Biochem. J. (1986) 234, 49–57 (Printed in Great Britain)

The *serC–aroA* operon of *Escherichia coli*

A mixed function operon encoding enzymes from two different amino acid biosynthetic pathways

Kenneth DUNCAN and John R. COGGINS*
Department of Biochemistry, University of Glasgow, Glasgow G12 8QQ, Scotland, U.K.

Sub-cloning experiments aimed at precisely locating the *E. coli aroA* gene, which encodes the shikimate pathway enzyme 5-enolpyruvylshikimate 3-phosphate synthase, showed that in certain constructions, which remain capable of complementing an auxotrophic *aroA* mutation, expression of *aroA* is reduced. DNA sequence analysis revealed that a sequence approx. 1200 base pairs (bp) upstream of *aroA* is necessary for its expression. An open reading frame was identified in this region which encodes a protein of 362 amino acids with a calculated M_r of 39834 and which ends 70 bp before the start of the *aroA* coding sequence. This gene has been identified as *serC*, the structural gene for 3-phosphoserine aminotransferase, an enzyme of the serine biosynthetic pathway. Both genes are expressed as a polycistronic message which is transcribed from a promoter located 58 bp upstream of *serC*. Evidence is presented which confirms that the *aroA* and *serC* genes constitute an operon which has the novel feature of encoding enzymes from two different amino acid biosynthetic pathways.

INTRODUCTION

The *Escherichia coli aroA* gene encodes the enzyme 5-enolpyruvylshikimate 3-phosphate synthase (EPSP synthase; EC 2.5.1.19; alternative name: 3-phosphoshikimate 1-carboxyvinyltransferase) which is the sixth enzyme on the seven-step early common pathway for aromatic amino acid biosynthesis (the pre-chorismate pathway). This enzyme has recently attracted considerable attention both because of its mechanism (Grimshaw *et al.*, 1982) and because it is the site of inhibition in plants of the commercially successful herbicide glyphosate (Steinrucken & Amrhein, 1980, 1984; Boocock & Coggins, 1983; Anton *et al.*, 1983; Mousdale & Coggins, 1984).

The gene is located at 20 min on the linkage map of the *E. coli* chromosome, between *pdxC* and *himD* (Bachmann, 1983). Mutations in *aroA* have been mapped and correlated with the loss of EPSP synthase activity. It has also been shown that the genes coding for the pre-chorismate pathway enzymes are widely scattered about the genome (Pittard & Wallace, 1966). Subsequent studies showed that in *E. coli* the pre-chorismate pathway enzyme activities are separable (Berlyn & Giles, 1969), whereas in the fungus *Neurospora crassa*, and in yeast, the enzymes catalysing the five central steps of the pathway are organised on a single pentafunctional polypeptide chain, the *arom* enzyme complex (Giles *et al.*, 1967; Lumsden & Coggins, 1977; Gaertner & Cole, 1977; Larimer *et al.*, 1983; Lambert *et al.*, 1985). Thus EPSP synthase exists in two distinct forms either as a monofunctional enzyme, for example in *E. coli* and plants (Mousdale & Coggins, 1984) or as a functional domain of a multifunctional enzyme as in *Neurospora* (Coggins *et al.*, 1985).

The monofunctional form of EPSP synthase has been purified to homogeneity from *E. coli* K12 (Lewendon & Coggins, 1983) and in milligram quantities from an overproducing strain of *E. coli* (Duncan *et al.*, 1984*a*). It has also been purified from pea seedlings (Mousdale & Coggins, 1984) and from *Klebsiella pneumoniae* (Anton *et al.*, 1983; Steinrucken & Amrhein, 1984). The *aroA* gene has been cloned from *E. coli* (Rogers *et al.*, 1983; Duncan & Coggins, 1984; Duncan *et al.*, 1984*a*) and from *S. typhimurium* (Comai *et al.*, 1983). The source of the *aroA* gene cloned in our laboratory was the specialized transducing phage λp*serC* (Kitikawa *et al.*, 1980; Schnier & Isono, 1982). A *Pst*I fragment of λp*serC* was cloned into pAT153 (Twigg & Sherratt, 1980) and it was shown that *E. coli* cells harbouring the resulting recombinant plasmid, pKD501, overexpress EPSP synthase approx. 100-fold compared with *E. coli* K12 (Duncan *et al.*, 1984*a*). This level of overexpression was consistent with the expected copy number of pAT153 (approx. 100) and the fact that *aroA* is expressed constitutively (Tribe *et al.*, 1976).

Sub-cloning experiments, aimed at locating *aroA* on the *Pst*I fragment prior to DNA sequence analysis, showed that expression of *aroA* was affected by cloning from a *Cla*I site in the phage. Subsequent DNA and protein sequencing (Duncan *et al.*, 1984*b*) showed that the startpoint of the *aroA* coding region was located 700 bp away from this site. The present paper describes the precise location of the *aroA* gene and the analysis of the DNA sequence flanking the 5′ side of the gene. The results reveal that *aroA* is co-transcribed as a polycistronic mRNA with a neighbouring gene. This gene has been identified as *serC*, which encodes the enzyme 3-phosphoserine aminotransferase (PSAT; EC 2.6.1.52).

Abbreviations used: kb, kilobase pairs; bp, base pairs; ORF, open reading frame; EPSP, 5-enolpyruvylshikimate 3-phosphate; PSAT, 3-phosphoserine aminotransferase.

* To whom reprint requests should be sent.

MATERIALS AND METHODS

Sub-cloning

All restriction enzymes were purchased from Bethesda Research Laboratories, and digestions were performed under the conditions recommended by the manufacturer. T4 DNA ligase was purchased from Amersham International; calf-intestinal phosphatase was from Boehringer. Phage and plasmid DNAs were prepared as described in Maniatis *et al.* (1982). The isolation of particular DNA fragments for sub-cloning after electrophoresis through low-melting-temperature agarose (BRL) was carried out as described in Maniatis *et al.* (1982). Bacterial transformation was by the method of Dagert & Ehrlich (1979).

Cloning in M13

The paired vectors M13mp8 and M13mp9 (Messing & Vieira, 1982) were used. All manipulations were carried out as described in the "M13 cloning and sequencing handbook" (Amersham International).

DNA sequence analysis

All the components for DNA sequence analysis were purchased in the form of kits from Amersham International; sequencing reactions were carried out as described in the "M13 cloning and sequencing handbook" (Amersham). [α-^{35}S]dATPαS (Amersham SJ304) was used throughout as label. Electrophoresis was on 6% polyacrylamide/8 M-urea linear or buffer gradient gels (0.4 cm × 20 cm × 40 cm).

Polyacrylamide-gel electrophoresis in the presence of SDS

Polyacrylamide-gel electrophoresis in the presence of SDS was performed by the method of Laemmli (1970) using a 12% running gel, with a 3% stacking gel.

Enzyme assays

Crude extracts were prepared from 100 ml stationary-phase cultures grown in liquid minimal medium. Cells were harvested, washed in 50 ml of 200 mM-Tris/HCl (pH 7.5)/200 mM-KCl/1 mM-EDTA/1 mM-dithiothreitol and resuspended in 5 ml of the same buffer. The cells were then disrupted by sonication and extracts were centrifuged (200000 *g*, 4 °C, 2 h) to remove debris and membrane-bound enzymes.

EPSP synthase. This was assayed in the reverse direction at 25 °C as described by Lewendon & Coggins (1983).

PSAT. The assay for PSAT activity is based on that described in Hirsch-Kolb & Greenberg (1971), and was modified by Dr. A. Lewendon. The formation of 2-oxoglutarate from 3-hydroxypyruvate was coupled to the glutamate dehydrogenase reaction. Assay mixes (1 ml) contained: 50 mM-Tris/HCl (pH 8.2), 32 mM-ammonium acetate, 2 mM-glutamate, 0.2 mM-NADH, 2.5 mM-3-hydroxypyruvate and 0.02 mg (2 units) of glutamate dehydrogenase (Boehringer). The reaction at 25 °C was initiated by the addition of 3-hydroxypyruvate and the absorbance change at 340 nm was measured.

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

Protein concentration was determined by the method of Bradford (1976).

S1 nuclease mapping

RNA was prepared from *E. coli* AB2829/pKD501 as follows. L broth (100 ml) containing tetracycline (20 μg/ml) was inoculated with *E. coli* AB2829/pKD501 and growth continued till $A_{650} = 0.3$. The cells were harvested and resuspended in 6 ml of 0.02 M-sodium acetate (pH 5.5)/0.5% SDS/1 mM-EDTA. Phenol (12 ml) was added and the tube was incubated at 60 °C for 10 min with gentle shaking. The sample was centrifuged and the aqueous phase re-extracted with phenol. The aqueous phase was again separated and the RNA recovered by ethanol precipitation. Following resuspension in 3 ml of the same buffer the ethanol precipitation/resuspension was repeated twice.

Hybridization and S1 nuclease digestion. RNA (50 μg) and probe DNA (1 μg) were co-precipitated by the addition of ethanol and resuspended in 50 μl of 80% (v/v) formamide (deionized)/0.02 M-Pipes (pH 6.4)/0.6 M-NaCl/4 mM-EDTA. The mixture was incubated at 55 °C for 3 h. Then 300 μl of ice-cold S1 buffer [30 mM-sodium acetate (pH 4.6)/50 mM-NaCl/1 mM-$ZnSO_4$/5% (v/v) glycerol] containing 2000 units of S1 nuclease (BRL) was added and incubation was continued at 43 °C for 45 min. Digestion products were extracted with phenol/chloroform and ethanol-precipitated before analysis on polyacrylamide gels.

RESULTS

Sub-cloning of the *aroA* gene

The 4.6 kb *Pst*I fragment, inserted in pKD501 (Duncan *et al.*, 1984*a*) is considerably larger than the size required to encode EPSP synthase, the *aroA* gene product. The restriction map of the genomic insert of λp*serC* (Schnier & Isono, 1982) was examined for restriction sites which would allow parts of the *Pst*I fragment to be cloned. Phage λp*serC* DNA was digested with three combinations of enzymes and the DNA fragments obtained were cloned into pAT153. *aroA* recombinant plasmids were isolated and characterized as described in Duncan *et al.* (1984*a*) (Fig. 1). These constructions (pKD503, pKD504 and pKD505) placed the boundary of the *aroA* gene between the *Cla*I site and the *Pst*I site on the phage. A 1.9 kb *Cla*I–*Pvu*II fragment from pKD501 was inserted into pBR322 and it was shown that the resulting recombinant (pKD506) is also capable of complementing the mutation in *E. coli* AB2829.

The specific activity of EPSP synthase in crude extracts of cells carrying each of these plasmids was determined (Table 1). Two of the plasmids, pKD503 and pKD504, overexpress EPSP synthase to the same extent as the original plasmid pKD501 and the related plasmid pKD502, which carries the same insert as pKD501 but in the opposite orientation. Unexpectedly, however, pKD505 and pKD506 show only wild-type enzyme levels. There are therefore two classes of sub-clone: 'high-activity' (pKD501, pKD502, pKD503 and pKD504) and 'low-activity' (pKD505 and pKD506); the level of EPSP synthase in the low-activity plasmid-carrying cells is adequate to complement the *aroA* mutation of *E. coli* AB2829, and so all the plasmids are phenotypically aroA$^+$.

These results indicate that cloning from the *Cla*I site disrupts expression of the *aroA* gene. Since the *Cla*I site

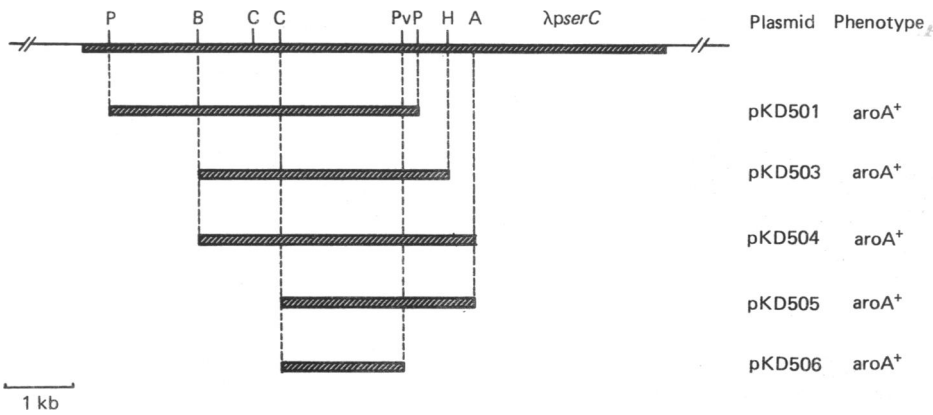

Fig. 1. Sub-cloning strategy for the *aroA* gene

pKD501–pKD505 were constructed by inserting fragments of λp*serC* DNA into pAT153; pKD506 is a derivative of pKD501; the *Cla*I–*Pvu*II fragment was inserted into pBR322. pKD502 (not shown) carries the same *Pst*I fragment as pKD501 but in the opposite orientation (Duncan *et al.*, 1984*a*). Only the relevant restriction enzyme sites (after Schnier & Isono, 1982) are shown. Key: P, *Pst*I; B, *Bgl*II; C, *Cla*I; Pv, *Pvu*II; H, *Hin*dIII; A, *Ava*I.

Table 1. Specific activity of EPSP synthase and PSAT in crude extracts of various *E. coli* strains, and in *E. coli* AB2829, transformed with a number of different recombinant plasmids

Values quoted are an average of duplicate results; n.d. indicates not determined.

Extract	EPSP synthase		PSAT	
	Specific activity (units/mg)	Relative activity	Specific activity (units/mg)	Relative activity
E. coli K12 (A.T.C.C. 14948)	0.004	1	0.011	1
E. coli AB2829	No activity		n.d.	
E. coli KL282	n.d.		No activity	
E. coli AB2829/pKD501	0.420	105	1.06	96
E. coli AB2829/pKD502	0.425	106	n.d.	
E. coli AB2829/pKD503	0.425	106	n.d.	
E. coli AB2829/pKD504	0.358	89	n.d.	
E. coli AB2829/pKD505	0.0044	1.1	n.d.	
E. coli AB2829/pKD506	0.0048	1.2	0.010	1

is outside the EPSP synthase coding sequence (Duncan *et al.*, 1984*b*) the low activity cannot be due to the production of incomplete polypeptide chain. A more likely explanation is that cloning from the *Cla*I site has disrupted the promoter of the *aroA* gene and that the observed low level of expression is from a plasmid promoter.

To investigate this further, the proteins present in crude extracts of the various plasmid-transformed strains were examined by SDS/polyacrylamide-gel electrophoresis. It was expected that a Coomassie Blue staining band corresponding to the overproduced EPSP synthase would be readily observed in the 'high-activity' crude extracts. Unfortunately, in wild type *E. coli* there is an abundant protein which co-migrates with EPSP synthase. This has been identified as elongation factor Tu which has an M_r of 43225 (Arai *et al.*, 1979) and under these conditions it is not separated from EPSP synthase which has an M_r of 46112 (Duncan *et al.*, 1984*b*). Thus it was impossible to observe directly changes in the amounts of EPSP synthase polypeptide produced by the various plasmid-carrying strains. However examination of the polypeptide chain patterns of high- and low-activity crude extracts (Fig. 2) revealed that the high-activity extracts always contained a second abundant protein of M_r 40000 which was absent from the low-activity extracts.

In order to establish that the *aroA* promoter is upstream of the *Cla*I site, the 0.5 kb *Cla*I fragment to the left of the *Cla*I–*Pvu*II region was cloned into pKD506. On transformation of *E. coli* AB2829 all the plasmid-carrying cells would be capable of growth on minimal medium and so a different selection system was required to identify high-activity clones specifically. It has been shown that pKD501-carrying cells are tolerant of high levels of the herbicide glyphosate (*N*-phosphonomethylglycine) in the medium (up to 50 mM) whereas growth of *E. coli* K12 is inhibited by 3 mM-glyphosate (Duncan *et al.*, 1984*a*). High-activity clones were therefore selected by plating recombinants on minimal agar containing glyphosate (20 mM). At least 800 colonies were screened

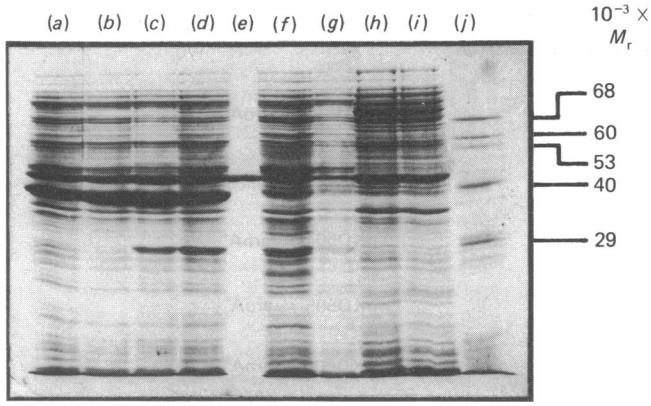

Fig. 2. SDS/polyacrylamide-gel electrophoresis of *E. coli* crude extracts

Track (*a*), *E. coli* AB2829/pKD501; track (*b*), *E. coli* AB2829/pKD502; track (*c*), *E. coli* AB2829/pKD503; track (*d*), *E. coli* AB2829/pKD504; track (*e*) EPSP synthase (*E. coli* enzyme purified by Dr. A. Lewendon; see Lewendon & Coggins, 1983); track (*f*), *E. coli* AB2829/pKD505; track (*g*), *E. coli* AB2829/pKD506; track (*h*), *E. coli* AB2829; track (*i*), *E. coli* K12; track (*j*), standards: bovine serum albumin (M_r 68000), catalase (M_r 60000), glutamate dehydrogenase (M_r 53000), aldolase (M_r 40000), carbonic anhydrase (M_r 29000).

in this way, but none were glyphosate-tolerant. Plasmid DNA was isolated from 20 transformants and digested with *Cla*I. All had the 0.5 kb *Cla*I insert; it was not possible to show directly that plasmids with the inserts in both of the possible orientations had been obtained since the only known restriction enzyme cleavage site is in the middle of the fragment. However it is exceedingly unlikely that all the inserts were in the 'wrong' orientation, which leads to the conclusion that inserting the *Cla*I fragment has no effect on *aroA* expression and indicates that the *aroA* promoter is still further upstream.

DNA sequence analysis

DNA sequence analysis was first concentrated on the 1.9 kb *Cla*I–*Pvu*II fragment cloned in pKD506. The DNA sequence of this region was determined using the M13/dideoxy chain-termination method (Sanger, 1981; Messing, 1983). The coding region of the *aroA* gene was identified by comparison of the translated DNA sequence with the *N*-terminal protein sequence of EPSP synthase. The start point of the *aroA* coding region is an ATG codon approx. 700 bp to the right of the *Cla*I site and the gene runs left to right, i.e. away from the *Cla*I site. This work and the *aroA* coding sequence are described in Duncan *et al.* (1984*b*).

The 700 bp DNA sequence between the *Cla*I site and the *aroA* startpoint was examined for a possible promoter sequence, similar to the consensus sequence reported in Hawley & McClure (1983), but none was found. Translation of the sequence revealed an open reading frame (ORF) which extends from the *Cla*I site to a point 70 bp before the *aroA* coding region startpoint. This, together with the observation that insertion of the *Cla*I fragment to the left of the *Cla*I site has no effect on *aroA* expression, suggested that *aroA* was part of an operon and was co-ordinately expressed with the 40000-M_r protein from some promoter upstream of the ORF.

The DNA sequences of the *Cla*I fragment and of a *Hpa*II fragment which extends 286 bp to the left of this fragment were determined. The sequencing strategy employed is summarized in Fig. 3, and the DNA sequence obtained is shown in Fig. 4. Translation of this sequence shows that there is an ORF which begins at position 326 in the sequence and is continuous with the ORF upstream of *aroA*. The first ATG/Met codon in the ORF is at position 332 and is preceded by the sequence GAGG, which is a possible ribosome binding site (Shine & Dalgarno, 1975) (Fig. 4). The ORF encodes a 362 amino acid protein, with a calculated M_r of 39834, which correlates well with the M_r 40000 band observed on polyacrylamide-gel electrophoresis of the crude extract (Fig. 2). The codon utilization of the gene is shown in Table 2, along with that of the *aroA* gene. The overall

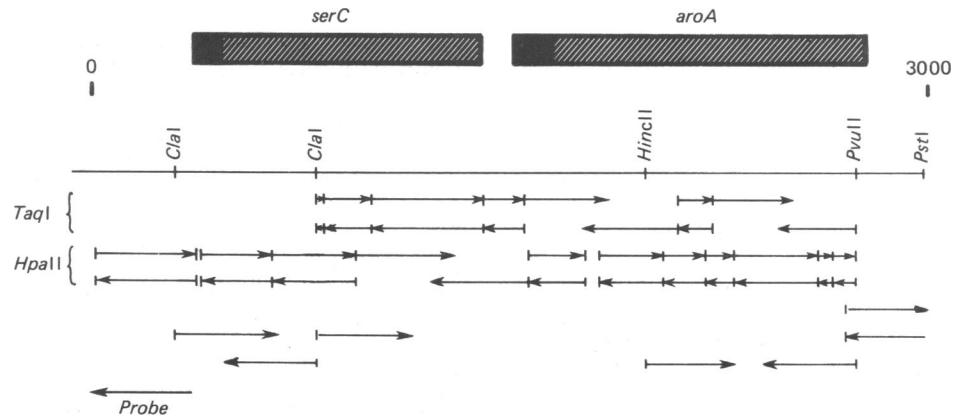

Fig. 3. Sequencing strategy for the entire *serC–aroA* region of the *E. coli* chromosome

Only the relevant restriction enzyme sites are shown. Arrows indicate direction and extent of sequences. The arrow labelled 'Probe' indicates the DNA fragment used to locate the *serC* promoter by S1 nuclease mapping. Left to right on this diagram corresponds to clockwise on the *E. coli* linkage map.

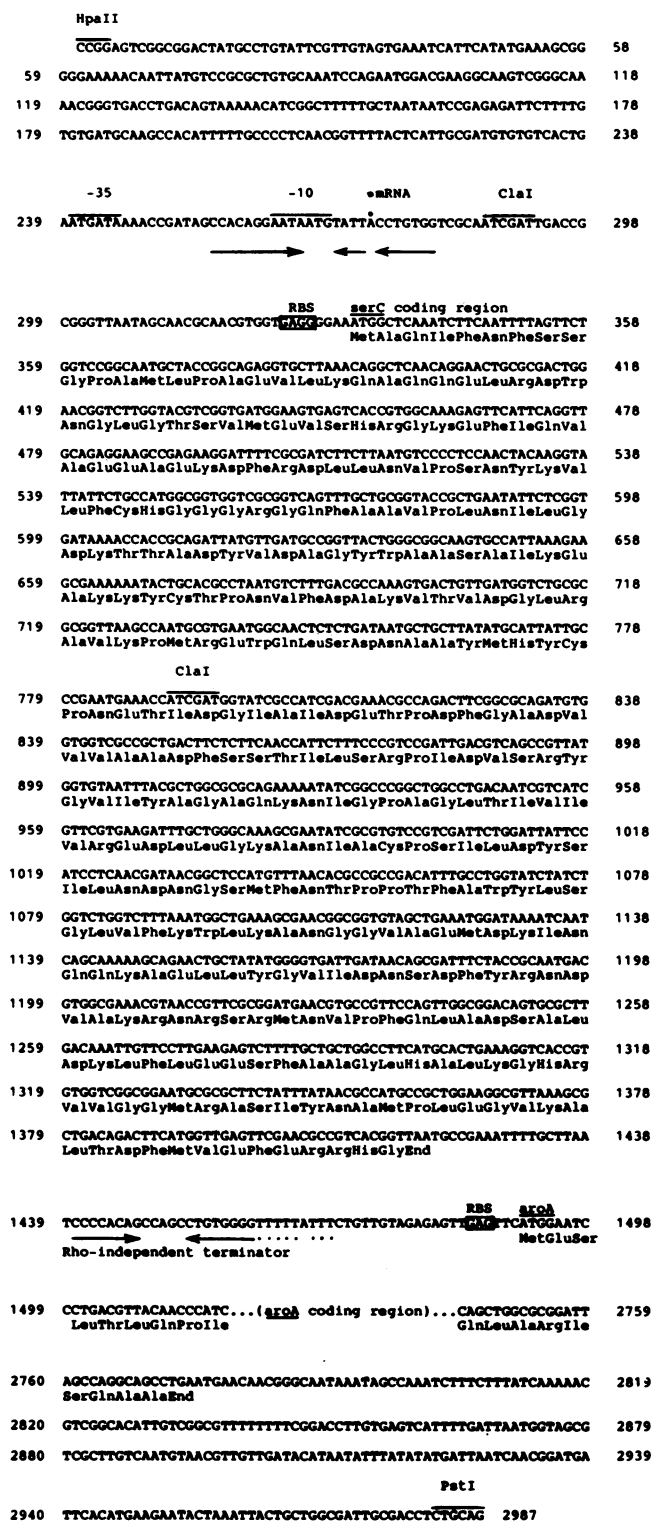

Fig. 4. Nucleotide sequence of the *serC–aroA* operon (excluding the *aroA* coding region, published in Duncan *et al.*, 1984*b*)

The *serC* structural gene and amino acid sequence predicted for PSAT are shown (positions 332–1420). The *serC* promoter elements (−35 and −10 regions) and the transcriptional startpoint are indicated. Potential ribosome binding sites (RBS) are also boxed, and regions of dyad symmetry are shown as arrows below the sequence.

pattern in the table is similar, and is similar to that found for *E. coli* genes in the moderately to weakly expressed category (Grosjean & Fiers, 1982; Gouy & Gautier, 1982).

Identification of the product of the ORF

The cloning and DNA sequencing carried out in this work, along with the data of Schnier & Isono (1982), have located the *aroA* gene with respect to one of its flanking markers, the *rpsA* gene (Bachmann, 1983). In the opposite direction, the most closely linked gene is *pdxC*, which is defined in Bachmann (1983) as a 'requirement for pyridoxine'; the nature of the mutation is unknown. Little is known about the biosynthesis of pyridoxine (vitamin B_6) in micro-organisms and it was assumed that the *pdxC* mutation was in the structural gene for a pyridoxine biosynthetic enzyme.

There is, however, evidence suggesting that the mutation which defines *pdxC* is an allele of the *serC* gene, which maps next to *pdxC*. The serine biosynthetic pathway converts 3-phosphoglycerate to serine in three steps. The enzymes catalysing the steps are encoded in the *serA*, *serB* and *serC* genes (Bachmann, 1983). The *serC* genotype, a mutation in the gene which encodes PSAT, was first described in *E. coli* B (Dempsey, 1969) and subsequently in *E. coli* K12 (Clarke *et al.*, 1973), where it was shown to be closely linked to *aroA*. *serA* and *serB* mutants require only serine for growth, but *serC* mutants display a double requirement for serine and pyridoxine.

Shimizu & Dempsey (1978) showed that *E. coli* K12 *serC* mutants revert easily to pyridoxine independence without regaining PSAT activity. When 3-hydroxypyruvate is used as a supplement, both the revertants and the parental types synthesize pyridoxine in normal amounts, but neither can use this compound to satisfy their serine requirement. As serine alone is inadequate to provide the nutritional requirement of *serC* mutants, these mutants must be unable to synthesise 3-hydroxypyruvate from serine. This suggests that, in normal *E. coli*, PSAT serves as a catalyst for transaminating small amounts of serine to 3-hydroxypyruvate, which is then used in pyridoxine biosynthesis. Transductional analysis was used to locate the secondary mutations in the revertant strains. Some revertant mutations were shown to be so closely linked to *aroA* that recombination between them and *aroA* is extremely unlikely. These observations suggested that the product of the ORF may be PSAT.

The identification of the gene as *serC* was carried out as follows. Firstly *E. coli* KL282 (serC) transformed with a number of plasmids was tested for its ability to grow on minimal medium (Table 3). Only those plasmids which carried the entire ORF were capable of relieving the auxotrophic mutation in *E. coli* KL282. Secondly, crude extracts of *E. coli* K12 and *E. coli* AB2829 transformed with various recombinant plasmids were prepared. The specific activity of PSAT was determined (Table 1); cells harbouring plasmids encoding the entire ORF overexpress PSAT approx. 100-fold.

PSAT has been purified to homogeneity in our laboratory. The native enzyme is a dimer of M_r 79000 and subunit M_r 39000. The amino acid composition of the enzyme was determined; this closely resembled the enzyme composition predicted from the DNA sequence. The *N*-terminal protein sequence confirmed that the *serC* gene coding region begins at the predicted position, except that the *N*-terminal methionine is cleaved from the

Table 2. Codon utilization in the *aroA* and *serC* genes

	T	*aroA*	*serC*	C	*aroA*	*serC*	A	*aroA*	*serC*	G	*aroA*	*serC*	
T	Phe	9	8	Ser	6	6	Tyr	9	8	Cys	2	1	T
	Phe	9	10	Ser	5	4	Tyr	4	5	Cys	4	3	C
	Leu	10	1	Ser	2	1	Term	—	—	Term	—	—	A
	Leu	5	3	Ser	2	3	Term	—	—	Trp	2	5	G
C	Leu	4	8	Pro	4	1	His	3	3	Arg	12	9	T
	Leu	3	3	Pro	3	1	His	5	3	Arg	7	7	C
	Leu	0	3	Pro	2	2	Gln	5	4	Arg	0	0	A
	Leu	26	12	Pro	9	11	Gln	7	7	Arg	2	1	G
A	Ile	17	9	Thr	7	1	Asn	9	10	Ser	1	3	T
	Ile	9	9	Thr	10	1	Asn	9	10	Ser	5	3	C
	Ile	0	0	Thr	7	3	Lys	14	17	Arg	0	0	A
	Met	14	11	Thr	10	4	Lys	3	3	Arg	1	0	G
G	Val	7	7	Ala	6	13	Asp	23	15	Gly	13	14	T
	Val	4	7	Ala	7	8	Asp	3	10	Gly	18	12	C
	Val	4	4	Ala	15	8	Glu	16	13	Gly	2	1	A
	Val	9	10	Ala	18	12	Glu	6	6	Gly	4	1	G

Table 3. Growth of *E. coli* KL282, untransformed and transformed with a number of recombinant plasmids

1, growth; 0, no growth.

Plasmid	Growth on: Minimal medium	Minimal medium +serine (20 μg/ml) +pyridoxine (75 μg/l)
pAT153	0	1
pKD501	1	1
pKD506	0	1
Untransformed	0	1

enzyme post-translation (A. Lewendon & J. R. Coggins, unpublished work).

Analysis of the DNA sequence, and identification of the *serC* promoter

Analysis of the 70 bp of DNA which separate the *serC* and *aroA* genes (Fig. 4) revealed a G+C-rich inverted repeat followed by a run of T residues, which is typical of an *E. coli* rho-independent transcription terminator (Rosenberg & Court, 1979). This structure may be important in attenuating *aroA* expression; the relative amounts of each gene product isolated from *E. coli* show that *serC* is expressed at a higher level than *aroA*. This suggests that two messages are produced from the same promoter upstream of *serC*, which is consistent with preliminary evidence from Northern blot analysis. *serC* may thus be expressed from either a monocistronic mRNA which is terminated at the termination sequence, or from a polycistronic mRNA which is transcribed by read-through at the terminator and which encodes both the *serC* and *aroA* genes.

In order to identify the promoter sequence for the operon, the S1 nuclease mapping technique was employed to determine the transcription startpoint. A 363 bp *Hpa*II fragment, cloned into M13mp8 (Fig. 3) was used as a probe for this work. The fragment was oriented in the clone in such a way that primed-synthesis of the second strand of single stranded template DNA using the Klenow fragment of DNA polymerase and a mixture of unlabelled dATP, dGTP and dTTP and (α-^{32}P]dCTP resulted in production of a probe which was labelled only on the strand complementary to the mRNA. The labelled *Hpa*II fragment was recovered by electrophoresis on low-melting-temperature agarose after cleaving the newly synthesized double stranded molecule with *Eco*RI and *Hin*dIII to cut the vector sequences flaking the insert. This *Hpa*II probe was then hybridized to RNA prepared from an exponentially growing culture of *E. coli* AB2829/pKD501 and digested with S1 nuclease. The pattern of protected DNA fragments was analysed by electrophoresis on a buffer-gradient 6% polyacrylamide/8 M-urea DNA sequencing gel, alongside a sequencing ladder. Fig. 5 shows the resulting autoradiograph. A ladder of bands is observed which is probably due to degradation of the mRNA at the 5′-terminus, prior to hybridization. The longest member of this group has a length of 90 nucleotides and corresponds to protection by full length mRNA. As is commonly found at the 5′ end of *E. coli* mRNAs this transcript begins with an A residue (at position 274 on Fig. 4) and is preceded by the sequence AATAATGTA (positions 263–271). The first seven nucleotides in this sequence are homologous to the consensus *E. coli* promoter −10 sequence (TATAATG: Hawley & McClure, 1983). Upstream of the −10 region, at an 'ideal' separation of 17 bp, is the sequence ATGATA (position 240–245) which is homologous to the consensus *E. coli* promoter −35 sequence (TTGACA). It therefore appears that the promoter of the *serC* gene is very similar to the consensus *E. coli* promoter both in sequence and in the spacing of the various elements. The 5′-untranslated region is 58 nucleotides in length. An alternative −10 region is the sequence TAATGTA (positions 265–271); although this is a satisfactory −10 sequence it does not have the ideal separation of 17 nucleotides from the −35 site.

A search for inverted repeats in the promoter region revealed a 26 bp sequence which can form a stem–loop structure (Fig. 4). The −10 region of the promoter is then located at the top of the stem–loop. This is a common feature of the operator sites of genes which are regulated by a repressor. Examples of this in aromatic biosynthesis include the *trp* operon (Platt, 1978), the *aroG* gene (Davis

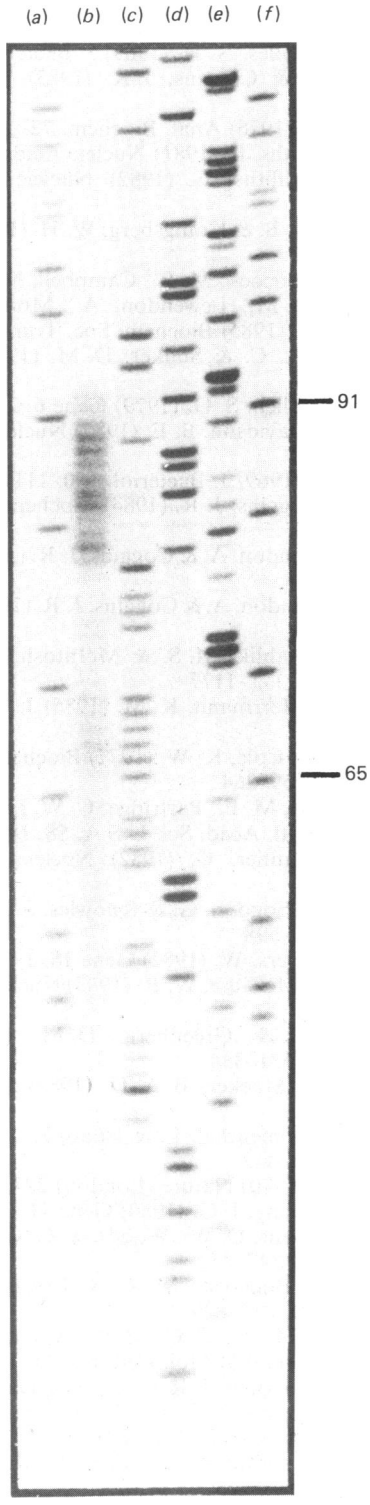

Fig. 5. S1 nuclease mapping of the transcription startpoint for the *serC* gene

Track (*a*), markers (the A track of a DNA sequencing ladder); track (*b*), ^{32}P-labelled DNA, hybridized to RNA prepared from *E. coli* AB2829/pKD501 and digested with S1 nuclease; tracks (*c–f*), markers (the T, G, C and A tracks, respectively) of a DNA sequencing ladder. Numbers indicate the length in nucleotides of the markers.

& Davidson, 1982) and the *aroH* gene (Zurawski *et al.*, 1981). The −10 region of the *trp* operon promoter is located at the top of a stem–loop to which the *trp* repressor is known to bind, preventing RNA polymerase from binding to the promoter. These observations suggest that expression of the *serC–aroA* operon might be regulated by a repressor. At the present time there is no direct evidence to support this hypothesis as no regulatory mutants of either *aroA* or *serC* have been described.

The *aroA* and *serC* genes are linked *in vivo*

It is conceivable that the deletion of a piece of DNA between the transcription terminator located downstream of *serC* and the start of the *aroA* coding region may have removed the *in vivo* promoter of the *aroA* gene. This could have occurred during the aberrant excision event which lead to the formation of λp*serC*, or during propagation of the phage prior to selection. To confirm that the *aroA* and *serC* genes are truly linked *in vivo* and that the observations described above are not an artefact introduced during manipulation of the genomic DNA, the *aroA* region was recloned directly from an *E. coli* K12 derivative (ATCC 14948). A *Pst*I fragment of *E. coli* K12 genomic DNA carrying the *aroA* gene was isolated by 'shotgun' cloning of a *Pst*I digest of *E. coli* DNA into the *Pst*I site of pAT153 and selecting for complementation of the *aroA* mutation in *E. coli* AB2829. The recombinant plasmid obtained, pKD508, was digested with *Cla*I and *Pvu*II and the 1.9 kb fragment equivalent to the insert in pKD506 was isolated. A *Taq*I digest of this DNA revealed the presence of the 124 bp fragment which spans the *serC–aroA* intergene region (results not shown) and confirmed the close arrangement of the two genes.

DISCUSSION

These results show that in *E. coli* expression of the *aroA* gene, encoding the pre-chorismate pathway enzyme EPSP synthase, is linked to the expression of the *serC* gene, which encodes PSAT, an enzyme catalysing a reaction on the serine biosynthetic pathway. The two genes constitutes a 'mixed function' operon since their products are involved in different biosynthetic pathways. The *E. coli* biosynthetic operons which have been studied in detail, for example the *trp* and *his* operons, consist of genes coding for enzymes which catalyse reactions on the same pathway. Induction of expression of these operons is necessary in order that the end products of the pathways may be synthesized, and each one of the gene products must be produced, otherwise the pathway will not function. For this reason it is widely believed that co-ordination of gene expression, in response to the needs of the cell, is the principal function of these operons. At first sight the *serC–aroA* operon does not appear to fit this hypothesis. EPSP synthase and PSAT are enzymes that catalyse reactions on separate pathways, and there is evidence in the literature that both enzymes are synthesized constitutively (Tribe *et al.*, 1976; McKitrick & Pizer, 1980). However, the end products of the two pathways, serine and chorismate, are substrates for the biosynthetic pathway leading to enterochelin, which is the ionophore required for the uptake of iron from the environment. The *serC–aroA* operon may provide the means of co-ordinating the expression of these two genes

so that enterochelin biosynthesis can proceed efficiently in response to iron starvation.

It is interesting to note that during growth on medium deficient in iron there are a number of changes in *E. coli* gene expression. Transcription of some of the enterochelin biosynthetic genes which code for enzymes catalysing the reactions after chorismate and serine is switched on from iron-regulated promoters (Laird *et al.*, 1980; Laird & Young, 1980; Fleming *et al.*, 1983). A number of tRNA species remain specifically undermodified and this affects aromatic amino acid transport and attenuation at the *trp* and *phe* operons (Buck & Griffiths, 1981, 1982). Also transcription of the *aroF* gene (encoding the tyrosine-repressible 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase) is derepressed (McCray & Herrmann, 1976). These effects together decrease the requirement for aromatic amino acids, allow for an increase in chorismate synthesis and increase enterochelin synthesis.

The synthesis of enterochelin requires both serine and chorismate in equimolar amounts. Studies by McKitrick & Pizer (1980) have indicated that there is a constant, but rapidly turning over, pool of serine within the cell. Apart from protein synthesis, serine is required for cysteine and phospholipid biosynthesis, and the conversion of serine to glycine provides most of the 'one-carbon' pool. A rapid increase in enterochelin biosynthesis would greatly diminish the serine pool. The linking of the expression of a serine biosynthetic pathway gene with the expression of an aromatic biosynthetic gene may serve to maintain serine levels when the requirement for enterochelin synthesis is high.

There is no evidence that attenuation plays a role in the expression of the *serC–aroA* operon. The sequence in the 5′ leader region cannot be folded to form structures analogous to those found in the *trp* operon (Yanofsky, 1981); nor is there a 'leader peptide' located in this region. The inverted repeat at the −10 region of the promoter may be involved in a different type of regulatory mechanism. This may be the operator site for an as yet unidentified repressor. The repressor which controls expression of the other enterochelin biosynthetic genes is not known. The sequences of iron-regulated promoters have not been examined in detail; only the *aroF* promoter sequence has been published but there is no similar stem–loop structure at the −10 region of the promoter (Garner & Herrmann, 1985). It is possible that the genes are expressed at a low constitutive level under conditions where there is sufficient iron and that this amount of each enzyme can satisfy the requirements for both serine and chorismate. Only under iron stress are the pools of serine and chorismate seriously depleted, and since there is no evidence that EPSP synthase or PSAT are subject to feedback inhibition the iron-starved cells require a means of simultaneously raising the levels of these two enzymes. The *serC–aroA* operon affords a convenient means for achieving this.

Since this work was completed the existence of the *serC–aroA* operon has also been reported in *Salmonella typhimurium* (Hoiseth & Stocker, 1985).

REFERENCES

Anton, D. L., Hedstrom, L., Fish, S. M. & Abeles, R. H. (1983) Biochemistry **22**, 5903–5908

Arai, K., Clark, B. F. C., Duffy, L., Jones, M. D., Kaziro, Y., Laursen, R. A., L'Italien, J., Miller, D. L., Nagarkatti, S., Nakamura, S., Nielsen, K. M., Petersen, T. E., Takahashi, K., & Wade, M. (1979) Proc. Natl. Acad. Sci. U.S.A. **77**, 1326–1330

Bachmann, B. (1983) Microbiol. Rev. **44**, 180–230

Berlyn, M. B. & Giles, N. H. (1969) J. Bacteriol. **99**, 222–230

Boocock, M. R. & Coggins, J. R. (1983) FEBS Lett. **154**, 127–133

Bradford, M. M. (1976) Anal. Biochem. **72**, 248–254

Buck, M. & Griffiths, E. (1981) Nucleic Acids Res. **9**, 401–414

Buck, M. & Griffiths, E. (1982) Nucleic Acids Res. **10**, 2609–2624

Clarke, S. J., Low, B. & Konigsberg, W. H. (1973) J. Bacteriol. **113**, 1091–1095

Coggins, J. R., Boocock, M. R., Campbell, M. S., Chaudhuri, S., Lambert, J. M., Lewendon, A., Mousdale, D. M. & Smith, D. D. S. (1985) Biochem. Soc. Trans. **13**, 299–303

Comai, L., Sen, L. C. & Stalker, D. M. (1983) Science **221**, 370–371

Dagert, M. & Ehrlich, S. D. (1979) Gene **6**, 23–28

Davis, W. D. & Davidson, B. E. (1982) Nucleic Acids Res. **10**, 4045–4058

Dempsey, W. B. (1969) J. Bacteriol. **100**, 1114–1115

Duncan, K. & Coggins, J. R. (1984) Biochem. Soc. Trans. **12**, 274–275

Duncan, K., Lewendon, A. & Coggins, J. R. (1984*a*) FEBS Lett **165**, 121–127

Duncan, K., Lewendon, A. & Coggins, J. R. (1984*b*) FEBS Lett. **170**, 59–63

Fleming, T. P., Nahlik, M. S. & McIntosh, M. A. (1983) J. Bacteriol. **156**, 1171–1177

Garner, C. C. & Herrmann, K. M. (1985) J. Biol. Chem. **260**, 3820–3825

Gaertner, F. H. & Cole, K. W. (1977) Biochem. Biophys. Res. Commun. **75**, 259–264

Giles, N. H., Case, M. E., Partridge, C. W. H. & Ahmed, S. I. (1967) Proc. Natl. Acad. Sci. U.S.A. **58**, 1453–1460

Gouy, M. & Gautier, C. (1982) Nucleic Acids Res. **10**, 7055–7074

Grimshaw, C. E., Sogo, S. G. & Knowles, J. R. (1982) J. Biol. Chem. **257**, 596–598

Grosjean, H. & Fiers, W. (1982) Gene **18**, 199–209

Hawley, D. K. & McClure, W. R. (1983) Nucleic Acids Res. **11**, 2237–2255

Hirsch-Kolb, H. & Greenberg, D. M. (1971) Methods Enzymol. **17B**, 331–334

Hoiseth, S. K. & Stocker, B. A. D. (1985) J. Bacteriol. **163**, 355–361

Kitakawa, M., Blumenthal, L. & Isono, K. (1980) Mol. Gen. Genet. **180**, 343–349

Laemmli, U. K. (1970) Nature (London) **227**, 680–685

Laird, A. J. & Young, I. G. (1980) Gene **11**, 359–366

Laird, A. J., Ribbons, D. W., Woodrow, G. C. & Young, I. G. (1980) Gene **11**, 347–357

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

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York

McCray, J. W. & Herrmann, K. M. (1976) J. Bacteriol. **125**, 608–615

McKitrick, J. C. & Pizer, L. I. (1980) J. Bacteriol. **141**, 235–245

Messing, J. (1983) Methods Enzymol **101**, 20–78

Messing, J. & Vieira, J. (1982) Gene **19**, 269–276

Mousdale, D. M. & Coggins, J. R. (1984) Planta **160**, 78–83

Pittard, J. & Wallace, B. J. (1966) J. Bacteriol **91**, 1494–1508

Platt, T. (1978) in The Operon (Millar, J. H. & Reznikoff, W. S., eds.), Cold Spring Harbor Laboratory, New York

Rogers, S. G., Brand, L. A., Holder, S. B., Sharps, E. S. & Brackin, M. J. (1983) Appl. Environ. Microbiol. **46**, 37–43

Rosenberg, M. & Court, D. (1979) Annu. Rev. Genet. **13**, 19–53
Sanger, F. (1981) Science **214**, 1205–1210
Schnier, J. & Isono, K. (1982) Nucleic Acids Res. **10**, 1857–1865
Shimizu, S. & Dempsey, W. B. (1978) J. Bacteriol. **134**, 944–949
Shine, J. & Dalgarno, L. (1975) Nature (London) **254**, 34–38
Steinrucken, H. C. & Amrhein, N. (1980) Biochem. Biophys. Res. Commun. **94**, 1207–1212
Steinrucken, H. C. & Amrhein, N. (1984) Eur. J. Biochem. **143**, 341–349
Tribe, D. E., Camakaris, H. & Pittard, J. (1976) J. Bacteriol. **127**, 1085–1097
Twigg, A. J. & Sherratt, D. (1980) Nature (London) **283**, 216–218
Yanofsky, C. (1981) Nature (London) **289**, 751–758
Zurawski, G., Gunsalus, R. P., Brown, K. D. & Yanofsky, C. (1981) J. Mol. Biol. **145**, 47–73

Received 28 August 1985; accepted 14 October 1985