The complete amino acid sequence of *Escherichia coli* 5-enolpyruvylshikimate 3-phosphate synthase

Kenneth Duncan, Ann Lewendon and John R. Coggins*

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, Scotland

Received 12 March 1984

The amino acid sequence of the enzyme 5-enolpyruvylshikimate 3-phosphate synthase (EPSP synthase) from *Escherichia coli* has been determined using a combined strategy involving amino acid and nucleotide sequencing techniques. The complete polypeptide chain consists of 427 amino acids and has a calculated M_r of 46 112.

5-Enolpyruvylshikimate 3-phosphate synthase *3-Phosphoshikimate 1-carboxyvinyltransferase*
aroA *gene* Escherichia coli *Shikimate pathway* *Glyphosate*

1. INTRODUCTION

The mechanism of the shikimate pathway enzyme EPSP synthase (EC 2.5.1.19, alternative name 3-phosphoshikimate 1-carboxyvinyltransferase) is being actively studied in a number of laboratories [1–3]. Much of the current interest follows the discovery that glyphosate (*N*-phosphonomethylglycine, 'Roundup'), a successful, broad-spectrum, post emergence herbicide, acts by inhibiting this enzyme [2,4–8]. Until recently, detailed mechanistic studies of EPSP synthase have been hindered by a lack of adequate quantities of purified enzyme.

This situation has now been remedied by the development of a simple purification procedure which gives homogeneous preparations of both the *Escherichia coli* [9] and pea seedling [10] enzymes and by the application of this procedure to the purification of the *E. coli* enzyme from an overproducing strain [11]. The availability of milligram quantities of pure EPSP synthase and of the cloned gene (*aroA*) [11] have now allowed us to establish its complete amino acid sequence.

* To whom correspondence should be addressed

Abbreviations: EPSP, 5-enolpyruvylshikimate 3-phosphate; [α-^{35}S]dATPαS, deoxyadenosine 5′-α-[^{35}S]thiotriphosphate

2. MATERIALS AND METHODS

2.1. *Reagents*

Restriction enzymes were purchased from Bethesda Research Laboratories (Cambridge). All the reagents for M13 cloning and sequencing were purchased in the form of kits from Amersham plc (Amersham). [α-^{35}S]dATPαS was also obtained from Amersham plc. Bacteriological reagents were from Difco (Detroit). 2 × TY medium contained 16 g bactotryptone, 10 g yeast extract and 5 g NaCl per litre.

2.2. *Purification of EPSP synthase*

EPSP synthase was purified from *E. coli* strain AB2829/pKD501 as in [11].

2.3. *Automatic amino acid sequence determination*

This was carried out as in [12] using a Beckman Model 890 liquid phase sequencer. The phenylthiohydantoin samples were analysed by chromatography on a Waters Resolve C_{18} reverse phase column with a pH 5.0 acetate–acetonitrile buffer system as in [13]. Amino acid analysis after back hydrolysis

Published by Elsevier Science Publishers B.V.
00145793/84/$3.00

with HI was used to confirm any doubtful residues [14].

2.4. *Amino acid analysis*

Samples of performic acid-oxidised EPSP synthase [15] were hydrolysed and analysed on an LKB Model 4400 amino acid analyser as in [16].

2.5. *Preparation of DNA for sequence analysis*

The *Cla*I–*Pvu*II insert in pKD506 was isolated by restriction endonuclease digestion followed by electrophoresis in low-melting-point agarose [17]. This was then digested (without further purification) with the enzymes *Tag*I and *Hpa*II. Electrophoresis on 2% agarose revealed that the fragment had been broken down into at least 6 *Tag*I and 9 *Hpa*II fragments. An aliquot of each digestion mix containing approximately 100 ng of fragments was purified by phenol and chloroform extraction and ethanol precipitation. After recovery the fragments were ligated to *Acc*I-cleaved M13mp8 [18] and transformed into *E. coli* strain JM101 [14]. A 0.8 kb *Hin*cII–*Pvu*II fragment of pKD506 was isolated in a similar way from a low-melting-point-agarose gel, ligated to *Sma*I cleaved M13mp8 and transformed into JM101. After overnight incubation, clear M13 plaques were picked and grown at 37°C in 1.5 ml of fresh 2 × TY medium containing 0.1 ml/10 ml of an overnight culture of JM101 for 5 h. Single strand recombinant M13 DNA was purified as in [19].

2.6. *DNA sequencing*

The Sanger di-deoxy chain termination method was used [20]. The conditions for annealing and the sequencing reactions were carried out as described in the Amersham M13 Cloning and Sequencing Handbook [19]. [α-^{35}S]dATPαS was used as label. Electrophoresis was on 6% polyacrylamide gels (20 × 40 × 0.04 cm); both linear and buffer gradient gels were used [19]. Gels were dried on Whatman 3MM paper using a Bio-Rad Model 1125 gel dryer before overnight autoradiography.

3. RESULTS AND DISCUSSION

The *aroA* gene has been previously located on a 4.6 kb *Pst*I fragment subcloned from the λ-transducing phage λ-p*serC* [11]. Further subcloning of λ-p*serC* has narrowed down the location of the *aroA* gene to a 1.9 kb *Cla*I–*Pvu*II fragment (K. Duncan and J.R. Coggins, unpublished). This fragment has been cloned into pBR322 [21] and the resulting recombinant plasmid (pKD506) is able to transform the *aroA*$^-$ *E. coli* strain AB2829 so that it will grow on minimal medium [11]. The sequencing work was done on this *Cla*I–*Pvu*II fragment.

The first step in our sequencing strategy was to establish the amino terminal sequence of EPSP synthase using the liquid phase sequencer. The resulting sequence is shown in fig.1. The identification of the first 12 amino acid residues was unambiguous. Residue 13 could not be identified directly as the phenylthiohydantoin. However, back hydrolysis with HI followed by amino acid analysis showed that aspartic acid was the most abundant amino acid present. The sequence then continued unambiguously as far as residue 25 which could not be identified. Residue 33 gave alanine and valine in ratio 2:1 and was tentatively identified as alanine. Residues 39 and 42 could not be identified

```
1                                                    10
Met - Glu - Ser - Leu - Thr - Leu - Gln - Pro - Ile - Ala -

11                                                   20
Arg - Val - (-) - Gly - Thr - Ile - Asn - Leu - Pro - Gly -
          (Asp)
21                                                   30
Ser - Lys - Ser - Val - (-) - Asn - His - Ala - Leu - Leu -

31                                                   40
Leu - Ala - Ala - Leu - Ala - His - Gly - Val - (-) - Val -
          (Val)
41              44
Leu - (-) - Asn - Leu -
```

Fig.1. The N-terminal amino acid sequence of *E. coli* EPSP synthase. The sequence was determined on a liquid phase sequencer as described in section 2. The initial amount of protein sequencing was 19 nmol and the repetitive yield from residue 1–44, by least square regression analysis, was 93% (correlation coefficient 0.95). Residues 11, 27 and 36 were identified only after back hydrolysis with HI followed by amino acid analysis. The gaps and the aspartic acid and valine detected at positions 13 and 33 are discussed in the text.

but the sequence could then be followed up to residue Leu-44. The open reading frame encoding the *aroA* gene could now be identified by comparison of this N-terminal amino acid sequence with sequences downstream of methionine codons in the DNA sequence.

The overall strategy for the nucleotide sequence analysis is shown in fig.2. *Tag*I and *Hpa*II clones were sequenced and the sequences entered as separate files on a Digital PDP11-34 computer. The sequences were linked up using Staden's programmes [22] until both strands of the DNA were complete and all the restriction sites used in the cloning had been overlapped (see fig.2). The open reading frame was found to run past the *Pvu*II site used to construct pKD506 (fig.2). In order to sequence beyond this site, a *Hae*II fragment of pKD502 [11] was isolated. This DNA fragment extends from position 1228 in fig.3 to the *Hae*III site at position 3757 in the sequence of pBR322 [21]. The fragment was cloned into *Sma*I-cleaved M13mp8 and the sequence of both stands determined.

The complete nucleotide sequence of the *aroA* gene together with the deduced amino acid sequence for EPSP synthase is shown in fig.3. In the N-terminal region there is exact agreement between the experimentally determined amino acid sequence and the amino acid sequence deduced from the DNA sequence. The predicted amino acid sequence corresponds to a 427 amino acid polypeptide chain of calculated M_r 46112. This is in good agreement with the subunit M_r value of 49000 determined by SDS-polyacrylamide gel electrophoresis [9] and with the native M_r of 42000 determined by gel permeation chromatography on a TSK G2000SW column [11]. A comparison of the experimentally determined amino acid composition with the amino acid composition predicted from the nucleotide sequence is shown in table 1. The overall agreement is very good. Thus the protein chemical results are entirely consistent with the predicted amino acid sequence and there can be no doubt that the sequenced gene is *aroA*.

Now that the complete amino acid sequence of EPSP synthase is known it will be possible to use

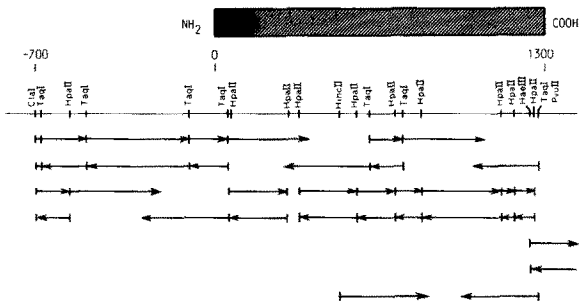


Fig.2. Overall strategy for the sequence analysis of the *aroA* gene. The sequence corresponding to the coding region is indicated by the box; the solid part of the box corresponds to the N-terminal amino acid sequence. Only the positions of the relevant restriction sites are indicated; the nucleotide numbers correspond to those given in fig.3. The horizontal arrows indicate the direction and extent of sequence determination.

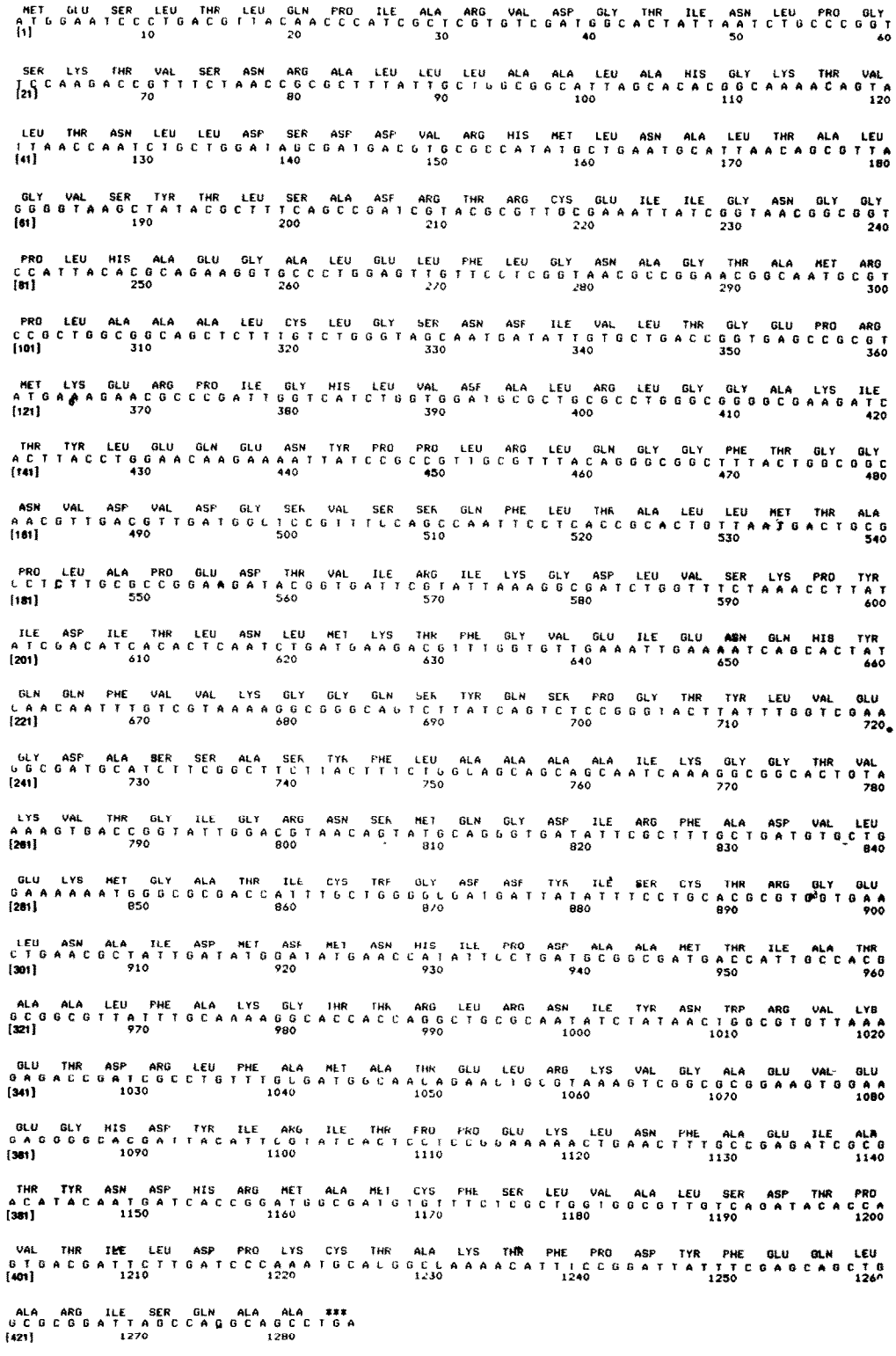
Table 1

The amino acid composition of *E. coli* EPSP synthase compared with the amino acid composition deduced for EPSP synthase from the *E. coli aroA* gene sequence

Amino acid	Relative amino acid composition based on Leu = 48 residues	Theoretical amino acid composition predicted from the DNA sequence
Asp	41.9	44
Thr[a]	31.1	34
Ser[a]	19.7	21
Glu	38.8	34
Pro	18.1	18
Gly	42.8	37
Ala	44.1	46
Cys[b]	4.9	6
Val	21.7	24
Met[c]	13.6	14
Ile	24.2	26
Leu	48.0	48
Tyr	13.1	13
Phe	13.2	13
His	8.1	8
Lys	17.0	17
Arg	17.2	22
Trp	nd	2

[a] Experimental values were extrapolated to zero time
[b] Determined as cysteic acid
[c] Determined as methionine sulphone

Samples were analysed in duplicate after hydrolysis of performic acid oxidised-protein with 6 M HCl at 105°C for 24, 48, 72 and 96 h. The 8 experimental values were simply averaged except where indicated in the footnotes



←

Fig.3. The complete nucleotide sequence of the *E. coli aroA* gene and the corresponding amino acid sequence of *E. coli* EPSP synthase. Nucleotides are numbered in the 5′ to 3′ direction beginning with the first residue of the ATG triplet encoding the N-terminal methionine. The bracketed numbers refer to the amino acid positions in the sequence.

a combination of protein chemical and genetic techniques to locate the active site and learn more about the enzyme's mechanism and in particular how it is inhibited by glyphosate [2–5]. In this respect it is of interest to note that mutant bacterial strains [7,8] and a mutant plant cell line [7] have been reported to be resistant to glyphosate. In one case this resistance was claimed to be due to a modification of the enzyme that diminished its affinity for glyphosate but did not significantly effect its catalytic properties [8]. If this can be confirmed then it may be possible to engineer strains of crop plants that are glyphosate-resistant by the insertion and expression of a modified EPSP synthase gene coding for a glyphosate-insensitive form of the enzyme.

ACKNOWLEDGEMENTS

We would like to thank Bryan Dunbar, John Fothergill and Linda Fothergill for assistance with the amino acid sequencing which was carried out in the SERC funded protein sequencing facility at Aberdeen University. We also thank Jim Jardine, Glasgow University, for carrying out the amino acid analyses. This work was supported by grants to J.R.C. from the Science and Engineering Research Council.

REFERENCES

[1] Grimshaw, C.E., Sogo, S.G. and Knowles, J.R. (1982) J. Biol. Chem. 257, 596–598.
[2] Boocock, M.R. and Coggins, J.R. (1983) FEBS Lett. 154, 127–133.
[3] Anton, D.L., Hedstrom, L., Fish, S.M. and Abeles, R.H. (1983) Biochemistry 22, 5903–5908.
[4] Amrhein, N., Schab, J. and Steinrucken, H.C. (1980) Nature 67, 356–357.
[5] Steinrucken, H.C. and Amrhein, N. (1980) Biochem. Biophys. Res. Commun. 94, 1207–1212.
[6] Rogers, S.G., Brand, L.A., Holder, S.B., Sharps, E.S. and Brackin, M.J. (1983) Appl. Environ. Microbiol. 46, 37–43.
[7] Amrhein, N., Johänning, D., Schab, J. and Schulz, A. (1983) FEBS Lett. 157, 191–196.
[8] Camai, L., Sen, L.C. and Stalker, D.M. (1983) Science 221, 370–371.
[9] Lewendon, A. and Coggins, J.R. (1983) Biochem. J. 213, 187–191.
[10] Mousdale, D.M. and Coggins, J.R. (1984) Planta 160, 78–83.
[11] Duncan, K., Lewendon, A. and Coggins, J.R. (1984) FEBS Lett. 165, 121–127.
[12] Smith, M.A., Gerrie, L.M., Dunbar, B. and Fothergill, J.E. (1982) Biochem. J. 207, 253–260.
[13] Carter, P.E., Dunbar, B. and Fothergill, J.E. (1983) Biochem. J. 215, 565–571.
[14] Smithies, O., Gibson, D., Fanning, E.M., Goodfliesh, R.M., Gilman, J.G. and Ballantyne, D.L. (1971) Biochemistry 10, 4912–4921.
[15] Hirs, C.H.W. (1967) Methods Enzymol. 11, 197–199.
[16] Lumsden, J. and Coggins, J.R. (1978) Biochem. J. 169, 441–444.
[17] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor, New York.
[18] Messing, J. and Vieira, J. (1982) Gene 19, 269–276.
[19] M13 Cloning and Sequencing Handbook (1983) Amersham plc, Amersham, England.
[20] Sanger, F. (1981) Science 214, 1205–1210.
[21] Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 77–90.
[22] Staden, R. (1980) Nucleic Acids Res. 8, 3673–3694.