The overexpression, purification and complete amino acid sequence of chorismate synthase from *Escherichia coli* K12 and its comparison with the enzyme from *Neurospora crassa*

Peter J. WHITE, Gary MILLAR and John R. COGGINS

Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K.

The enzyme chorismate synthase was purified in milligram quantities from an overproducing strain of *Escherichia coli*. The amino acid sequence was deduced from the nucleotide sequence of the *aroC* gene and confirmed by determining the *N*-terminal amino acid sequence of the purified enzyme. The complete polypeptide chain consists of 357 amino acid residues and has a calculated subunit M_r of 38183. Cross-linking and gel-filtration experiments show that the enzyme is tetrameric. An improved purification of chorismate synthase from *Neurospora crassa* is also described. Cross-linking and gel-filtration experiments on the *N. crassa* enzyme show that it is also tetrameric with a subunit M_r of 50000. It is proposed that the subunits of the *N. crassa* enzyme are larger because they contain a diaphorase domain that is absent from the *E. coli* enzyme.

INTRODUCTION

Chorismate synthase (5-enolpyruvylshikimate 3-phosphate phosphate-lyase; EC 4.6.1.4), the seventh and final enzyme in the shikimate pathway, catalyses the conversion of 5-enolpyruvylshikimate 3-phosphate into chorismate, the key branch-point intermediate in aromatic biosynthesis (Haslam, 1974; Weiss & Edwards, 1980). This enzyme, which to date is the least studied of all the enzymes of the shikimate pathway, has two unusual properties. Firstly, it has a reaction mechanism that is stereochemically ambiguous (Floss *et al.*, 1972; Ganem, 1978), and, secondly, it requires a reduced flavin cofactor although the reaction results in no net overall change in redox state.

The enzyme has been characterized from three microbial sources [*Escherichia coli* (Morell *et al.*, 1967), *Neurospora crassa* (Welch *et al.*, 1974; Boocock, 1983) and *Bacillus subtilis* (Hasan & Nester, 1978)] and more recently from one plant source [*Pisum sativum* (Mousdale & Coggins, 1986)]. The enzymes from these sources exhibit a marked dissimilarity in their ability to generate the reduced flavin cofactor necessary for catalysis. Both the *N. crassa* and *B. subtilis* chorismate synthases appear to be associated with specific flavin reductase (diaphorase) activities that can generate the essential reduced cofactor via the oxidation of nicotinamide nucleotides under aerobic conditions. In contrast, the partially purified *E. coli* and pea enzymes could only be assayed under strictly anaerobic conditions in the presence of chemically or enzymically reduced flavin (Morell *et al.*, 1967; Mousdale & Coggins, 1986; White *et al.*, 1987).

In an attempt to understand the mechanism of action of chorismate synthase the *E. coli* gene (*aroC*) that encodes chorismate synthase has been cloned and overexpressed (Millar *et al.*, 1986b). The present paper reports the purification of the overexpressed *E. coli* enzyme, the sequence of the *aroC* gene and the *N*-terminal sequence analysis and quaternary structure of

the purified *E. coli* enzyme. A new purification procedure for the *N. crassa* enzyme is also presented, and this has allowed comparison of these two enzymes at the levels of primary and quaternary structure.

MATERIALS AND METHODS

Reagents

All the reagents except those specified below were obtained from Boehringer Corp., Lewes, East Sussex, U.K., Sigma Chemical Co., Poole, Dorset, U.K., or BDH Chemicals, Poole, Dorset, U.K. DEAE-Sephacel, Sephacryl S-300 and phenyl-Sepharose CL-4B were obtained from Pharmacia, Milton Keynes, Bucks., U.K.; cellulose phosphate (P11) and DEAE-cellulose (DE52) were from Whatman Biochemicals, Maidstone, Kent, U.K.; clostripain was from Worthington Biochemical Corp., Freehold, NJ, U.S.A. The barium salt of 5-enolpyruvylshikimate 3-phosphate, prepared as described in Coggins *et al.* (1987), was a gift from Mrs. S. Muir, and dimethyl suberimidate dihydrochloride was prepared as described by Davies & Stark (1970).

Subcloning

All restriction enzymes were purchased from Bethesda Research Laboratories (Paisley, Scotland, U.K.) and were used as recommended by the suppliers. Bacteriophage-T4 DNA ligase was obtained from Amersham International (Amersham, Bucks., U.K.). DNA fragments were isolated by using low-melting-temperature agarose (Bethesda Research Laboratories) as described in Maniatis *et al.* (1982).

Cloning in M13

Extensive use was made of the Messing series of paired M13 vectors (Messing & Vieira, 1982) to clone regions of the 1.69 kbp genomic insert of plasmid pGM602 (see Fig. 2). All manipulations were as previously described (Millar *et al.*, 1986a).

DNA sequence analysis

All DNA sequencing employed the dideoxy chain-termination method (Sanger *et al.*, 1977; Biggin *et al.*, 1983). The conditions for annealing and for sequencing reactions were as described previously (Duncan *et al.*, 1984). [α - 35 S]Thio]dATP obtained from Amersham International was used as the incorporated label. Electrophoresis was in 8 M-urea/6% (w/v) polyacrylamide linear- or buffer-gradient gels (0.4 cm $\times$ 40 cm $\times$ 20 cm).

Each sequenced M13 clone was entered as a single file into a Digital PDP11-34 computer. The sequences were linked up by using Staden's (1980) programs.

Coupled transcription/translation *in vitro*

Bacterial cell-free coupled transcription/translation system was obtained from Amersham International (Amersham no. N380) and used as recommended by the supplier. [35 S]Methionine-labelled proteins expressed from the plasmid DNA by using this system were examined by polyacrylamide-gel electrophoresis in the presence of SDS and subsequent autoradiography (-70°C).

Assay of chorismate synthase activity

E. coli chorismate synthase was assayed anaerobically, by measuring the appearance of chorismate in a stopped-time method as described previously (White *et al.*, 1987). *N. crassa* chorismate synthase was assayed by monitoring the formation of chorismate (ϵ_{275} 263 M $^{-1}$ ·cm $^{-1}$; Gibson, 1970) spectrophotometrically at 275 nm in a total volume of 1 ml. The assay mixture contained (final concentrations) 50 mM-triethanolamine hydrochloride/KOH buffer, pH 7.0, 50 mM-KCl, 20 μM -NADPH, 10 μM -FMN and 50 μM -5-enolpyruvylshikimate 3-phosphate.

One unit of enzyme activity is defined as the amount of enzyme catalysing the conversion of 1 μmol of substrate/min.

Purification of chorismate synthase from *E. coli*

All steps after the breaking of the cells were performed at 4°C .

Step 1: extraction and centrifugation. A 20 g batch of *E. coli* (strain AB2849/pGM602) was suspended in 10 ml of 50 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol and 1.2 mM-phenylmethanesulphonyl fluoride (buffer A) and broken by two passages through a French pressure cell. The extract was diluted to 80 ml with the above buffer, and 0.5 mg of deoxyribonuclease I was added. After 1 h stirring the resulting suspension was centrifuged at 28000 g for 30 min. Chorismate synthase was purified from the resulting supernatant.

Step 2: fractionation with $(\text{NH}_4)_2\text{SO}_4$. The crude extract was made 1 M with respect to benzamidine and adjusted carefully to 35% saturation with solid $(\text{NH}_4)_2\text{SO}_4$ (208 g/l). After stirring for 20 min the precipitate was removed by centrifugation at 28000 g for 30 min. The supernatant was adjusted to 60% saturation (390 g/l) and stirred for 20 min. The precipitated protein was collected by centrifugation at 28000 g, resuspended in buffer A and dialysed overnight against 2 litres of

50 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol, 1.2 mM-phenylmethanesulphonyl fluoride and 50 mM-KCl (buffer B).

Step 3: DEAE-Sephacel chromatography. The dialysed material was loaded on to a column of DEAE-Sephacel (bed volume 60 ml) equilibrated in buffer B. The column was washed with buffer until the A_{280} of the eluate had fallen to a value below 0.3. The column was then eluted with a 400 ml linear gradient of 50–250 mM-KCl in 50 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol (flow rate 70 ml/h). Fractions (5 ml) containing chorismate synthase activity were pooled and dialysed overnight against 2 $\times$ 2 litres of 10 mM-potassium phosphate buffer, pH 6.6, containing 0.4 mM-dithiothreitol (buffer C).

Step 4: cellulose phosphate (P11) chromatography. The dialysed extract was loaded on to a column (bed volume 25 ml) of cellulose phosphate equilibrated in buffer C and washed with the same buffer until the A_{280} of the eluate had fallen to a value below 0.05. The column was then eluted with a 400 ml linear gradient of 10–400 mM-potassium phosphate buffer, pH 6.6, containing 0.4 mM-dithiothreitol (flow rate 20 ml/h). Fractions (5 ml) containing chorismate synthase activity were pooled and dialysed overnight against 2 $\times$ 1 litre of 50 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol. The protein was concentrated by adsorption on a 1 ml (bed volume) DEAE-Sephacel column and eluted with the above buffer containing 1.0 M-KCl. The concentrated enzyme was then dialysed overnight against 500 ml of 50 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol and 0.8 M- $(\text{NH}_4)_2\text{SO}_4$.

Step 5: phenyl-Sepharose chromatography. The dialysed material was loaded on to a column of phenyl-Sepharose (bed volume 1 ml) equilibrated in 50 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol and 0.8 M- $(\text{NH}_4)_2\text{SO}_4$. Chorismate synthase activity was eluted with a decreasing linear gradient of $(\text{NH}_4)_2\text{SO}_4$ (0.8–0.0 M) in the above buffer (flow rate 10 ml/h). Fractions (1.5 ml) containing chorismate synthase activity were pooled and dialysed against 1 litre of 50 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol. The enzyme was concentrated by adsorption on a 1 ml (bed volume) DEAE-Sephacel column as before. The concentrated protein was dialysed against 50 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol and 50% (v/v) glycerol and stored at -20°C .

Purification of chorismate synthase from *N. crassa*

All steps were carried out at 4°C unless otherwise stated.

Step 1: extraction and centrifugation. A 90 g batch of powdered freeze-dried *N. crassa* mycelia was stirred gently into 1500 ml of an extraction buffer, which consisted of 100 mM-potassium phosphate, pH 7.0. This buffer was supplemented with 5 mM-EDTA, 0.4 mM-dithiothreitol and 1.2 mM-phenylmethanesulphonyl fluoride. Stirring was continued for 1 h. The extract was then centrifuged at 10000 g for 30 min. The supernatant was passed through several layers of cheesecloth and was the crude extract from which the enzyme was purified.

Step 2: 'negative' chromatography on DEAE-cellulose. The crude extract was pumped at 400 ml/h through a column (15 cm × 61 cm²) of DE52 DEAE-cellulose pre-equilibrated with 100 mM-potassium phosphate buffer, pH 7.0, containing 0.4 mM-dithiothreitol and 1.2 mM-phenylmethanesulphonyl fluoride. The column was washed with equilibration buffer until all significant chorismate synthase activity was eluted.

Step 3: fractionation with (NH₄)₂SO₄. The material eluted from the DE52 DEAE-cellulose column was made 1 M with respect to benzamidinium and adjusted to 40 % saturation with solid (NH₄)₂SO₄ (242 g/l). After the mixture had been stirred for 20 min the precipitated proteins were removed by centrifugation at 10000 *g* for 30 min. The supernatant was adjusted to 50 % saturation with solid (NH₄)₂SO₄ (63 g/l) and stirred for 20 min. The precipitated protein was collected by centrifugation at 10000 *g* for 30 min and redissolved in 100 ml of 100 mM-potassium phosphate buffer, pH 7.0, containing 0.4 mM-dithiothreitol and 1.2 mM-phenylmethanesulphonyl fluoride. The dissolved protein was then dialysed overnight against 2 × 1 litres of 50 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol and 1.2 mM-phenylmethanesulphonyl fluoride (buffer D).

Step 4: DEAE-Sephacel chromatography. The dialysed protein was spun briefly at 10000 *g* to remove any precipitated proteins and then applied to a DEAE-Sephacel column (12 cm × 5 cm²) that had been pre-equilibrated with buffer D. The column was then washed with buffer D containing 30 mM-KCl (flow rate 200 ml/h) until the *A*₂₈₀ was below 0.15. The protein was eluted with a 1-litre linear gradient of 30–300 mM-KCl in buffer D (flow rate 100 ml/h; 10 ml fractions). The fractions containing the highest enzyme activity were pooled and dialysed against 2 × 2 litres of 10 mM-potassium phosphate buffer, pH 6.5, containing 0.4 mM-dithiothreitol and 1.2 mM-phenylmethanesulphonyl fluoride (buffer E).

Step 5: phosphocellulose chromatography. The dialysed protein from the previous step was applied to a phosphocellulose column (25 cm × 1 cm²) that had been pre-equilibrated with buffer E. The protein was eluted with two 200 ml linear gradients of potassium phosphate in buffer E: 10–100 mM (flow rate 30 ml/h; 10 ml fractions) and 100–400 mM (flow rate 12 ml/h; 5 ml fractions). Chorismate synthase activity was eluted in the second gradient, and fractions of high activity were pooled and dialysed overnight against 1 litre of buffer D. The protein was concentrated by adsorption on a 1 ml (bed volume) DEAE-Sephacel column and eluted with buffer D containing 1 M-KCl. The concentrated enzyme was dialysed against 1 litre of 20 mM-Bis-Tris/HCl buffer, pH 6.0, containing 0.4 mM-dithiothreitol and 1.2 mM-phenylmethanesulphonyl fluoride (buffer F) before chromatography on Mono Q.

Step 6: chromatography on Mono Q. This step was carried out at room temperature with a Pharmacia fast protein liquid chromatography system. The enzyme from the previous step was applied to a Mono Q anion-exchange column and eluted with a linear gradient of 0–300 mM-KCl in buffer F (flow rate 1 ml/min, 0.5 ml fractions). The fractions containing the highest choris-

mate synthase activity were pooled and dialysed overnight against 500 ml of 50 mM-Tris/HCl buffer, pH 7.5, containing 50 % (v/v) glycerol and 0.4 mM-dithiothreitol and then stored at –20 °C.

Protein determination

Protein was determined by the method of Bradford (1976), with bovine serum albumin as standard.

Polyacrylamide-gel electrophoresis

Electrophoresis was performed by the method of Davis (1964) in 5 % polyacrylamide gels at 4 °C. After electrophoresis gels were stained either for protein or for diaphorase activity. Electrophoresis in the presence of SDS was performed by the method of Laemmli (1970), with a 3 % stacking gel and a 10 % running gel, or by the method of Weber & Osborn (1969) in 7.5 % tube gels.

Activity staining

Non-denaturing tube gels were stained for diaphorase activity at 21 °C in a cocktail consisting of 0.25 mM-NADPH, 10 μM-FMN, 60 μM-2,6-dichloroindophenol and 1 mM-Thiazolyl Blue Tetrazolium in 100 mM-Tris/HCl buffer, pH 8.8 (Boocock, 1983).

Cross-linking with dimethyl suberimidate

This was carried out in 0.1 M-triethanolamine buffer, pH 8.0, as described previously (Lumsden & Coggins, 1977) except that the final concentrations were 0.3 mg of protein/ml, 20 mM-dimethyl suberimidate and 40 mM-NaCl. After 1 h the cross-linking reaction was terminated by either the addition of 1 M-NH₄HCO₃ to a final concentration of 0.5 M or dialysis against 0.05 M-sodium phosphate buffer, pH 7.2. The cross-linked proteins were then analysed by electrophoresis in 0.05 M-sodium phosphate buffer, pH 7.2, in the presence of SDS (Weber & Osborn, 1969).

Determination of *M_r* by gel filtration

The native *M_r* values of both enzymes were estimated by gel-permeation chromatography at 21 °C on a Superose 6 column in a Pharmacia fast protein liquid chromatography apparatus. The column was eluted with 50 mM-Tris/HCl buffer, pH 7.5, containing 150 mM-KCl and 0.4 mM-dithiothreitol (flow rate 0.5 ml/min, fraction size 0.25 ml). The eluate was monitored at 280 nm, and the column was calibrated with the following proteins: yeast hexokinase (*M_r* 104000); pig muscle lactate dehydrogenase (*M_r* 144000); rabbit muscle pyruvate kinase (*M_r* 237000); horse apoferritin (*M_r* 440000).

Automatic amino acid sequence determination

A 30 nmol sample of *E. coli* chorismate synthase, which had been reduced and carboxymethylated by the method of Lumsden & Coggins (1978), was sequenced on a Beckman model 890 liquid-phase sequencer as described previously (Smith *et al.*, 1982). The phenylthiohydantoin samples were analysed by chromatography on a Waters Resolve C₁₈ reverse-phase column with a pH 5.0 acetate/acetonitrile buffer system (Carter *et al.*, 1983). The initial amount for protein sequencing was 16 nmol and the repetitive yield from residues 1–30 was 93.61 % by regression analysis.

The peptides isolated from *N. crassa* chorismate synthase were sequenced on an Applied Biosystems

Fig. 1. DNA sequencing strategy for the *E. coli aroC* gene

The location of the *aroC* gene within the 1.69 kbp SalI–ClaI insert of pGM602 (Millar *et al.*, 1986b) is represented by hatching. The deletions (pGM603 and pGM604) employed to map the *aroC* gene within pGM602 are shown at the top of the Figure. The length and orientation of M13 sub-clones used in the DNA sequence analysis are represented by arrows and the relevant restriction sites are indicated. The position of a potential terminator structure (Ψ) is indicated (see the text and Fig. 6).

model 4708 gas phase sequencer as described by Russell *et al.* (1986).

Preparation of peptides from *N. crassa* chorismate synthase

A 6 nmol sample of *N. crassa* chorismate synthase, which had been reduced and carboxymethylated by the method of Lumsden & Coggins (1978), was digested with clostripain by following the procedure of Mitchell & Harrington (1971). Clostripain was first activated by preincubation in 1 mM-calcium acetate/2 mM-dithiothreitol overnight at 4 °C. The digestion was carried out in 75 mM-sodium phosphate buffer, pH 7.6, containing 7.5 mM-dithiothreitol at a final protein concentration of 5 mg/ml. The previously activated clostripain was added to give a protein/clostripain ratio of 50:1 (w/w). Digestion was allowed to proceed for 1 h at 21 °C and for a further 3.5 h at 37 °C. The reaction was stopped by freezing at –20 °C. The resulting peptides were separated by reverse-phase chromatography on a Waters μ Bondapak C₁₈ column.

RESULTS

Previous work from this laboratory (Millar *et al.*, 1986b) had indicated that the plasmid pLC33-1 (Clarke & Carbon, 1976) carried the region of the *E. coli* chromosome at or near minute 50 where the *aroC* gene had been mapped (Bachmann, 1983; Hagervall & Björk, 1984). Furthermore, a 1.69 kbp SalI–ClaI sub-clone of pLC33-1, cloned in pAT153 (Twigg & Sherratt, 1980), had also been shown to retain the ability to complement the *E. coli aroC*⁻ mutant AB2849 (Pittard & Wallace, 1966) strain and to overexpress chorismate synthase activity (Millar *et al.*, 1986b).

Further restriction mapping and deletion analysis of this plasmid (designated pGM602; Millar *et al.*, 1986b) (see Fig. 1) identified EcoRV and NruI sites within the genomic insert (Fig. 1). Removal of the region of pGM602 bounded by the SalI and NruI restriction sites

had no effect on the resultant plasmid's (pGM603; Fig. 1) ability to complement the auxotrophic requirements of *E. coli* AB2849. However, deletion of the region bounded by the SalI and EcoRV sites (pGM604; Fig. 1) abolished the complementation of *E. coli* AB2849.

Coupled transcription/translation *in vitro* of plasmids pGM602 and pAT153 established that the insert carried by pGM602 was expressing a protein of an apparent M_r of 38000–40000 (Fig. 2). This observation was consistent with previous data (Millar *et al.*, 1986b), which had identified an overexpressed polypeptide of M_r 40000 in crude extracts prepared from *E. coli* AB2849/pGM602.

A mini-library of bacteriophage M13 clones carrying sub-segments of the 1.69 kbp SalI–ClaI insert of pGM602 was constructed (Fig. 1) and the DNA sequence of each was determined. The sequences obtained were aligned until the entire 1.69 kbp (with the exception of 30 bp at the redundant SalI end) had been sequenced on both strands, and all restriction sites had been overlapped (Fig. 1). The resulting sequence is shown in Fig. 3.

Both strands were examined for open reading frames by searching for translational start and stop codons. Only one open reading frame was found that was sufficiently long to encode an enzyme with an M_r of 38000–40000. When translated from the first initiation codon (AUG position 492) this was found to encode a polypeptide chain of 357 amino acid residues with a calculated M_r of 38183. Immediately preceding the putative start codon is the sequence GGAG (position 483–486, Fig. 3), which exhibits reasonable homology with the consensus *E. coli* ribosome-binding-site sequence GGAGG (Shine & Dalgarno, 1975).

The position and orientation of this putative *aroC* coding sequence was consistent with the deletion analysis summarized in Fig. 1. Similarly the size of protein encoded (M_r 38183) showed good agreement with that observed both *in vivo* (Millar *et al.*, 1986b) and *in vitro* from pGM602 (Fig. 2).

The program TESTCODE (Devereux *et al.*, 1984) was

Fig. 2. Transcription/translation *in vitro* of plasmid pGM602 to give the encoded proteins

Incorporation *in vitro* of [³⁵S]methionine into pAT153-(control) and pGM602-encoded proteins was investigated. *E. coli* chorismate synthase (*aroC*) and vector-encoded β -lactamase (*amp^r*) protein products are indicated. SDS/polyacrylamide-gel electrophoresis was carried out on a 12.5% (w/v) gel and autoradiography was at -70 °C with the use of intensifier screens.

applied to both strands of the DNA sequence. TESTCODE searches for statistical order based on the non-random utilization of codons observed in *bona fide* genes and provides an objective means of identifying protein-coding sequences. The output when the TESTCODE is applied to the *aroC* sequence is shown in Fig. 4. Through most of the putative coding sequence the running value of the algorithm predicts greater than 95% probability that the sequence is a genuine protein-coding sequence.

To confirm that this open reading frame encoded chorismate synthase the enzyme was purified from the overproducing strain AB2849/pGM602 (see Table 1). Three chromatographic steps were required to obtain electrophoretically homogeneous enzyme: ion-exchange chromatography on DEAE-Sephacel, anion-exchange chromatography on cellulose phosphate and finally hydrophobic interaction chromatography on phenyl-Sepharose. A polyacrylamide-gel electrophoretogram showing the protein components at each stage of the purification is shown in Fig. 5. The chorismate synthase content in *E. coli* K12 is very low and difficult to quantify and no previous purification has been reported. From the overproducing strain the enzyme still required over 600-fold purification; the final specific activity was 14.8 units/mg.

The location of the correct open reading frame for the

1 GTCCAGCGCGGTGGATATCTCTCCAGACGGCGTGGCGGTTGCTGAACAGCAATCGGAAG
60 ACACGGCTCTGATCCACAAAGCTATTCCGATTGCTTCGCGATCTGTTCCGGGACTGCCGAA
120 AGTCGACTAGGACCTGATTGTCTACTAACCCGGCGCTATGTCGATGCGAAGATATGTCGCGAC
180 CTGCCAAACAATACCGCCACGACGCCGGAATGGGCTGGCATCTGGCACTGACCGGCTGAC
240 AACTGACGGCTGGCATTTCCGGTAACGGCGCAGATTACCTTGCCTGATGATGGCGTGTGA
300 TTGTGAAGTCGGCAACAGCATGTGACTCTTATGGAAACAATACCGGATGTTCCGTTCA
360 CCTGGCTGGAGTTTGATAACGGCGCGGATGCTGTGTTTATGCTCACCAAGAGCAGCTTA
420 TTGCGGACGAGAACATTTCGGGATTTATAAAGATTAGCTAAACACCGCAACACAACAAT
480 AACGGAGCGCTGATGCTGGAAACAACAATTGGACAACCTTTTGGCGGTAAACCACTTCGGC
RBS MetAlaGlyAsnThrIleGlyGlnLeuPheArgValThrThrPheGly [16]
540 GAATCGACGGGCTGGCCCTCGGCTCATCTCGATCGGTGTTCGGGACGAGCATTCGGCTG
GluSerHisGlyGlyGlyCysIleValAspGlyValProProGlyIleProLeu [36]
600 ACCGAACGGCAGCTGCAACATGACCTTCGACCTGCTCCGCTCGGACATCGGCTATACC
ThrGluAlaAspLeuGlnHisAspAspArgArgGlyProThrThrArgGlyThrThr [56]
660 ACCACGGCGCGGACCGGACCGTCAGTCAAAATCTCTCCGGTGTGTTTGAAGCGGCTACT
ThrGlnArgArgGluProAspGlnValValLysLeuSerGlyValPheGluGlyValThr [76]
720 ACCGGCACCAGCATGGCTGTGTGATCGGAAAACATGACACGGGCTCTCAGGATTACAGT
ThrGlyThrSerIleGlyLeuLeuGlnAsnThrAspGlnArgSerGlnAspAspTyrSer [96]
780 CGGATTAAAGGAGTTTTCGCTCAGGCGATCCGAGTACACATACGAAACAAAATACGGT
AlaIleLysAspValPheArgProGlyHisAlaAspTyrThrThrGlyGlnLysTyrGly [116]
840 CTGCGCGATTATCGCGGCGGTGGAGGCTTCTCCGCGCGGAAACCGCATCGCGCTGGG
LeuArgAspTyrArgGlyGlyGlyArgSerAlaArgGluThrAlaMetArgValAla [136]
900 GCAGGAGCTATGCCAAAAAATATCTCGCGGAGAAATTTGGTATTGAAATCCGTGGCTGC
AlaGlyAlaIleAlaLysLysTyrLeuAlaGluLysPheGlyIleGluIleArgGlyCys [156]
960 CTGACCCAGGCGGGCAGATCTCCGCTGGATATCAAGACTGCTGCGAGTGGAGCAAAAT
LeuThrGlnMetGlyAspIleProAspPheLysAspTyrPheGlnValValAlaLeuGlnAsn [176]
1020 CCGTTTTTTTGGCGGACCGCCGACAAAAATCGACGGTTAGACAGTTGATCGCTCGGCTG
ProPhePheCysProAspProAspLysIleAspAlaLeuAspGluLeuMetArgAlaLeu [196]
1080 AAAAAAGCGGGGACTCCATCGGCGCTAAAGTCACCGTTGTTGGCAATCGGGTCTCTGCC
LysLysGlyLysAspSerIleGlyAlaValThrValAlaSerGlyValProAla [216]
1140 GGACTTGGCAGCGGCTGTTCAGCGGCTGGATGCTGACATCGCCCATCGGCTGATGAGC
GlyLeuGlyGluProValPheAspArgLeuAspAlaAspIleAlaHisAlaLeuMetSer [236]
1200 ATCAACGGGCTGAAAGCGGTGAAATTCGGCAGCGGTTTGACGTGGTGGGCTCGCGGCG
IleAsnAlaGlyGlyValGluGlyValGluGlyAlaGlyValThrValAlaLeuArgGly [256]
1260 AGCCAGAACCCGATGAAATCACCAAGACGGTTTCAGAGCAACCATCGCGGCGGATT
SerGlnAsnArgAspGluIleThrLysAspGlyPheGlnSerAsnHisAlaGlyGlyIle [276]
1320 CTGCGGGTATCAGCAGCGGGCAGCAAAATCTTGCCCATGTCGGCTGAACACGACCTCC
LeuGlyGlyIleSerSerGlyGlnGlnIleIleAlaHisMetAlaLysGlyProThrSer [296]
1380 AGCATTACCGTGGCGGCTGTACCATTAACCGCTTTGGCGAAGTGTGAGATGATCACC
SerIleThrValProGlyArgThrIleAsnArgPheGlyGluGluValGluMetIleThr [316]
1440 AAGGCGCTCAGCATGCTGTGTCGGGATCGGCGAGTCCGCGTCCGCAAGCGAAGATGCT
LysGlyArgHisAspProGlyValGlyIleArgAlaValProIleAlaGluAla [336]
1500 GCGGATGCTTTATGATGATCAGCTTACCGGCAACCGGCGCAAAATTCGGATGCTGAAGAC
GlyAspArgPheAsnGlySerProValThrAlaThrGlyAlaLysCysArgCysGluAsp [356]
1560 TGATATTCCACGCTGTGAAAAATGAAATAAACCGCGATTTCGGCTCTGGCTCTGCTGC
1620 CAGTAGGCGCAGCTCGCAGCGGACGGCTGGCAAAAAATAACCCAACCTGTCCCGGTAG
1680 GGCCAAAATCGA [590]

Fig. 3. Nucleotide sequence (sense strand) of the *aroC* gene

The complete *aroC* coding sequence together with flanking 5' (upstream) and 3' (downstream) sequences is shown. The *aroC* structural gene and predicted amino acid sequence are indicated (positions 493–1562). The numbered amino acids in the *aroC* coding region are bracketed ([]). The translational initiation codon is preceded by a proposed ribosome-binding site (RBS). The location of a potential inverted repeat structure 3' to the *aroC* gene is indicated by overlining (see Fig. 6).

DNA sequence of *E. coli* chorismate synthase was confirmed by direct *N*-terminal amino acid sequencing of the purified enzyme in a liquid-phase sequencer. The *N*-terminal amino acid was found to be alanine, indicating that the methionine residue corresponding to the start codon had been lost after translation. The first 30 amino acid residues were identified and agreed exactly with the protein sequence predicted for the open reading frame from the DNA sequence.

Extensive searches upstream of the translational start site (AUG position 492; Fig. 3) failed to identify any

Fig. 4. Application of Fickett's TESTCODE statistic to the sequence of the 1.69 kb *SalI*-*ClaI* fragment

The resulting plot is divided into three regions for which the statistic makes a prediction. Region A is the region within which 95 % of all *bona fide* coding regions fall and where 95 % of non-coding regions do not fall. Region C is the opposite. The middle ground (B) is the region where the statistic does not have a significant ability to predict coding and non-coding sequences. The *aroC*-gene coding region is indicated by the hatched arrow at the top of the Figure. Stop codons ($\diamond$) and Met codons (I) are also shown for all three reading frames.

exact matches with the consensus -35 and -10 *E. coli* promoter elements (Hawley & McClure, 1983). Two candidates each with limited homology to the conserved hexamers and each showing the optimal 16–17 bp separation between hexanucleotide sequences (Mulligan *et al.*, 1986) were identified by eye. The first had the sequence $^{420}\text{TTGCCG}$ -17 bp- ATTTAT^{448} which preceded the triplet GAT at position 453, and the second had the sequence $^{372}\text{TTGATA}$ -16 bp- GTTTAT^{399} which preceded the triplet CAC at position 404 (all numbering as in Fig. 3). Neither of these proposed 'promoters' (and their inferred transcriptional start sites) has been verified experimentally.

Downstream of the *aroC* termination codon is a structure (Fig. 6) that is similar to the rho-independent

Fig. 5. Purification of chorismate synthase from *E. coli* strain AB2849/pGM602

This 10 % denaturing gel monitors the purification of chorismate synthase. Track A, crude extract; track B, 35–60 %-satn.-(NH_4) $_2\text{SO}_4$ fraction; track C, enzyme eluted from DEAE-Sephacel column; track D, enzyme eluted from cellulose phosphate column; track E, enzyme eluted from phenyl-Sepharose column; track F, M_r markers.

terminators of *E. coli*. This structure has a calculated ΔG of -99.6 kJ/mol (-23.8 kcal/mol) but lacks the stretch of uridine residues normally found on the 3' side of the stem (Tinoco *et al.*, 1973). The pattern of codon utilization of the *E. coli aroC* gene is shown in Table 2. The *aroC* gene clearly falls into the weakly expressed category as defined by Grosjean & Fiers (1982). In this respect the *aroC* gene is similar to the *aroA*, *aroB*, *aroD*, *aroE* and *aroL* genes previously sequenced in this laboratory (Duncan *et al.*, 1984; Millar & Coggins, 1986; Millar *et al.*, 1986a; Duncan *et al.*, 1986; Anton & Coggins, 1988), an observation consistent with the low constitutive levels of expression observed for the *aro* genes (Tribe *et al.*, 1976).

To facilitate comparison with the *N. crassa* chorismate synthase a sample of this enzyme was also required. In our experience *N. crassa* chorismate synthase is highly susceptible to endogenous proteinases (P. J. White,

Table 1. Purification scheme for *E. coli* chorismate synthase

The results presented are for a typical purification starting from 20 g of *E. coli* AB2849-pGM602 cells. For experimental details see the text.

Step	Volume (ml)	Concn. of protein (mg/ml)	Total protein (mg)	Activity (units/ml)	Total activity (units)	Specific activity (units/mg)	Purification (fold)	Yield (%)
1: crude extract	60	27.5	1650	0.60	36	0.022	1	100
2: 35–60 %-satd. (NH_4) $_2\text{SO}_4$	44	23.0	1012	0.69	30.5	0.030	1.4	85
3: DEAE-Sephacel	44	4.0	176	0.50	22.0	0.125	5.7	61
4: cellulose phosphate	46	0.08	3.7	0.56	26.0	7.03	319	72
5: phenyl-Sepharose	29	0.04	1.16	0.58	16.8	14.8	658	47

M. R. Boocock & J. R. Coggins, unpublished work) and since the only published method for purifying this enzyme (Welch *et al.*, 1974) predated the recognition of the proteinase problem in *N. crassa* (Lampkin *et al.*, 1976; Gaertner & Cole, 1976; Lumsden & Coggins,

Fig. 6. Putative *aroC*-gene terminator

The possible *aroC*-gene terminator sequence shown is the inverted repeat sequence 3' to the end of the *aroC* gene (indicated by overlining in Fig. 3) as a stem-loop structure.

Table 2. Codon utilization in the *E. coli aroC* gene

	T	C	A	G	
T	Phe 9	Ser 2	Tyr 3	Cys 2	T
	Phe 3	Ser 4	Tyr 4	Cys 4	C
	Leu 1	Ser 1	ochre 0	opal 1	A
	Leu 3	Ser 3	ambre 0	Trp 1	G
C	Leu 1	Pro 3	His 5	Arg 9	T
	Leu 6	Pro 2	His 2	Arg 15	C
	Leu 0	Pro 2	Gln 5	Arg 1	A
	Leu 11	Pro 10	Gln 9	Arg 0	G
A	Ile 13	Thr 2	Asn 3	Ser 2	T
	Ile 14	Thr 15	Asn 6	Ser 7	C
	Ile 0	Thr 2	Lys 15	Arg 0	A
	Met 7	Thr 3	Lys 1	Arg 0	G
G	Val 9	Ala 5	Asp 10	Gly 8	T
	Val 6	Ala 6	Asp 18	Gly 25	C
	Val 1	Ala 5	Gln 13	Gly 6	A
	Val 7	Ala 12	Gln 7	Gly 4	G

1977) a new procedure involving a full antiproteinase strategy similar to that employed in the purification of the *aroM* multifunctional enzyme from the same organism (Lumsden & Coggins, 1977) was developed. The enzyme required to be purified over 3000-fold from crude extract. The purification procedure is outlined in Table 3 and a polyacrylamide gel electrophoretogram showing the protein components at each stage is shown in Fig. 7. The purified enzyme had a specific activity of 32.1 units/mg of protein, which is more than 10-fold greater than the previously published value (Welch *et al.*, 1974).

The subunit M_r values of both enzymes were estimated by polyacrylamide-gel electrophoresis in the presence of SDS. The *E. coli* chorismate synthase was found to have subunit M_r 38000 and the *N. crassa* enzyme subunit M_r 50000. Determination of the quaternary structure of each enzyme was undertaken by using gel-filtration and cross-linking studies. Molecular exclusion chromato-

Fig. 7. Purification of chorismate synthase from *N. crassa*

This 10% denaturing gel monitors the purification of chorismate synthase. Track A, crude extract; track B, negative DE52 DEAE-cellulose chromatography; track C, 40–50% satn.-(NH₄)₂SO₄ fraction; track D, enzyme eluted from DEAE-Sephacel column; track E, enzyme eluted from cellulose phosphate column; track F, enzyme eluted from Mono Q column; track G, M_r markers.

Table 3. Purification scheme for *N. crassa* chorismate synthase

The results represent a purification from 90 g of freeze-dried powdered mycelium. For experimental details see the text.

Step	Volume (ml)	Concn. of protein (mg/ml)	Total protein (mg)	Activity (units/ml)	Total activity (units)	Specific activity (units/mg)	Purification (fold)	Yield (%)
1: crude extract	1320	5.6	7392	0.054	72	0.010	1.0	100
2: negative DE52 DEAE-cellulose	1240	4.0	4960	0.058	72	0.014	1.4	100
3: 40–50% satd. (NH ₄) ₂ SO ₄	105	10.3	1082	0.57	60	0.06	6.0	83
4: DEAE-Sephacel	86	2.6	224	0.60	52	0.23	23	72
5: cellulose phosphate	30	0.36	10.8	1.2	36	3.33	333	50
6: Mono Q	1.5	0.28	0.42	9.0	13.5	32.1	3210	19

Fig. 8. Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulphate of *E. coli* and *N. crassa* chorismate synthase crosslinked with dimethyl suberimide

The 7.5% gels were run as described in the Materials and methods section and stained for protein. Track A, 40 μ g of cross-linked *E. coli* chorismate synthase; track B, 8 μ g of *E. coli* chorismate synthase; track C, 36 μ g of cross-linked *N. crassa* chorismate synthase; track D, 10 μ g of *N. crassa* chorismate synthase.

Fig. 9. Electrophoresis of *N. crassa* chorismate synthase

These non-denaturing 7.5% tube gels of the purified *N. crassa* chorismate synthase were stained (A) for diaphorase activity and (B) for protein as described in the Materials and methods section.

graphy indicated a native M_r for *E. coli* chorismate synthase of 144 000 and for the *N. crassa* enzyme 198 000 (results not shown). The tetrameric nature of each enzyme was confirmed by cross-linking with dimethyl suberimide and the results are shown in Fig. 8.

The purified *N. crassa* chorismate synthase has an intrinsic flavin reductase (diaphorase) activity. Native polyacrylamide-gel electrophoresis of the enzyme shows a single protein band that co-migrates with a single diaphorase activity band (Fig. 9). The purified *E. coli* chorismate synthase has no detectable diaphorase activity.

Fig. 10. Alignment of the sequence of *E. coli* chorismate synthase with the three peptides generated by complete clostripain digest of *N. crassa* chorismate synthase

The sequences were aligned using the Bestfit sequence comparison program of the University of Wisconsin Genetics Computer Group. Identical amino acids are boxed; a conservative substitution is represented by a vertical line and an asterisk indicates a residue that could not be unambiguously identified from amino acid sequencing. The single-letter amino acid abbreviations are used.

Direct *N*-terminal sequencing of the *N. crassa* enzyme was unsuccessful, presumably because the *N*-terminus is blocked. A complete clostripain digestion of the reduced and carboxymethylated enzyme was therefore carried out and the resulting peptides were separated by reverse-phase h.p.l.c. Three of the purified peptides were sequenced and show homologies with the sequence of the *E. coli* enzyme (Fig. 10).

DISCUSSION

The complete amino acid sequence of *E. coli* chorismate synthase (the *aroC*-gene product) has been determined by a combined nucleotide and direct amino acid sequencing strategy. Construction of an overexpressing strain has also permitted for the first time the purification to homogeneity of this unusual enzyme. Homogeneous *N. crassa* chorismate synthase of high specific activity was obtained by using an improved purification procedure, and this has allowed a comparative study to be made of these two enzymes. Detailed structural analysis revealed both enzymes to be tetrameric but to differ significantly in subunit M_r . The *E. coli* enzyme has a subunit M_r of 38 000 whereas the *N. crassa* chorismate synthase is somewhat larger with a subunit M_r of 50 000. The two enzymes appear to have regions of homologous primary structure. Three peptides obtained by the proteolytic digest of the *N. crassa* chorismate synthase show significant homology with stretches of the *E. coli* sequence. These regions of sequence homology occur towards the *C*-terminus of the *E. coli* enzyme. All the chorismate synthases that have been studied to date have an absolute requirement for a reduced flavin cofactor during the reaction. This cofactor is not stoichiometrically consumed during catalysis and has a role that has not yet been unambiguously identified. The *N. crassa* chorismate synthase has the ability to generate this cofactor via an intrinsic flavin reductase (diaphorase) activity (Welch *et al.*, 1974), and native polyacrylamide-gel electrophoresis of the enzyme shows a single protein band that co-migrates with a single diaphorase activity band. All the evidence available indicates that the *N. crassa* enzyme is a homotetramer consisting of four identical polypeptide chains. Each chain is bifunctional,

carrying both chorismate synthase and diaphorase activity.

The *E. coli* enzyme has no detectable diaphorase activity and will only catalyse the formation of chorismate when supplied with exogenous reduced flavin. This reduced flavin can be generated enzymically, for example by pig heart lipoamide dehydrogenase in a reaction mixture scrubbed of dissolved O_2 (Morell *et al.*, 1967), or chemically via the strong reducing agent $Na_2S_2O_4$ (Morell *et al.*, 1967; White *et al.*, 1987). Neither the *E. coli* nor the *N. crassa* enzyme shows activity in the absence of exogenously supplied flavin. The diaphorase activity of the bifunctional *N. crassa* enzyme shows absolute specificity for the physiological substrates FMN and NADPH (Welch *et al.*, 1974; Boocock, 1983). This intrinsic *N. crassa* diaphorase activity can only catalyse the reduction of FMN in the presence of NADPH; FAD is not reduced. However, the *N. crassa* enzyme will catalyse the formation of chorismate if supplied directly with either FMNH₂ or FADH₂ (P. J. White, unpublished work). In contrast, the *E. coli* enzyme is unable to generate reduced flavin via the oxidation of reduced nicotinamide nucleotides; it can synthesize chorismate only when supplied directly with either FMNH₂ or FADH₂. Although the enzymes show activity when supplied with either of the reduced flavin cofactors, both chorismate synthases show a preference for FMNH₂ over FADH₂ (P. J. White, unpublished work).

The ability of both the *N. crassa* and *E. coli* enzymes to utilize directly either FMNH₂ or FADH₂ in the absence of any physiological reducing power suggests the existence of a catalytic domain on each enzyme that has common characteristics. This domain catalyses the formation of chorismate from 5-enolpyruvylshikimate 3-phosphate in the presence of a reduced flavin cofactor (physiologically FMNH₂). On the basis of this evidence and our limited sequence results we propose that the *N. crassa* polypeptide comprises two separate domains: one, which is homologous to the *E. coli* enzyme, carries the chorismate synthase activity and requires a reduced flavin cofactor, and the other, which has not yet been characterized, carries the diaphorase activity. It is proposed that the *E. coli* polypeptide described in this paper has only a chorismate synthase domain and cannot generate reducing power intrinsically. The smaller subunit size of the *E. coli* enzyme may reflect a 'missing' diaphorase domain. Further sequence analysis should allow this domain to be more precisely localized within the *N. crassa* chorismate synthase polypeptide.

It is of interest to compare two other chorismate synthases that have been characterized with the results reported here. The *B. subtilis* enzyme exists as a bifunctional unit comprising two separable polypeptide chains. These chains catalyse distinct reactions: (a) the reduction of either FAD or FMN to reduced flavin via the oxidation of NADPH, and (b) the conversion of 5-enolpyruvylshikimate 3-phosphate into chorismate (Hasan & Nester, 1978).

A higher-plant chorismate synthase from *Pisum sativum* has been assayed in intact chloroplasts (Mousdale & Coggins, 1986) and shows activity only in the presence of exogenously supplied reduced flavin. The four enzymes described here seem to exist in one of two forms: (a) bifunctional enzymes possessing both chorismate synthase and a specific diaphorase activity in covalent or non-covalent association (exemplified by the

N. crassa and the *B. subtilis* enzymes) or (b) monofunctional enzymes possessing only a chorismate synthase activity (exemplified by the *E. coli* and the *Pisum sativum* enzymes).

The enzymic source of physiological reducing power for the monofunctional enzymes has not yet been identified. Some micro-organisms, for example *Aerobacter aerogenes* and *E. coli*, have been shown to possess strong diaphorase activities (Bernofsky & Mills, 1966; Morell *et al.*, 1967; P. J. White, unpublished work), and it is conceivable that *in vivo* the monofunctional chorismate synthases are able to obtain reduced flavin endogenously from such a source. Interestingly, cell-free extracts of *N. crassa* show very little diaphorase activity (D. M. Mousdale, unpublished work), and this may explain why the *N. crassa* chorismate synthase is found as a bifunctional enzyme in covalent association with a specific diaphorase activity.

We thank Professor J. E. Fothergill and Mr. B. Dunbar (University of Aberdeen Protein Sequencing Facility) for determining the amino acid sequences and Dr. D. M. Mousdale for invaluable advice at many stages during this work. We are also grateful to the University of Glasgow for a studentship (G.M.) and to the Science and Engineering Research Council for a studentship (P.J.W.) and for financial support.

REFERENCES

- Anton, I. A. & Coggins, J. R. (1988) *Biochem. J.* **249**, 319–326
- Bachmann, B. (1983) *Microbiol. Rev.* **44**, 180–270
- Bernofsky, C. & Mills, R. C. (1966) *J. Bacteriol.* **92**, 1404–1414
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965
- Boocock, M. R. (1983) Ph.D. Thesis, University of Glasgow
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
- Carter, P. E., Dunbar, B. & Fothergill, J. E. (1983) *Biochem. J.* **215**, 565–571
- Clarke, L. & Carbon, J. (1976) *Cell* **9**, 91–99
- Coggins, J. R., Boocock, M. R., Chaudhuri, S., Lambert, J. M., Lumsden, J., Nimmo, G. A. & Smith, D. D. S. (1987) *Methods Enzymol.* **142**, 325–341
- Davies, G. E. & Stark, G. R. (1970) *Proc. Natl. Acad. Sci. U.S.A.* **66**, 651–656
- Davis, B. J. (1964) *Ann. N.Y. Acad. Sci.* **121**, 404–427
- Devereux, J., Haerberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
- Duncan, K., Lewendon, A. & Coggins, J. R. (1984) *FEBS Lett.* **165**, 121–127
- Duncan, K., Chaudhuri, S., Campbell, M. S. & Coggins, J. R. (1986) *Biochem. J.* **238**, 475–485
- Floss, H. G., Onderka, D. K. & Carroll, M. (1972) *J. Biol. Chem.* **247**, 736–743
- Gaertner, F. H. & Cole, K. W. (1976) *Arch. Biochem. Biophys.* **177**, 566–573
- Ganem, B. (1978) *Tetrahedron* **34**, 3353–3383
- Gibson, F. (1970) *Methods Enzymol.* **17**, 362–364
- Grosjean, H. & Fiers, W. (1982) *Gene* **18**, 199–209
- Hagervall, T. G. & Björk, G. R. (1984) *Mol. Gen. Genet.* **196**, 201–207
- Hasan, N. & Nester, E. W. (1978) *J. Biol. Chem.* **253**, 4999–5004
- Haslam, E. (1974) *The Shikimate Pathway*, Butterworths, London
- Hawley, D. K. & McClure, W. R. (1983) *Nucleic Acids Res.* **11**, 2237–2255
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lampkin, S. L., Cole, K. W., Vitto, A. & Gaertner, F. H. (1976) *Arch. Biochem. Biophys.* **177**, 561–565

- Lumsden, J. & Coggins, J. R. (1977) *Biochem. J.* **161**, 599–607
- Lumsden, J. & Coggins, J. R. (1978) *Biochem. J.* **169**, 441–444
- Maniatis, T., Fritsch, E. F. & Sambrook, H. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276
- Millar, G. & Coggins, J. R. (1986) *FEBS Lett.* **200**, 11–17
- Millar, G., Lewendon, A., Hunter, M. G. & Coggins, J. R. (1986a) *Biochem. J.* **237**, 427–437
- Millar, G., Anton, I. A., Mousdale, D. M., White P. J. & Coggins, J. R. (1986b) *Biochem. Soc. Trans.* **14**, 262–263
- Mitchell, W. M. & Harrington, W. F. (1971) *Methods Enzymol.* **19**, 635–642
- Morell, H., Clark, M. J., Knowles, P. F. & Sprinson, D. B. (1967) *J. Biol. Chem.* **242**, 82–90
- Mousdale, D. M. & Coggins, J. R. (1986) *FEBS Lett.* **205**, 328–332
- Mulligan, M. E., Brosius, J. & McClure, W. R. (1986) *J. Biol. Chem.* **260**, 3529–3538
- Pittard, J. & Wallace, B. J. (1966) *J. Bacteriol.* **91**, 1494–1508
- Russell, G. A., Dunbar, B. & Fothergill-Gilmore, L. A. (1986) *Biochem. J.* **236**, 115–126
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254**, 34–38
- Smith, M. A., Gerrie, L. M., Dunbar, B. & Fothergill, J. E. (1982) *Biochem. J.* **207**, 253–260
- Staden, R. (1980) *Nucleic Acids Res.* **8**, 3673–3694
- Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London)* **246**, 40–41
- Tribe, D. E., Camakaris, H. & Pittard, J. (1976) *J. Bacteriol.* **127**, 1085–1097
- Twigg, A. J. & Sherratt, D. (1980) *Nature (London)* **283**, 216–218
- Weber, K. & Osborn, M. (1969) *J. Biol. Chem.* **244**, 4406–4412
- Weiss, U. & Edwards, J. M. (1980) *The Biosynthesis of Aromatic Compounds*, pp. 103–133, John Wiley and Sons, New York
- Welch, G. R., Cole, K. W. & Gaertner, F. H. (1974) *Arch. Biochem. Biophys.* **165**, 505–518
- White, P. J., Mousdale, D. M. & Coggins, J. R. (1987) *Biochem. Soc. Trans.* **15**, 144–145

Received 30 September 1987; accepted 30 November 1987