Biochem. J. (1986) 237, 427–437 (Printed in Great Britain)

The cloning and expression of the *aroL* gene from *Escherichia coli* K12

Purification and complete amino acid sequence of shikimate kinase II, the *aroL*-gene product

Gary MILLAR,* Ann LEWENDON,* Michael G. HUNTER† and John R. COGGINS*‡
*Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K., and
†Searle Research and Development, P.O. Box 53, Lane End Road, High Wycombe, Bucks, HP12 4HL, U.K.

The *aroL* gene encoding the enzyme shikimate kinase II was cloned from *Escherichia coli* K12. Construction of over-expressing strains permitted for the first time the purification to homogeneity of a monofunctional shikimate kinase. The complete amino acid sequence of shikimate kinase II was determined by a combined nucleotide and direct amino acid sequencing strategy. *E. coli* shikimate kinase II is a monomeric enzyme containing 173 amino acid residues with a calculated M_r 18937. The amino acid sequence contains a region homologous with other kinases and ATP-requiring enzymes. Evidence is presented suggesting that the transcriptional start site of the *aroL* gene is located within a potential operator site.

INTRODUCTION

Shikimate kinase (EC 2.7.1.71) catalyses the phosphorylation of shikimic acid to yield shikimate 3-phosphate. This reaction is the fifth step of the early common pathway for aromatic amino acid biosynthesis (the shikimate pathway) in micro-organisms. One of the remarkable features of this pathway is the very different structural organization of the enzymes in bacteria and fungi. It has been shown that in *Neurospora crassa* the central five enzymes of the seven-step pathway occur as a multifunctional enzyme consisting of two identical pentafunctional polypeptide chains of M_r 165000 (Lumsden & Coggins, 1977, 1978; Gaertner & Cole, 1977; Smith & Coggins, 1983; Lambert *et al.*, 1985). A similar multifunctional enzyme occurs in *Saccharomyces cerevisiae* (Larimer *et al.*, 1983; K. Duncan, R. M. Edwards & J. R. Coggins, unpublished work) and in a number of other species of fungi (Strauss, 1979; Bode & Birnbaum, 1981; Kinghorn & Hawkins, 1982; Nakanishi & Yamamoto, 1984). In contrast, in *Escherichia coli* the corresponding enzymes are separable (Berlyn & Giles, 1969) and presumably monofunctional. Four of these monofunctional *E. coli* enzymes, 3-dehydroquinate synthase (Frost *et al.*, 1984), 3-dehydroquinase (Chaudhuri *et al.*, 1986), shikimate dehydrogenase (Chaudhuri & Coggins, 1985) and 5-enolpyruvylshikimate-3-phosphate synthase (Lewendon & Coggins, 1983; Duncan *et al.*, 1984), have been purified and characterized. As part of a detailed comparative study of the *N. crassa* and *S. cerevisiae* multifunctional enzymes with the five corresponding monofunctional *E. coli* enzymes we required to purify and characterize *E. coli* shikimate kinase.

Two shikimate kinase activities have been detected in *E. coli* (Ely & Pittard, 1979) and in *Salmonella typhimurium* (Morell & Sprinson, 1968; Berlyn & Giles, 1969). Neither of the *E. coli* isoenzymes has been purified and in only one case, shikimate kinase II, has the gene (*aroL*) been located (Ely & Pittard, 1979). In *Bacillus subtilis* there is a single shikimate kinase, which is a component of a trifunctional multienzyme complex. This complex contains a bifunctional polypeptide carrying 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase and chorismate mutase activities as well as a monofunctional shikimate kinase polypeptide (Nakatsukasa & Nester, 1972). The kinase polypeptide, which is active only in the complex, has been purified to homogeneity and has an M_r of 10000 (Huang *et al.*, 1975). In *B. subtilis* the regulatory properties of shikimate kinase suggest that it may represent a crucial allosteric step in the shikimate pathway (Huang *et al.*, 1975; Nasser *et al.*, 1969). The reason for the existence of two shikimate kinase isoenzymes in *E. coli* is not known, but it has been suggested that shikimate may be a branch-point intermediate for two distinct pathways (Weiss & Edwards, 1980).

The chromosomal location of the *aroL* gene is known (Ely & Pittard, 1979); it maps close to the *proC* and *phoA* genes at 9 min on the *E. coli* chromosome (Bachmann, 1983). The present paper reports the cloning and sequence of the *E. coli aroL* gene and the over-production, purification to homogeneity and *N*-terminal sequencing of the gene product, shikimate kinase II.

MATERIALS AND METHODS

Media and bacterial strains

Cells were prepared by growth at 37 °C on minimal media (Maniatis *et al.*, 1982) supplemented with 0.2% glucose containing, where appropriate, an antibiotic or nutritional supplement. Amino acid supplements were added at 50 μg/ml (final concentration). Bacterial strains used are listed in Table 1, and the plasmids used are listed in Table 2.

Transformation of *E. coli*

Competent cells were prepared and transformations carried out as described by Daghert & Ehrlich (1979).

Abbreviations used: bp, base-pairs; f.p.l.c., fast protein liquid chromatography.

‡ To whom correspondence should be addressed.

Table 1. Bacterial strains used

Organism	Genotype	Source/reference
E. coli K12	Wild-type A.T.C.C. 14948, F^-, λ lysogenic	American Type Culture Collection
E. coli HB101	F^-, *pro*, *leu*, *thi*, *lac*Y, *hsd*R, *end*A, *rec*A, *rpsL*20, *ara*14, *gal*K2, *xyl*5, *mt*l1, *sup*E44	Bolivar & Backman (1979)
E. coli JM101	Δ(*lac pro*), *thi*, *sup*E, F'*tra*D36, *pro*AB, *lac*I^q, *lac*Z, ΔM15	Messing *et al.* (1981)
E. coli HW0927	*pro*C32, *met*E70, *trp*E38, *ara*-14, *lac*Z36, *mtl*-1, *xyl*5, *thi*1, *pur*E42, *rec*A1, a216, *tsx*-67, *rpsl*109, *ton*A, *sup*E	Searle
E. coli HW1045	*tyr*R, *tyr*A, *trp*R	Searle
E. coli HW87	*ara*D139 (*ara-leu*), Δ7697, *lac*ZIPOZY, Δ74, *gal*U, *Gal*K, *hsd*R, *rps*L, *srl*, *rec*A56	Searle

Table 2. Plasmids used

Plasmid	Description	Source
pAT153	amp^R, tet^R	Twigg & Sherratt (1980)
pMH423	10 kbp *Eco*RI fragment carrying *aroL* in pAT153	Present study
pGM424	2.7 kbp *Bam*HI subclone of *aroL* in pAT153	Present study
pGM425	*Pvu*II deletion of pGM424	Present study
pGM63A	*Bam*HI deletion of pMH423	Present study

Transformed cells were selected for antibiotic-resistance on L agar (Maniatis *et al.*, 1982) supplemented with ampicillin (50 μg/ml), carbenicillin (200 μg/ml) or tetracycline (20 μg/ml).

Bacteriophage-M13-transformed cells were plated on H agar (1.2%, w/v) containing 1% (w/v) bactotryptone and 0.8% NaCl.

Molecular-biological techniques

Restriction endonucleases and bacteriophage-T4 DNA ligase were purchased from Bethesda Research Laboratories (Paisley, Renfrewshire, Scotland, U.K.), avian-myeloblastosis-virus reverse transcriptase was from NBL (Cramlington, Northumbria, U.K.), Klenow fragment of *E. coli* DNA polymerase was from Amersham International (Amersham, Bucks., U.K.) and coupling enzymes (pyruvate kinase/lactate dehydrogenase) were from Boehringer (Lewes, East Sussex, U.K.). Digests were performed under conditions recommended by the suppliers, and restricted DNA products were separated by electrophoresis on 1% (w/v) agarose gels. DNA fragments were isolated from low-melting-temperature agarose by phenol and chloroform extraction before ethanol precipitation (Maniatis *et al.*, 1982).

All DNA sequencing was by the di-deoxy chain-termination method (Sanger *et al.*, 1977; Biggin *et al.*, 1983). The conditions for annealing and for the sequencing reactions and the electrophoretic conditions were as described previously (Duncan *et al.*, 1984). [α-[^{35}S]Thio]dATP obtained from Amersham International was used as label.

Extensive use was made of the Messing series of M13 vectors (Messing & Vieira, 1982) to subclone segments of the 1.5 kbp *Bam*HI–*Pvu*II insert of plasmid pGM425. The sequencing strategy is summarized in Fig. 2. *Sau*3A, *Taq*I and *Bam*HI M13 clones were sequenced, and the sequences were entered as separate files into a Digital PDP11-34 computer. The sequences were linked up by using Staden's (1980) programs.

For transcript mapping RNA was prepared from exponentially growing *E. coli* HB101/pGM425 cells as described by Aiba *et al.* (1981). A synthetic oligonucleotide primer, 25 nucleotide residues long

(sequence

5'-TACCACAGCCCCGAGGCCCGATCAG-OH-3')

and complementary to nucleotides 18–43 in the *aroL*-gene coding sequence, was annealed to the RNA at 55 °C for 30 min in 10 mM-Tris/HCl buffer, pH 8.5, containing 10 mM-$MgCl_2$. Primer extension synthesis of the run-off transcripts was carried out by adding 10 μCi of [α-[^{35}S]thio]dATP and dGTP, dCTP and dTTP to a final concentration of 0.1 mM each and avian-myeloblastosis-virus reverse transcriptase (23 units). Incubation was at 30 °C for 30 min. Run-off transcripts were analysed by the use of polyacrylamide sequencing gels.

Determination of shikimate kinase activity in crude extracts

Crude extracts of untransformed wild-type *E. coli* and of plasmid-transformed strains were prepared by two passages through a French press at 6200 kPa (900 lbf/in^2) in 50 mM-Tris/HCl buffer, pH 7.5, containing 5 mM-$MgCl_2$. After centrifugation at 100000 *g* for 2 h the supernatants from these extracts were subjected to chromatography on a Mono-Q (Pharmacia) f.p.l.c. column. Bacterial protein [5 mg as determined by the Bradford (1976) method] was applied to the column, which had been previously equilibrated with 50 mM-Tris/HCl buffer, pH 7.5, containing 5 mM-$MgCl_2$. The column was then washed with a linear gradient of 0–300 mM-KCl in 50 mM-Tris/HCl buffer, pH 7.5, con-

taining 5 mM-$MgCl_2$; a constant flow rate of 1 ml/min was used throughout, and 1 ml fractions were collected. Fractions containing shikimate kinase activity were identified by assay.

Assay of shikimate kinase activity

Shikimate kinase activity was assayed at 25 °C by coupling the release of ADP to the pyruvate kinase (EC 2.7.1.40) and lactate dehydrogenase (EC 1.1.1.27) reactions. Shikimate-dependent oxidation of NADH was monitored at 340 nm ($\epsilon = 6180\ M^{-1} \cdot cm^{-1}$). The assay mixture contained (final concentrations) 50 mM-triethanolamine hydrochloride/KOH buffer, pH 7.0, 50 mM-KCl, 5 mM-$MgCl_2$, 1.6 mM-shikimic acid, 2.5 mM-ATP, 1 mM-phosphoenolpyruvate, 0.1 mM-NADH, 3 units of pyruvate kinase/ml and 2.5 units of lactate dehydrogenase/ml.

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

Purification of shikimate kinase

All steps after cell breakage were performed at 0–4 °C.

Extraction and centrifugation. *E. coli* HW87/pMH423 cells (16 g) were passed twice through a French pressure cell at 6200 kPa, and were then extracted with 30 ml of 50 mM-Tris/HCl buffer, pH 7.5, containing 50 mM-KCl, 5 mM-$MgCl_2$ and 0.4 mM-dithiothreitol (buffer A). Deoxyribonuclease I (Sigma type IV) (0.5 mg) was added and the extract was stirred for 1 h, then centrifuged at 100000 *g* for 2 h. The resulting supernatant (the crude extract) was dialysed for 4 h against buffer A (2000 ml).

DEAE-Sephacel chromatography. The dialysed crude extract was loaded on to a column of DEAE-Sephacel (10 cm × 3.5 cm diam.) equilibrated in buffer A. The column was washed with 350 ml of buffer A, and then the shikimate kinase activity eluted with a linear gradient (600 ml) of 0–300 mM-KCl in buffer A. The flow rate was 50 ml/h, and 6 ml fractions were collected. Fractions containing shikimate kinase activity were pooled before chromatography on phenyl-Sepharose.

Phenyl-Sepharose chromatography. Solid $(NH_4)_2SO_4$ was added to the pooled fractions from step 2 to give a final concentration of 164 g/l (30% saturation), and after being stirred for 15 min the solution was centrifuged at 23000 *g* for 15 min. The supernatant was loaded on to a column of phenyl-Sepharose CL-4B (4 cm × 2 cm diam.) equilibrated in 100 mM-Tris/HCl buffer, pH 7.5, containing 1.2 M-$(NH_4)_2SO_4$ and 0.4 mM-dithiothreitol. The column was then washed with 150 ml of this buffer, and the shikimate kinase activity was eluted with a linear gradient (400 ml) of 1.2–0.0 M-$(NH_4)_2SO_4$ in 100 mM-Tris/HCl buffer, pH 7.5, containing 0.4 mM-dithiothreitol. The flow rate throughout was 60 ml/h, and 5 ml fractions were collected.

Fractions containing shikimate kinase activity were pooled and concentrated by vacuum dialysis. The concentrated phenyl-Sepharose pool (approx. 6 ml) was dialysed overnight against 500 ml of buffer A containing 10% (v/v) glycerol.

Gel filtration on Sephacryl S200. The concentrated enzyme solution was applied to a Sephacryl S200 (superfine grade) column (85 cm × 2.5 cm diam.) that had been equilibrated in buffer A, and the enzyme was eluted with the same buffer; the flow rate was 8 ml/h, and 2 ml fractions were collected. Fractions containing shikimate kinase activity were dialysed against 500 ml of buffer A containing 50% (v/v) glycerol before long-term storage at −20 °C.

Protein determination

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

Determination of M_r by gel filtration

The native M_r was estimated by gel-permeation chromatography at 25 °C on a Superose 12 column in a Pharmacia f.p.l.c. 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 [M_r values from Weber & Osborn (1969) unless otherwise stated]: chicken ovalbumin (M_r 45000); *E. coli* 3-dehydroquinate synthase (M_r 38900; G. Millar & J. R. Coggins, unpublished work); bovine erythrocyte carbonic anhydrase (M_r 29000); sperm-whale myoglobin (M_r 17200); horse heart cytochrome *c* (M_r 12500).

Amino acid analysis

Samples of performic acid-oxidized shikimate kinase (Hirs, 1967) were hydrolysed and analysed on an LKB model 4400 amino acid analyser as described previously (Lumsden & Coggins, 1978).

Automatic amino acid sequence determination

A 20 nmol portion of *E. coli* shikimate kinase II, which had been reduced and carboxymethylated by the method of Lumsden & Coggins (1978), was sequenced by using 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 C_{18} reverse-phase column with a pH 5.0 acetate/acetonitrile buffer system (Carter *et al.*, 1983).

RESULTS

Cloning of the *aroL* gene

The *aroL* gene has been mapped to 9 min on the *E. coli* genetic map (Ely & Pittard, 1979; Bachmann, 1983). Since a suitable genetic background to enable direct selection of *aroL* clones was not available, an adjacent selectable locus (*proC*) was cloned. The *proC* gene was directly selected by screening a number of gene banks of *E. coli* W3110 DNA cloned into pAT153, for plasmids capable of complementing the *proC* lesion in the *E. coli* strain HW0927. The gene banks were transformed into *E. coli* HW0927, and *carb* resistant transformants were selected on L-agar plates supplemented with carbenicillin. The transformants were replica-plated on to minimal medium lacking proline as a supplement to identify those colonies that had gained an intact *proC* gene, and DNA was prepared from these colonies. Restriction-endonuclease analysis indicated that there were two classes of *proC* clones. The first class, isolated from an *Eco*RI library, was shown to be related to the second class, isolated from a *Bgl*II library, by Southern-blot analysis. One plasmid from the *Eco*RI library, which contained a

10 kbp insert, was chosen for further analysis. *E. coli* strains harbouring this plasmid (pMH423) had a 45-fold-elevated level of shikimate kinase activity, which clearly demonstrated the presence of the *aro*L gene within the insert. The precise locations of *pro*C and *aro*L clones in plasmid pMH423 were ascertained by sub-cloning fragments and testing both for *pro*C-gene complementation and shikimate kinase over-expression.

Plasmid pMH423 was digested with *Bam*HI restriction endonuclease, and a 2.7 kbp and a 6.3 kbp fragment were recovered from a low-melting-point agarose gel. The 2.7 kbp fragment was ligated to *Bam*HI (within the tetracycline-resistance gene)-treated plasmid vector pAT153 (Twigg & Sherratt, 1980), and the 6.3 kbp fragment was recircularized with bacteriophage-T4 DNA ligase. Both preparations were used to transform $CaCl_2$-treated *E. coli* HB101 cells, and the transformation mix was plated on L-agar supplemented with 50 μg of ampicillin/ml. After overnight growth, 100 colonies were replica-plated on to L-agar containing tetracycline (20 μg/ml). Ten colonies with an amp^r, tet^s phenotype were selected, DNA was prepared (Holmes & Quigley, 1981) and the restriction-endonuclease-digest pattern was characterized. Plasmids pGM63A and pGM424 represent the *Bam*HI-deleted recircularized derivative of plasmid pMH423 and the 2.7 kbp *Bam*HI sub-clone in pAT153 respectively. Plasmid pGM425 was constructed in a similar way by deleting 1.2 kbp of DNA between two *Pvu*II sites in the insert of plasmid pGM424 (Fig. 1).

The *Bam*HI sites of plasmid pMH423 mapped close to, but separate from, the two *Pst*I sites known to reside within the *proC* coding region (Deutch *et al.*, 1982). Plasmid pGM424 could no longer relieve the auxotrophic requirements of the pro*C E. coli* strain HW0927. Plasmid pGM63A still carried the region of the *E. coli* chromosome encoding *proC* and could overcome the nutritional constraints of *E. coli* HW0927 (Table 3).

Enzyme assays of partially purified cell extracts revealed that bacterial strains transformed with plasmids pMH423, pGM424 and pGM425 all over-expressed shikimate kinase activity relative to the wild-type level (Table 3), whereas plasmid pGM63A did not confer a similar phenotype upon its host strain.

Transcription–translation analysis *in vitro* of these putative *aroL* clones revealed that plasmids pMH423, pGM424 and pGM425 were all expressing a 19 kDa protein from the cloned DNA; a 25 kDa protein also observed from plasmids pMH423 and pGM424 was absent from plasmid pGM425 (results not shown).

Over-expression of shikimate kinase

Because of interfering NADH oxidase activity it is difficult to measure the low activity of shikimate kinase found in wild-type-cell extracts by using the coupled assay

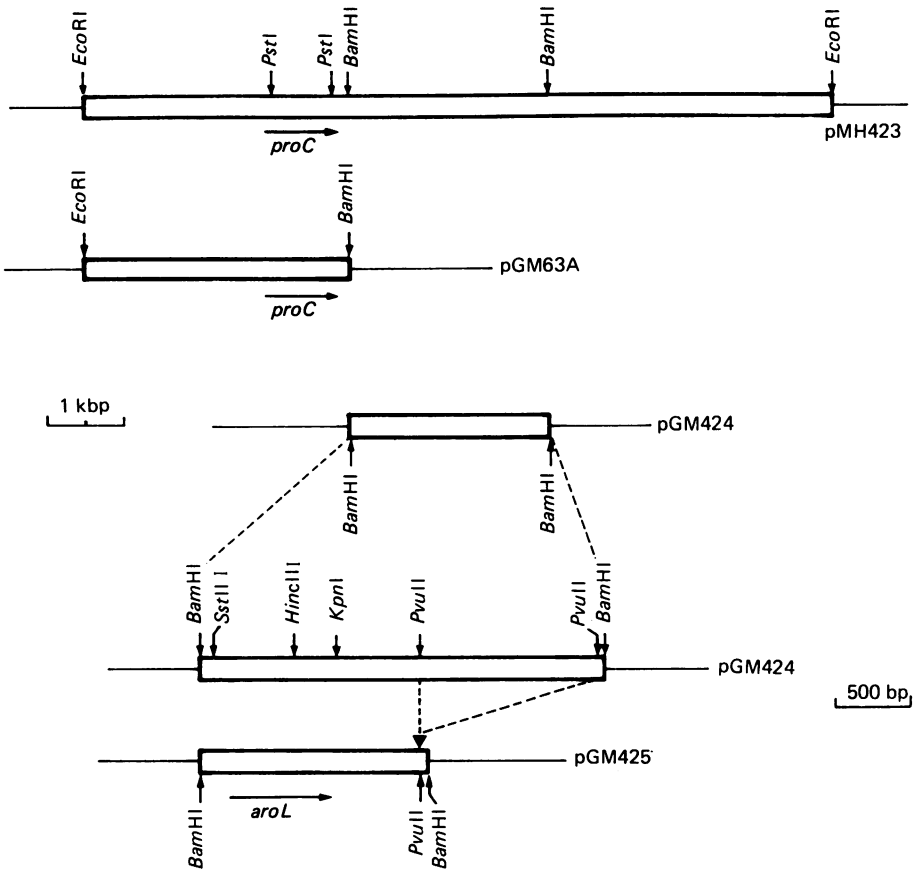

Fig. 1. Subcloning strategy for *E. coli* K12 *aro*L gene

Boxed areas represent *E. coli* K12 genomic DNA inserts cloned into plasmid vector pAT153 (continuous lines). The restriction sites used to generate and orientate the successive subclones are indicated. The locations of *proC* and *aroL* coding regions are marked. The deletion employed in generating pGM425 is shown as ▼.

Table 3. Over-expression of shikimate kinase in plasmid-transformed strains of *E. coli*

Shikimate kinase activity in cell extracts of *E. coli* HB101, both transformed and untransformed, was determined before (crude extract) and after chromatography on a Pharmacia Mono-Q f.p.l.c. anion-exchange column as described in the Materials and methods section. The recovery of activity and the degree of over-expression of shikimate kinase activity in the transformed strains are shown. Abbreviation: N.D., not determined.

Plasmid classification	*proC* complementation	Shikimate kinase specific activity (units/mg of protein)		Activity recovered (%)	Shikimate kinase over-expression (fold)
		Crude extract	F.p.l.c. fraction		
None (wild-type)	N.D.	N.D.	0.08	N.D.	1
pMH423	+	0.35	3.55	100	45
pGM63A	+	N.D.	N.D.	N.D.	N.D.
pGM424	−	1.36	8.3	75	103
pGM425	−	1.73	17.6	83	220

procedure. The enzyme can, however, be readily assayed after ion-exchange chromatography on a Mono-Q (Pharmacia) f.p.l.c. column. This column resolves the two shikimate kinase isoenzymes, which had previously been separated on a different type of ion-exchange column by Ely & Pittard (1979). The position of elution of the over-expressed shikimate kinase II was first established with extracts of the over-producing strain (HB101/pMH423). This then allowed the location and determination of the shikimate kinase II activity in wild-type (*E. coli* K12) crude extracts (Table 3). *E. coli* HB101 (recA) was used as a host for plasmids carrying the *aroL* gene to prevent recombinations with the *aroL* chromosomal locus. *E. coli* HB101 had the same amount of shikimate kinase II activity as the wild-type (K12) strain. The identical behaviour of the wild-type shikimate kinase II and the plasmid-encoded shikimate kinase II on these very-high-resolution f.p.l.c. columns indicated that the over-produced enzyme was the same as, or was at least very similar to, the wild-type enzyme.

Plasmid-pGM424-encoded shikimate kinase activities were also determined in crude extracts of transformed hosts *E. coli* HB101 and *E. coli* HW1045 (tyrR). Shikimate kinase activities were elevated by 50% in the repressor-deficient background relative to the 'wild-type' (HB101) value. This increased shikimate kinase activity is on top of the already greatly over-expressed enzyme activity in the plasmid-transformed 'wild-type' strain. This observation clearly indicates that the plasmid-encoded *aroL* gene is subject to regulation by the *tyrR*-gene product and as such probably carries the necessary sequences for repressor binding.

DNA sequence determination of *aroL* gene

The 1.5 kbp *Bam*HI–*Pvu*II region of chromosomal DNA cloned within plasmid pGM425 (Fig. 1) was identified as being likely to carry the *E. coli aroL* coding sequence. A library of bacteriophage M13 clones carrying sub-segments of this region was constructed (see the Materials and methods section) and the DNA sequence of each was determined by the primer-extension dideoxy method of Sanger *et al.* (1977). The sequences obtained were aligned until 1330 bp had been determined from the *Bam*HI site towards the *Pvu*II site. All restriction sites used for cloning were overlapped and the sequence was determined on both strands for 1200 bp.

Both strands were examined for open reading frames by searching for translational start and stop codons. Only one open reading frame capable of encoding a polypeptide of molecular mass greater than 10 kDa was identified. When translated from the first methionine codon the sequence predicted a polypeptide chain of M_r 19068. This open reading frame extended from positions 286 to 810 in the sequence (Fig. 3) running towards the *Pvu*II site from the *Bam*HI site. Evidence that this was a coding region was obtained by using a computer

Table 4. *N*-Terminal amino acid sequence of *E. coli* shikimate kinase II

The sequence was determined, as described in the Materials and methods section, on 20 nmol of protein. The repetitive yield from residues 1–24, by least-squares regression analysis, was 97% (correlation coefficient 0.83).

Residue no.	Amino acid identified (as phenylthiohydantoin)	Yield (nmol)
1	Thr	2.3
2	Gln	13.0
3	Pro	8.5
4	Leu	11.5
5	Phe	12.3
6	Leu	11.5
7	Ile	10.6
8	Gly	10.0
9	Pro	8.0
10	Arg	12.2
11	Gly	7.6
12	Cys	5.2
13	Gly	7.8
14	Lys	5.6
15	Thr	2.9
16	Thr	3.3
17	Val	7.1
18	Gly	6.7
19	Met	6.9
20	Ala	6.3
21	Leu	6.7
22	Ala	6.2
23	Asp	6.3
24	Ser	0.5

program that examines the positional base frequencies in all three reading frames (Staden, 1984). This was confirmed by direct *N*-terminal amino acid sequencing of purified shikimate kinase II in a liquid-phase sequencer. The *N*-terminal amino acid was found to be threonine, indicating that the methionine residue corresponding to the start codon had been lost after translation. The first 24 amino acids were identified and agreed exactly with the protein sequence predicted for the open reading frame from the DNA sequence (Table 4).

Examination of the nucleotide sequence directly upstream of the initiation codon revealed a good consensus ribosome-binding site (Shine & Dalgarno, 1975), which has the sequence GGA (nucleotides 274–276). Analysis for potential promoter sequences by using the ANALYSEQ computer program (Staden, 1984) identified a region with extensive homology to the consensus prokaryotic promoter −35 and −10 regions. This indicated that the most likely transcription start site upstream of the *aroL* coding sequence was at positions 125–127 bp upstream of the translation initiation codon.

As is shown in Fig. 2, the entire *aroL* coding region (including 285 bp of upstream sequence and 400 bp of downstream sequence) has been sequenced on both strands. The complete sequence of the *aroL* coding region and 285 bp of upstream sequence are given in Fig. 3.

Transcript mapping

The 5′ end of the *aroL* transcript was located by electrophoresis of the oligonucleotide-primed reverse run-off transcript beside a sequence ladder of M13 DNA (Fig. 4). This established that an A residue 125 bp upstream of the initiation codon methionine (ATG) was the start of the *aroL* transcript (Fig. 4).

Analysis of the nucleotide sequence immediately upstream of the transcription start site revealed a −10 and a −35 region (16 bp apart) with considerable homology to the consensus sequence of *E. coli* promoter

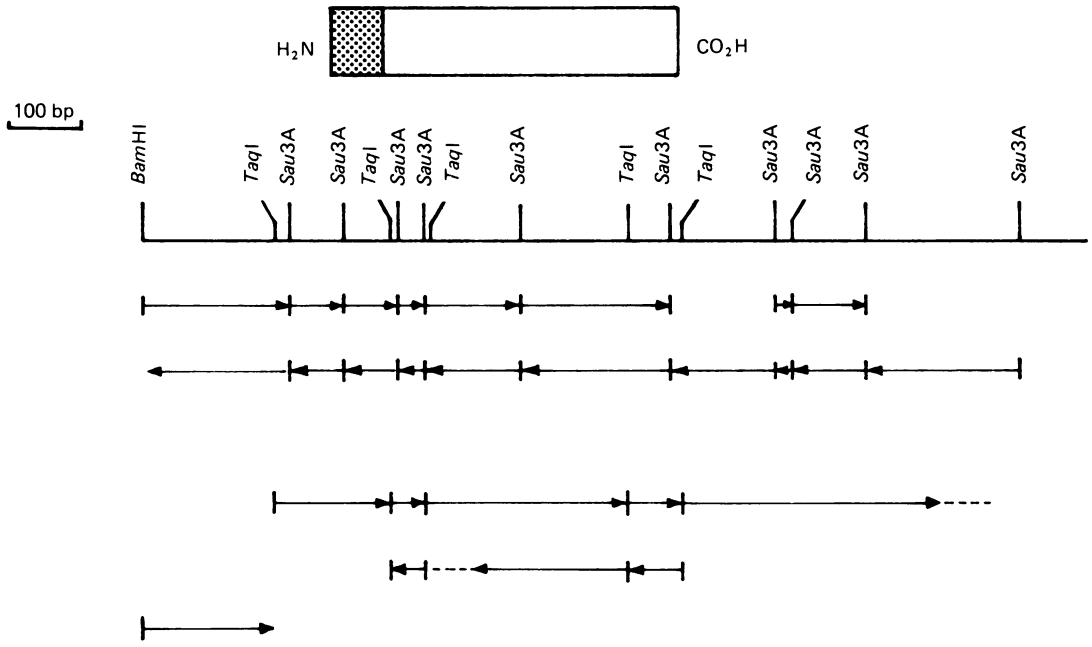

Fig. 2. Sequence analysis of the *E. coli aroL* gene

The boxed area above the restriction map represents the *aroL* coding region. The stippled area indicates the extent of *N*-terminal amino acid sequencing. The restriction map extends from the *Bam*HI site of pGM425 towards the *Pvu*II site. Arrows indicate the size and orientation of individual sequenced M13 clones.

sequences (Hawley & McClure, 1983) (Fig. 3). Overlapping the transcript start is an 18 bp imperfect inverted repeat sequence (L2) whose axis of symmetry is centred on position +5 (Fig. 3). This sequence exhibits striking homology both with the proposed *aroF* operator sequence (Fig. 5) (Garner & Herrmann, 1985) and with the consensus prokaryotic DNA-binding protein recognition sequence $TGTGTN_{(6-10)}ACACA$ (Gicquel-Sanzey & Cossart, 1982). Situated 52 bp upstream of L2, and overlapping the inferred −35 region of the *aroL* promoter, is the axis of another 18 bp inverted repeat (L1) sharing similar properties (Fig. 3). Garner & Herrmann (1985) have shown that the *tyrR*-gene-product-binding site within the *aroF* operator is composed of two 18 bp sequences located 51 bp apart. Since *aroL*-gene expression is also subject to *tyrR*-gene-product regulation (Ely & Pittard, 1979), this strongly suggests that these sequences (L1 and L2) may be involved in the binding of this regulatory protein. The location of the transcription start site and the −35 region of the promoter within these

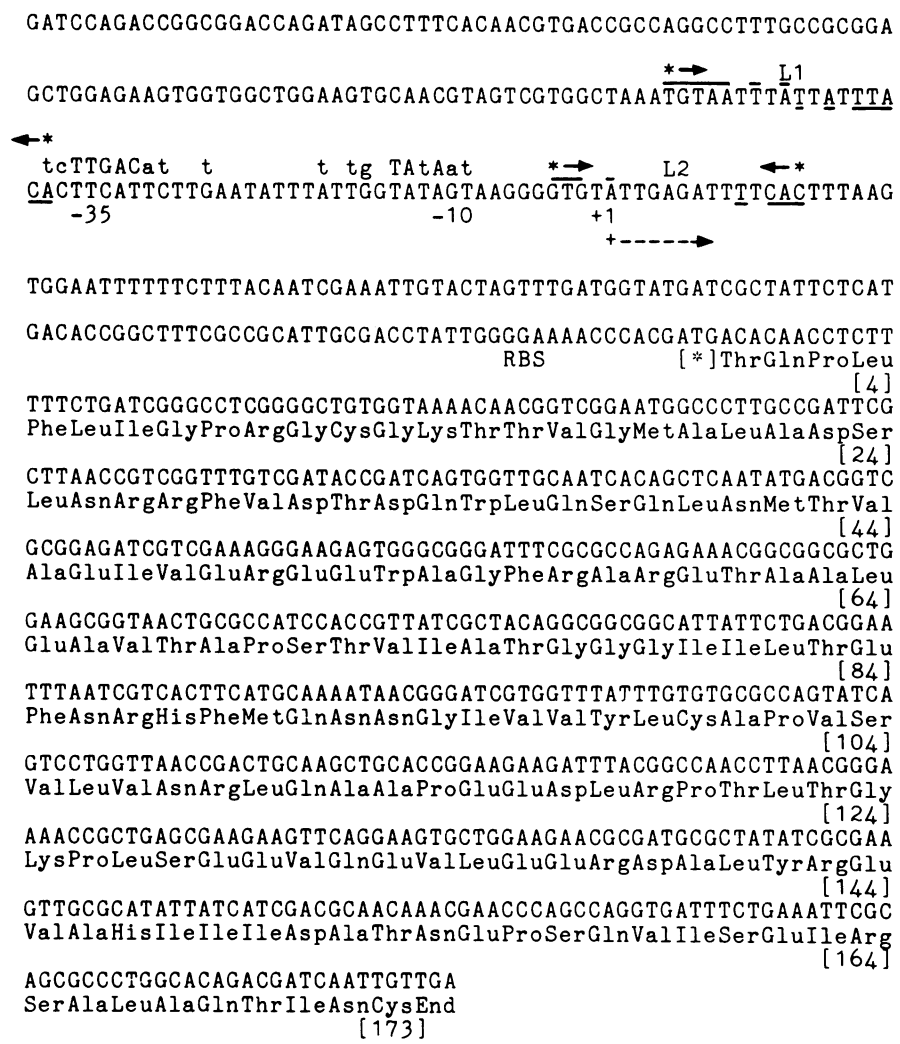

Fig. 3. Complete nucleotide sequence of *E. coli aroL* gene and upstream sequence

The complete DNA sequence, from the second G residue of the *Bam*HI recognition sequence, is given and numbered relative to the transcription start site (+1 +----►). Peptide sequence is shown above the nucleotide sequence and numbered ([ ]) relative to the *N*-terminal threonine residue [1]. The locations of the ribosome-binding site (RBS) and the initiation methionine residue lost post-translationally ([*]) are indicated. Regions of dyad symmetry (L1 and L2) are overlined and underlined on either side of their respective axes of symmetry and their boundaries indicated (*→, ←*). The promoter regions (−10 and −35) are shown with the consensus sequences given above.

possible regulatory regions strongly supports this hypothesis.

In addition to the full-length reverse run-off transcript (A), two other faint transcript bands (I and II) can be seen in Fig. 4, occurring 98 and 102 bp upstream of the translation start. Both are located in a region, between 87 and 105 bp upstream of the initiation codon (ATG), which shares some of the properties of the L1 and L2 sequences. This region has a potential to form secondary structures due to partial dyad symmetry, and this presumably accounts for the occurrence of premature transcript terminations at locations I and II.

Purification of shikimate kinase

Shikimate kinase was purified from an over-producing strain, *E. coli* HW87/pMH423, as detailed in Table 5. An overall yield of 35% was obtained, and a purification of 80-fold resulted in homogeneous shikimate kinase as judged by polyacrylamide-gel electrophoresis in the presence of SDS (gel not shown).

Structure of shikimate kinase II

The subunit M_r of *E. coli* shikimate kinase II was estimated to be 20000 by polyacrylamide-gel electrophoresis in the presence of SDS. The native M_r estimated by gel filtration on a Superose 12 column was $22000 \pm 5\%$. These results indicate that *E. coli* shikimate kinase II is a monomeric enzyme. Both values are in good agreement with the value predicted from the DNA sequence data of 18937.

The amino acid composition of an acid hydrolysate of shikimate kinase was determined (Table 6). Comparison with the values predicted from the deduced nucleotide sequence are also shown in Table 6. The overall agreement between the experimentally obtained and theoretically predicted values is excellent.

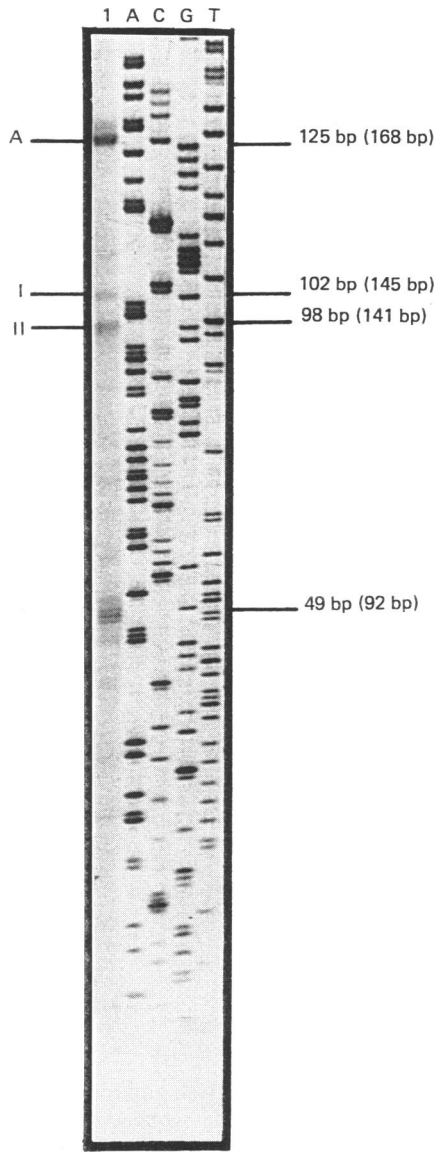

Fig. 4. Transcript mapping of the *E. coli aroL* gene

Tracks A, C, G and T represent the relevant sequencing tracks of bacteriophage M13 mp8 (see the Materials and methods section). Track 1 shows the reverse run-off transcript products using the oligonucleotide primer (2.5 ng) annealed to and extended along RNA (10 μg) prepared from *E. coli* HB101/pGM425. The full-length transcript is indicated as A; distances in bp upstream of the initiation codon ATG are given, and in parentheses the size of the nucleotide in the bacteriophage M13 mp8 sequence ladder is also shown for reference. The significance of bands I and II is discussed in the text.

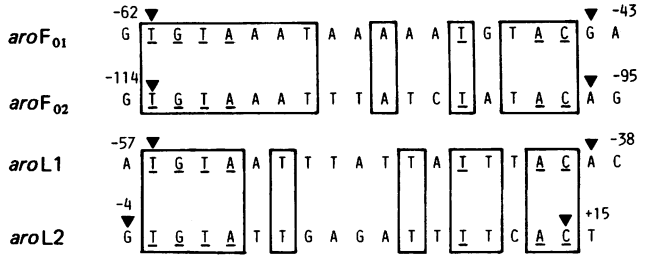

Fig. 5. Homologies of the *E. coli aroL*-gene upstream sequence with the control sequences of *aroF* gene

The *aro*F operator sequences implicated in *tyrR*-gene-product control (Garner & Herrmann, 1985) are shown and identical homologues boxed. Also shown are the L1 and L2 regions upstream of the *aroL* gene. Homologies between all four sequences when aligned are indicated by underlining. The limits of the 18 bp inverted repeat sequences are marked ▼. All the numbering is with respect to the transcriptional start of the respective genes.

DISCUSSION

The complete amino acid sequence of *E. coli* shikimate kinase II (the *aro*L-gene product) has been determined by a combined nucleotide and direct amino acid sequencing strategy. Construction of over-expressing strains has also permitted for the first time the purification to homogeneity of a monofunctional and monomeric shikimate kinase.

In the absence of available *aroL* mutants, complementation of a mutation in an adjacent marker (*proC*) on the *E. coli* chromosome was the basis for selection of potential *aroL* clones. The ability of these and subsequent sub-clones to over-express shikimate kinase activity was used as an additional selection criterion and facilitated the successful cloning and purification of the *aroL* gene and its gene product shikimate kinase II. The purified shikimate kinase is monomeric with an M_r of 20000; direct sequencing of the polypeptide chain revealed that the methionine residue encoded by the initiation codon ATG had been 'lost', presumably post-translationally. The M_r calculated from the predicted amino acid sequence of the enzyme is therefore 18937. This predicted M_r value and the observed M_r values agree well with the value of 20000 estimated previously by gel filtration of the partially purified enzyme (Ely & Pittard, 1979). It is noteworthy that three of the other pre-chorismate pathway enzymes in *E. coli* are also monomeric: 3-dehydroquinate synthase (Frost *et al.*, 1984), shikimate dehydrogenase (Chaudhuri & Coggins, 1985) and 5-enolpyruvylshikimate-3-phosphate synthase (Lewendon & Coggins, 1983).

Walker *et al.* (1982) and Finch & Emmerson (1984) have identified two highly conserved regions (A and B) in kinases and other ATP/ADP-requiring enzymes that are believed to contribute to the nucleotide-binding fold. Consensus sequences have been deduced for sequences A and B from several such proteins, and shikimate kinase exhibits at least one of these structural motifs. Sequence A is characterized by the consensus *G*-X_4-*G*-*K*-*T*-X_6-*I*/*V*; shikimate kinase contains the sequence *G*-P-R-G-C-*G*-*K*-*T*-T-V-G-M-A-L-A at residues 8–22. In adenylate kinase, where the tertiary structure is known (Pai *et al.*, 1977), sequence A corresponds to residues 15–29 (Fig. 6). A β-sheet region is formed by residues 10–14 in adenylate kinase, and this is immediately followed by a loop (residues 15–22) containing the conserved lysine residue. A conspicuous feature of the shikimate kinase sequence is the occurrence near the *N*-terminus of a region that is strongly homologous to the adenine-nucleotide-binding loop of adenylate kinase (Fig. 6). Secondary-structure predictions by using the method of Chou & Fasman (1978) for the shikimate kinase sequence indicates that

Table 5. Purification of shikimate kinase from *E. coli* HW87/pMH423

The results presented are for a purification from 16 g of cells. Full details are given in the Materials and methods section.

Step	Stage	Vol. (ml)	Protein (mg/ml)	Activity (units/ml)	Total activity (units	Yield (%)	Specific activity (units/mg)	Purification (fold)
1	Crude extract	48	6.8	5.66	272	100	0.83	1
2	DEAE-Sephacel	47	0.7	5.99	281	103	8.56	10.3
3	Phenyl-Sepharose	50	0.14	3.80	190	70	27.1	31.6
4	Sephacryl S200	10.5	0.13	9.04	95	35	69.5	80.7

Table 6. Amino acid composition of *E. coli* shikimate kinase compared with the amino acid composition deduced for shikimate kinase from the *E. coli aroL* gene sequence

Samples were analysed in duplicate after hydrolysis of performic acid-oxidized-protein with 6 M-HCl at 105 °C for 24, 48, 72 and 96 h. The eight experimental values were simply averaged except where indicated in the footnotes. Abbreviation: N.D., not determined.

Amino acid	Relative amino acid composition based on Leu = 17 residues	Theoretical amino acid composition predicted from the DNA sequence
Asx	13.6	14
Thr*	13.2	14
Ser*	7.3	8
Glx	27.1	26
Pro	8.8	8
Gly	10.2	10
Ala	18.3	18
Cys†	3.2	3
Val	14.5	15
Met‡	3.3	4
Ile	11.6	12
Leu	17.0	17
Tyr	2.1	2
Phe	5.5	5
His	2.1	2
Lys	2.4	2
Arg	12.5	12
Trp	N.D.	2

*Experimental values were extrapolated to zero time.
†Determined as cysteic acid.
‡Determined as methionine sulphone.

residues 2–7 have a high β-sheet probability and that the consensus sequence A homologue, residues 8–22, forms a turn or loop structure containing the conserved lysine residue. The occurrence of this binding fold very close to the *N*-terminus is not unique to pig adenylate kinase or *E. coli* shikimate kinase; in bacteriophage T_4 polynucleotide kinase an almost identical situation is observed (Midgley & Murray, 1985). The second conserved region in kinases (region B), which has the consensus sequence *R*/*K*-X_3-*G*-X_3-*L*-(hydrophobic)$_4$- followed by an aspartic acid residue, cannot be unambiguously located within the shikimate kinase sequence. One possible location is residues 138 (arginine) to 152 (aspartic acid).

In *E. coli* regulation of the carbon flow through the shikimate pathway occurs mainly at the first enzyme 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase (EC 4.1.2.15) (Gibson & Pittard, 1968; Camakaris & Pittard, 1983). Three 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase isoenzymes, products of the unlinked *aroF*, *aroG* and *aroH* genes, are each inhibited by tyrosine, phenylalanine and tryptophan respectively. Further regulation includes repression of the *aro*F and *aroG* genes by the *tyrR*-gene-product protein complexed with tyrosine and phenylalanine respectively (Camakaris & Pittard, 1983); *aroH* is repressed by the *trpR*-gene product (Pittard *et al.*, 1969). Regulation of activities subsequent to the first committed step is not a common feature of biosynthetic pathways. However, in both *E. coli* (Ely & Pittard, 1979) and *B. subtilis* (Huang *et al.*, 1975) regulation of aromatic biosynthesis also includes regulation of shikimate kinase activity. Expression of the structural gene of one of the *E. coli* shikimate kinase isoenzymes, the *aroL*-gene product, is regulated by the *tyrR*-gene-product protein with tyrosine or tryptophan as co-repressor (Ely & Pittard, 1979).

Transcript mapping has located the 5′ end of the *aroL*-gene mRNA and identified promoter sequences upstream. The existence of two inverted repeat sequences

Protein	Residues	Sequences	Reference
Adenylate kinase	8-29	S K I I F V V G G P G S G K G T Q C E K I V	(*a*)
*rec*A protein	59-80	G R I V E I Y G P E S S G K T T L T L Q V I	(*b*)
rho protein	172-193	G Q R G L I V A P P K A G K T M L L Q N I A	(*c*)
DNA B protein	223-244	S D L I I V A A R P S M G K T T F A M N L V	(*d*)
EF-Tu	11-32	H V N V G T I G H V D H G K T T L T A A I T	(*e*)
DNA helicase II	22-43	R S N L L V L A G A G S G K T R V L V H R I	(*f*)
Shikimate kinase	1-22	T Q P L F L I G P R G C G K T T V G M A L A	(*g*)

Fig. 6. Homologies between adenine-nucleotide-binding proteins and *E. coli* shikimate kinase II

All proteins are from *E. coli* except for adenylate kinase (pig). Strongly conserved regions are boxed. References: (*a*) Pai *et al.* (1977); (*b*) Horii *et al.* (1980); (*c*) Pinkham & Platt (1983); (*d*) Nakayama *et al.* (1984); (*e*) Zengel *et al.* (1984); (*f*) Finch & Emmerson (1984); (*g*) present work.

in the region of the *aroL* promoter suggests the existence of a potential *tyrR*-gene-product-binding region. The 3′ end of the L1 sequence overlaps the 5′ end of the −35 region, and the L2 sequence overlaps the transcription start site. Both sequences (L1 and L2) exhibit homology with the proposed regulatory *tyrR*-gene-product-binding site, which is found upstream of the *aroF* coding sequence (Garner & Herrmann, 1985) (Fig. 5). In both *aroF* and *aroL* genes the two 18 bp inverted repeats are separated by 51 or 52 bp; in both cases the proposed regulatory-protein-binding site is within, or overlaps, the promoter sequences of the respective genes.

Expression of the *aroL*-containing plasmid pGM424 in a tyrR background resulted in only a 50% elevation of shikimate kinase II activity compared with the over-expressed activity (pGM424) in the 'wild-type' background (*E. coli* HB101). De-repression would therefore appear to be confined to a subpopulation of the plasmid molecules, implying that the high-copy-number vector is titrating out the *tyrR*-gene-product repressor.

Note added in proof (received 7 May 1986)

Since this work was accepted for publication the cloning and sequencing of the *E. coli aroL* gene have been reported by another group (see Defeyter & Pittard, 1986).

It is a pleasure to acknowledge the helpful and constructive advice given to us by Dr. K. Duncan and Dr. R. M. Edwards. We thank Proessor J. E. Fothergill, Dr. L. A. Fothergill and Mr. B. Dunbar (University of Aberdeen Protein Sequencing Facility) for help with the amino acid sequencing and Mr. J. Jardine for the amino acid analysis. We also thank the University of Glasgow for a studentship (to G.M.) and the Science and Engineering Research Council for financial support.

REFERENCES

Aiba, H., Adhya, S. & deCrombrugghe, B. (1981) J. Biol. Chem. **256**, 11905–11910
Bachmann, B. (1983) Microbiol. Rev. **44**, 180–230
Berlyn, M. B. & Giles, N. H. (1969) J. Bacteriol. **99**, 222–230
Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3963–3965
Bode, R. & Birnbaum, D. (1981) Z. Allg. Mikrobiol. **21**, 417–422
Bolivar, F. & Backman, K. (1979) Methods Enzymol. **68**, 245–267
Bradford, M. M. (1976) Anal. Biochem. **72**, 248–254
Camakaris, H. & Pittard, J. (1983) in Amino Acids: Biosynthesis and Genetic Regulation (Herrmann, K. & Somerville, R. L., eds.), pp 339–350, Addison Wesley, London
Carter, P. E., Dunbar, B. & Fothergill, J. E. (1983) Biochem. J. **215**, 565–571
Chaudhuri, J. R. & Coggins, J. R. (1985) Biochem. J. **226**, 217–223
Chaudhuri, S., Lambert, J. M., McColl, L. A. & Coggins, J. R. (1986) Biochem. J., in the press
Chou, P. Y. & Fasman, G. D. (1978) Annu. Rev. Biochem. **47**, 251–277
Daghert, M. & Ehrlich, S. D. (1979) Gene **6**, 23–28
Defeyter, R. C. & Pittard, J. (1986) J. Bacteriol. **165**, 226–232
Deutch, A. H., Smith, C. J., Rushlow, K. E. & Kretschmer, P. J. (1982) Nucleic Acids Res. **10**, 7701–7714
Duncan, K., Lewendon, A. & Coggins, J. R. (1984) FEBS Lett. **165**, 121–127
Ely, B. & Pittard, J. (1979) J. Bacteriol. **138**, 933–943
Finch, P. W. & Emmerson, P. T. (1984) Nucleic Acids Res. **12**, 5789–5798
Frost, J. W., Bender, J. L., Kadonaga, J. T. & Knowles, J. R. (1984) Biochemistry **23**, 4470–4475
Gaertner, F. H. & Cole, K. W. (1977) Biochem. Biophys. Res. Commun. **75**, 259–264
Garner, C. C. & Herrmann, K. M. (1985) J. Biol. Chem. **260**, 3820–3825
Gibson, F. & Pittard, J. (1968) Bacteriol. Rev. **32**, 465–492
Gicquel-Sanzey, B. & Cossart, P. (1982) EMBO J. **1**, 591–595
Hawley, C. J. & McClure, W. R. (1983) Nucleic Acids Res. **11**, 2237–2255
Hirs, C. H. W. (1967) Methods Enzymol. **11**, 197–199
Holmes, D. S. & Quigley, M. (1981) Anal. Biochem. **114**, 193–197
Horii, T. Ogawa, T. & Ogawa, H. (1980) Proc. Natl. Acad. Sci. U.S.A. **80**, 4253–4257
Huang, L., Montoya, A. L. & Nester, E. W. (1975) J. Biol. Chem. **250**, 7675–7681

Kinghorn, J. R. & Hawkins, A. R. (1982) Mol. Gen. Genet. **186**, 145–152
Lambert, J. M., Boocock, M. R. & Coggins, J. R. (1985) Biochem. J. **226**, 817–829
Larimer, F. W., Morse, C. C., Beck, A. K., Cole, K. W. & Gaertner, F. H. (1983) Mol. Cell. Biol. **3**, 1609–1614
Lewendon, A. & Coggins, J. R. (1983) Biochem. J. **213**, 187–191
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, J. (1982) Molecular Cloning, Cold Spring Harbor, New York
Messing, J. & Vieira, J. (1982) Gene **19**, 269–276
Messing, J., Crea, R. & Seeburg, P. H. (1981) Nucleic Acids Res. **9**, 309–321
Midgley, C. A. & Murray, N. E. (1985) EMBO J. **4**, 2695–2703
Morell, H. & Sprinson, D. B. (1968) J. Biol. Chem. **243**, 676–682
Nakanishi, H. & Yamamoto, M. (1984) Mol. Gen. Genet. **195**, 164–169
Nakatsukasa, W. M. & Nester, E. W. (1972) J. Biol. Chem. **247**, 5972–5979
Nakayama, M., Arai, N., Bond, M. W., Kaziro, Y. & Arai, K. (1984) J. Biol. Chem. **259**, 97–101
Nasser, D., Henderson, G. & Nester, E. W. (1969) J. Bacteriol. **98**, 44–50
Pai, E. F., Sachsenheimer, W., Schirmer, H. & Schultz, G. E. (1977) J. Mol. Biol. **114**, 37–45
Pinkham, J. L. & Platt, T. (1983) Nucleic Acids Res. **11**, 3531–3545
Pittard, J., Camakaris, H. & Wallace, B. J. (1969) J. Bacteriol. **97**, 1242–1247
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) 34–38
Smith, D. D. S. & Coggins, J. R. (1983) Biochem J. **213**, 405–415
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
Staden, R. (1984) Nucleic Acids Res. **12**, 505–509
Strauss, A. (1979) Mol. Gen. Genet. **172**, 233–242
Twigg, A. J. & Sherratt, D. (1980) Nature (London) **283**, 216–218
Walker, J. E., Saraste, M., Runswick, M. J. & Gray, M. J. (1982) EMBO J. **1**, 945–951
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
Zengel, J. M., Archer, R. H. & Lindahl, L. (1984) Nucleic Acids Res. **12**, 2181–2192

Received 29 November 1985/27 January 1986; accepted 17 March 1986