J., Jakes, R. and Atkinson, A. (1975) Eur. J. Biochem. 59, 327–333.
[16] Gray, W.R. and Hartley, B.S. (1963) Biochem. J. 89, 379–380.
[17] Morris, H.R., Williams, D.H., Midwinter, G.G. and Hartley, B.S. (1974) Biochem. J. 141, 701–713.
[18] Calendar, R. and Berg, P. (1966) Biochemistry 5, 1681–1686.
[19] Chousterman, S. and Chapeville, G. (1973) Eur. J. Biochem. 35, 51–56.
[20] Walter, P., Gangloff, J., Bonnet, J., Boulanger, Y., Ebel, J.-P. and Fasiolo, F. (1983) submitted.
[21] Hall, C.V., Van Cleemput, M., Muench, K.H. and Yanofsky, C. (1982) J. Biol. Chem. 257, 6132–6136.

[22] Putney, S., Royal, N., Neuman de Vegvar, H., Herlihy, W., Biemann, K. and Schimmel, P. (1981) Science 213, 1497–1501.
[23] Ikemura, T. (1981) J. Mol. Biol. 151, 389–409.
[24] Plumbridge, J.A., Springer, M., Graffe, M., Goursot, R. and Grunberg-Manago, M. (1980) Gene 11,33–42.
[25] Neidhardt, F.C., Bloch, P.L., Pederson, S. and Reek, S. (1977) J. Bacteriol. 129, 378–387.
[26] Chavancy, G. and Garel, J.P. (1981) Biochimie 63, 187–195.
[27] Orgel, L.E. (1963) Proc. Natl. Acad. Sci. USA 49, 517–521.
[28] Winter, G.P., Koch, G.L.E., Dell, A. and Hartley, B.S. (1979) in: Transfer RNA: Structure, Properties and Recognition (Schimmel, P.R. and Abelson, J.N. eds) pp. 255–265, Cold Spring Harbor Laboratory, New York.
[29] Winter, G.P. and Hartley, B.S. (1977) FEBS Lett. 82, 348–350.