The *purB* gene of *Escherichia coli* K-12 is located in an operon

Stephen M. Green,[†] Tahir Malik,[‡] Ian G. Giles and William T. Drabble

Author for correspondence: William T. Drabble. Tel: +44 1703 594279. Fax: +44 1703 594459.

Department of
Biochemistry, University of
Southampton, Biomedical
Sciences Building, Bassett
Crescent East,
Southampton SO16 7PX,
UK

The structural gene (*purB*) for succinyl-AMP (S-AMP) lyase and three additional ORFs are on the same DNA strand of the chromosome of *Escherichia coli*. Cassette mutagenesis and primer extension mapping demonstrated that *purB* is co-transcribed with an upstream gene (ORF23, or *ycfC*) encoding a 22.9 kDa membrane-associated protein of non-essential, but unknown, function unrelated to purine biosynthesis. The *purB* operon lies between *phoP* and an ORF expressing an essential function which may correspond to *asuE* (*trmU*). S-AMP lyase was purified to near homogeneity. The purified enzyme is a homotetramer of 50 kDa subunits, has a K_m for S-AMP of 3.7 μ M and a pH optimum of 7.4–7.6.

Keywords: *Escherichia coli*, *purB* gene, succinyl-AMP lyase, internal promoter, *phoP* gene

INTRODUCTION

The purine nucleotides AMP and GMP are synthesized through a branched multi-enzyme pathway. In *Escherichia coli*, the structural genes (*pur* and *gua*) for these enzymes have been mapped on the chromosome (Berlyn *et al.*, 1996), and occur at various loci either individually (*purT*, *purL*, *purC*, *purA*) or collectively within operons (*purF*, *purHD*, *purMN*, *purEK*, *guaBA*). The nucleotide sequences for all the genes are known but the previously published sequence and organization of the *purB* region (He *et al.*, 1992) differ significantly from those reported here. *purB* maps at 25.57' on the *E. coli* chromosome between *asuE* (*trmU*), the site of an antisuppressor mutation that affects tRNA modification (Sullivan *et al.*, 1985; Björk, 1995), and *phoP* (Groisman *et al.*, 1992; He *et al.*, 1992; Kasahara *et al.*, 1992), encoding a regulatory protein involved in stress responses (Miller, 1991). *purB* encodes succinyl-AMP (S-AMP) lyase (EC 4.3.2.2), a bifunctional enzyme that converts succinyl-AMP to AMP (the last reaction of AMP biosynthesis) and succinyl aminoimidazole carboxamide ribotide to aminoimidazole carboxamide ribotide (a precursor of both AMP and GMP).

Expression of the *pur* and *gua* genes is under multivalent regulation. General repression is mediated by the PurR protein (Meng *et al.*, 1990; He *et al.*, 1990), the product of the *purR* gene (Rolfes & Zalkin, 1988), which binds to the *pur* operator, a 16 bp palindrome (Schumacher *et al.*, 1994). The PurR co-repressors have been identified as hypoxanthine and guanine (Rolfes & Zalkin, 1990a; Meng & Nygaard, 1990). Binding of PurR at a *pur* operator within *purB* represses *purB* (He *et al.*, 1992; He & Zalkin, 1992). The *gua* operon encoding two enzymes specific for GMP biosynthesis is repressed, in addition, by DnaA protein (Tesfa-Selase & Drabble, 1992, 1996) and is regulated by stringent and growth-rate-dependent controls (Davies & Drabble, 1996). These controls link the biosynthesis of GMP to DNA replication and to stable RNA synthesis, respectively.

In this report, we describe the cloning and sequencing of *purB* and adjacent regions of the *E. coli* chromosome, and identify a new operon and its promoter. *purB* is shown to be the second gene of this two-gene transcriptionally linked operon and contains sites of potential regulatory significance, including a *pur* operator and a DnaA box. The promoter-proximal gene encodes a 22.9 kDa membrane-associated protein of unknown, but non-essential, function.

[†] **Present address:** Department of Molecular Microbiology, University of Southampton, Southampton General Hospital, Southampton SO16 6YD, UK.

[‡] **Present address:** Tampa Bay Research Institute, 10900 Roosevelt Blvd, St Petersburg, FL 33716, USA.

Abbreviation: S-AMP, succinyl-AMP.

The EMBL accession number for the sequence reported in this paper is X59307.

METHODS

Bacterial strains and plasmids. The strains of *E. coli* used are listed in Table 1. Plasmids are shown in Table 1 and Figs 1 and 5.

Table 1. Bacterial strains and plasmids

Strain/plasmid	Description	Source/reference
Strain		
JC7623	<i>recB21 recC22 sbcB15 thr-1 ara-14 leuB6 Δ(gpt-proA)62 lacY1 tsx-33 supE galK2 hisG4 rfbD1 rpsL21 kdgK51 xyl-5 mtl-1 argE3 thi-1</i>	Kushner <i>et al.</i> (1971)
JM109	[F' <i>traD36 lacI^q Δ(lacZ)M15 proAB⁺</i>] <i>recA1 endA1 gyrA96 thi hsdR17 (r_K⁻ m_K⁺) supE44 relA1 Δ(lac-proAB)</i>	Yanisch-Perron <i>et al.</i> (1985)
JM109(DE3)	JM109 (λ DE3)*	Studier & Moffatt (1986)
MC1061	F ⁻ <i>araD139 Δ(ara-leu)7697 Δ(lac)X74 galU galK hsdR (r_K⁻ m_K⁺) mcrB rpsL</i>	Casadaban & Cohen (1980)
MW1047	W3110 <i>purB747</i>	Laboratory strain
TG2	[F' <i>traD36 lacI^q Δ(lacZ)M15 proAB⁺</i>] <i>Δ(lac-proAB) supE thi recA srl::Tn10 Δ5(r_K m_K)</i>	Sambrook <i>et al.</i> (1989)
W3110	F ⁻ λ ⁻ (prototroph)	J. Guest, University of Sheffield, UK
Plasmid		
p124	See Fig. 1	Sullivan <i>et al.</i> (1985)
pACYC177	Small, medium-copy-number cloning vector	Chang & Cohen (1978)
pACYC184	Small, medium-copy-number cloning vector	Chang & Cohen (1978)
pBluescript KS(+)	Vector for expression of fusion proteins from T7 promoter	Stratagene
pHP45Ω	Plasmid carries spectinomycin-resistance cassette (Ω) for insertional mutagenesis	Prentki & Krisch (1984)
pNM481, pNM482	<i>lacZ</i> gene translational fusion vectors	Minton (1984)
pSG108	7.1 kbp <i>SaI</i> fragment from p124 cloned into <i>SaI</i> -cut pUC18	This work
pSG116	Derived from pSG108 (Fig. 1) by removal of inserted DNA leftwards of <i>NsiI</i> site and rightwards of <i>Clal</i> site	This work
pTM105	1.75 kbp <i>BglII</i> fragment of pSG108 cloned into <i>SmaI</i> site of pNM482	This work (Fig. 5)
pTM106	2 kbp <i>SmaI</i> ΩSp ^R cassette from pHP45Ω cloned into ORF15 <i>Asp718</i> site of pTM105	This work (Fig. 5)
pTM107	1.75 kbp <i>EcoRI</i> fragment of pTM105 cloned into <i>EcoRI</i> site of pACYC184	This work (Fig. 5)
pTM108	2 kbp <i>SmaI</i> ΩSp ^R cassette from pHP45Ω cloned into ORF23 <i>AflIII</i> site of pTM107	This work (Fig. 5)
pTM109	3.75 kbp <i>EcoRI</i> fragment (Sp ^R) of pTM108 cloned into <i>EcoRI</i> site of pNM481	This work (Fig. 5)
pTM110	1.70 kbp <i>BglII</i> - <i>EcoRI</i> fragment (ORF15 ORF23 <i>purB'</i>) cloned into <i>BamHI</i> - <i>EcoRI</i> -cut pBluescript KS(+)	This work
pTM114	4.58 kbp <i>HindIII</i> - <i>BamHI</i> fragment of pSG116 carrying Cm ^R cassette in <i>AflIII</i> site cloned into pUC19	This work
pTM115	<i>Asp718</i> site of ORF15 carried on pTM105 digested with <i>Asp718</i> , filled-in and religated	This work (Fig. 5)
pTM117	4.58 kbp <i>HindIII</i> - <i>BamHI</i> fragment of pSG116 carrying Cm ^R cassette in <i>Asp718</i> site cloned into pUC18	This work
pUC18, pUC19	High-copy-number cloning vectors	Messing (1991)

* Bacteriophage λ carrying the gene for T7 RNA polymerase integrated into the host genome.

Media. The enriched (L broth) and defined (phosphate-buffered minimal salts) media used have been described previously (Tesfa-Selase & Drabble, 1992). Liquid media were solidified with 1.5% (w/v) agar. Minimal medium was supplemented with Casamino acids (0.5%) or with individual amino acids (20 µg ml⁻¹). Antibiotics were used at the following final concentrations (µg ml⁻¹): ampicillin (Ap), 50; chloramphenicol (Cm), 30; spectinomycin, 100.

Preparation of plasmids and DNA manipulations. Plasmid DNA suitable for restriction endonuclease digestion, ligation, transformation and sequencing was prepared using DNA-purification columns (Promega). Large-scale isolation of plasmid DNA suitable for all applications and recombinant DNA techniques used for plasmid construction are described in Sambrook *et al.* (1989). DNA fragments were isolated from agarose gels using Prepagene (Northumbria Biologicals). Pro-

gressive shortening of plasmid DNA to produce a set of nested deletions for use in sequencing was performed using the reagents and protocols supplied with the Stratagene Exonuclease III/Mung Bean Nuclease Deletion kit.

Transformation. Cells were made competent by treatment with calcium chloride using a method based on that of Mandel & Higa (1970). Lac⁺ transformants were detected on media containing X-Gal (40 µg ml⁻¹).

DNA sequencing. Plasmid DNA was sequenced on both strands by the dideoxy chain-termination method using Sequenase Version 2.0 (United States Biochemicals). Sequencing transcripts were separated by electrophoresis through 0.4-mm-thick 6% (w/v) polyacrylamide gels containing 46% (w/v) urea. Compressions caused by formation of secondary structure in GC-rich regions of the sequence were resolved by substituting

dITP for dGTP in the labelling and termination reactions. For reading sequences up to 5 nt from the primer, manganese buffer was included in the labelling reaction. Oligonucleotide 17-mers used as sequencing primers, and other primers, were synthesized using an Applied Biosystems 381A DNA synthesizer by the phosphoramidate method and purified using oligonucleotide purification cartridges (Applied Biosystems).

Cell-free extracts. Extracts were prepared from 100 ml cultures grown in minimal medium supplemented with 0.5% Casamino acids. For growth under repressing conditions, 100 µg adenine ml⁻¹ was added. Bacteria were harvested at an OD₆₀₀ of 0.7–0.9, resuspended in 0.5 ml 30 mM potassium phosphate buffer, pH 7.0, and ultrasonically disrupted at 4 °C using an MSE Soniprep 150. Cell debris was removed by centrifugation (30000 g for 1 h) and the supernatant was dialysed against 30 mM phosphate buffer, pH 7.0. Protein concentration was measured by the Bradford (1976) method.

Assay for S-AMP lyase. S-AMP lyase was assayed by following the decrease in A_{280} (Pye Unicam SP8-400 spectrophotometer) due to the disappearance of succinyl-AMP. The specific activity of S-AMP lyase is defined as the loss of 1 µmol succinyl-AMP ($\epsilon_{280} = 10.7 \times 10^3 \text{ l mol}^{-1} \text{ cm}^{-1}$) (mg protein)⁻¹ min⁻¹ at 25 °C. The assay mixture (total volume 0.5 ml) contained 50 mM HEPES/KOH buffer, pH 7.2; 40 µM succinyl-AMP; and enzyme (approximately 150 µg protein for cell-free extracts, 2 µg protein for purified enzyme).

Electrophoresis of proteins. The protein content of extracts of disrupted cells was analysed by SDS-PAGE using a discontinuous buffer system and with appropriate size markers. Gels were stained with Coomassie Brilliant Blue R. The molecular mass of native S-AMP lyase was estimated using a non-denaturing gel system based on the method of Bryan (1977).

Purification and N-terminal sequence of S-AMP lyase. The enzyme was purified from extracts of TG2(pSG108) over-expressing the protein. The extract (1 ml containing 6 mg protein) was diluted twofold with 25 mM Tris/HCl (pH 8.0) (buffer A) and filtered through a 0.22 µm filter to remove residual debris. FPLC was carried out at room temperature. The FPLC column (Pharmacia; Mono Q anion-exchange resin) was equilibrated using the following sequence (flow rate 1 ml min⁻¹): buffer A (6 min), 25 mM Tris/HCl (pH 8.0) containing 1 M NaCl (buffer B) (6 min), buffer A (6 min). The protein sample was applied to the column and eluted with the following buffer sequence (flow rate 1 ml min⁻¹): buffer A (2 min), 0–50% buffer B/100–50% buffer A (30 min), buffer B (4 min), buffer A (6 min). Fractions containing the enzyme were pooled and desalted using a PD-10 gel-filtration column (Pharmacia). Purified S-AMP lyase was stored at –20 °C for up to 6 months without significant loss of enzyme activity.

N-terminal sequencing of the purified S-AMP lyase was by limited Edman degradation. Protein (5 µg) in 30 µl 10 mM phosphate buffer (pH 7.2) was diluted twofold with 50% (v/v) acetonitrile/1% (v/v) trifluoroacetic acid (TFA) solution. The protein solution was spotted onto a glass fibre disc and treated with TFA and Biobrene detergent. The disc was dried and placed in the reaction vial of an Applied Biosystems 477A Protein Sequencer. Seven cycles of Edman degradation were performed and the phenylthiohydantoin derivatives were separated using an Applied Biosystems 120A HPLC analyser.

Catalytic and kinetic properties of S-AMP lyase. S-AMP lyase was assayed over a pH range of 6.5–9.0. The buffers used were MES/KOH (pH 5.5, 6.0, 6.5 and 7.0), HEPES/KOH (pH 6.5–8.5 in 0.1 pH increments) and Tris/HCl (pH 8.0, 8.5 and

9.0). The mean activity at each pH value was calculated from duplicate measurements. S-AMP lyase was assayed at succinyl-AMP concentrations of 0.125, 0.25, 0.5, 1.0, 2.0 and 4.0 times the K_m (14 µM) for the S-AMP lyase of *Salmonella typhimurium* (Gots & Berberich, 1965). The mean of four assays at each concentration was determined.

Gene fusions. DNA fragments were placed in-frame with *lacZ* into plasmids pNM481 and pNM482 (Minton, 1984). β -Galactosidase was assayed as described by Tesfa-Selase & Drabble (1992) using strain MC1061 as host. The recorded β -galactosidase activities are means from at least two independent cultures; the overall variation was not greater than $\pm 10\%$.

Cassette mutagenesis. Insertion of the Ω fragment from plasmid pHP45 Ω (Prentki & Krisch, 1984) was used to disrupt the ORF15 and ORF23 reading frames (see Fig. 5). The Ω fragment carries the streptomycin/spectinomycin resistance gene of R100 flanked by short inverted repeats within which are transcriptional and translational termination signals and synthetic polylinkers.

The method of Kulakauskas *et al.* (1991) was used to move mutant alleles (carrying an inserted Cm^R cassette) from a plasmid onto the chromosome as follows. The 3.6 kbp ORF15–ORF23–*purB'* fragment of pSG116 was released by *Hind*III–*Bam*HI and cloned into *Hind*III–*Bam*HI-cut pACYC177. This vector contains no site for *A*flIII or *A*sp718 but they are present within ORF23 and ORF15, respectively, of the inserted fragment. The 977 bp *Sau*3A fragment of pACYC184, which carries Cm^R, was inserted at the *A*flIII site (in ORF23) or the *A*sp718 site (in ORF15) by blunt-end ligation. The mutated fragments were then released by *Hind*III–*Bam*HI treatment and cloned into pUC19 or pUC18 to give pTM114 (ORF23::Cm^R) and pTM117 (ORF15::Cm^R), respectively. Mixed λ phage lysates were obtained by infecting MC1061 harbouring either pTM114 or pTM117 with λ 7F9 (Kohara *et al.*, 1987). The lysates contained a mixture of λ 7F9 (no recombination), co-integrates (one cross-over between the phage and the plasmid) carrying both Cm^R and Ap^R, and λ 7F9 carrying the Cm^R cassette in the appropriate ORF (two cross-overs) but not Ap^R. The mixed lysates were used to transfer the mutation to the chromosomal ORF by homologous recombination by infecting JC7623 (*recBC sbcB*). Selection of transductants on Cm medium excluded transfer to the chromosome of non-recombinant phage and of co-integrates (which being derived from a ColE1-type plasmid are extremely unstable in a *recBC sbcC* strain).

Primer extension mapping. Experiments were carried out using a Primer Extension kit (Promega). Specific primers, labelled at their 5' ends with [γ -³²P]ATP, were hybridized, each at its optimum temperature (Sasse-Dwight & Gralla, 1991), to RNA isolated by the method of Wilkinson (1991) from TG2 carrying pSG108 (Fig. 1). Primer extension was initiated by the addition of avian myeloblastosis virus reverse transcriptase, and the DNA products were analysed on 6% (w/v) denaturing polyacrylamide gels alongside either sequencing ladders or the end-labelled DNA size markers (*Hinf*I fragments of ϕ X174 DNA) provided. Five primers were synthesized (see above). These hybridize to, and initiate reverse transcription from, RNA at specific regions within ORF15, ORF23 and the *purB* coding sequence (Fig. 3).

Expression of cloned genes and cell fractionation. A modification of the method of Tabor & Richardson (1985) was used for controlled high-level gene expression *in vivo*. DNA fragments cloned in the correct orientation into a Bluescript plasmid (Stratagene) were specifically expressed from a T7 promoter. The host strain JM109(DE3) carries an IPTG-inducible gene

Fig. 1. Restriction maps of p124 and pSG108. The pUC18 vector (2.7 kbp) of pSG108 is not shown. The location of the *purB* gene and its direction of transcription are indicated. Restriction enzyme sites: A, *AccI*; B, *BglII*; C, *Clal*; E, *EcoRI*; H, *HindIII*; K, *KpnI* (*Asp718*); N, *NsiI*; S, *Sall*; Sp, *SphI*. *E_v* and *H_v* indicate vector polylinker restriction sites.

for T7 RNA polymerase. JM109(DE3), with or without the appropriate plasmid, was grown in L broth to an OD_{600} of 0.6. Cells from 0.2 ml culture were washed in M9 medium (Sambrook *et al.*, 1989) and resuspended in 1 ml M9 medium supplemented with 1% methionine assay medium (Difco). Bacteria were then grown at 37 °C for 1 h when IPTG (0.5 mM) was added, followed by a further 15 min incubation before addition of rifampicin (200 μ g ml⁻¹). The cells were left at 37 °C for 30 min for exclusive T7 promoter expression. [³⁵S]Methionine (3.7 $\times 10^5$ Bq; 37 TBq mmol⁻¹) was present for the final 15 min after which time the bacteria were collected by centrifugation and resuspended in 50 μ l 1% (w/v) SDS containing 6 M urea and 0.1% 2-mercaptoethanol. Proteins were analysed by SDS-PAGE and Coomassie Brilliant Blue staining. Labelled proteins were visualized by autoradiography. The method of Ames *et al.* (1984) was used to release periplasmic proteins. The pellet of ³⁵S-labelled bacteria was treated for 15 min at room temperature with 20 μ l chloroform. After addition of 0.2 ml 10 mM Tris/HCl (pH 8.0) and mixing, the cells were collected by centrifugation (20 min). The supernatant (containing periplasmic proteins) was evaporated to dryness then redissolved in 30 μ l SDS-PAGE loading buffer. The cell pellet (containing the membrane fraction and cytoplasmic proteins) was resuspended in 30 μ l SDS-PAGE loading buffer. For the preparation of membrane fractions, labelled cells were collected by centrifugation, resuspended in 1 ml distilled water, and sonicated (MSE Soniprep 150, ten 30 s bursts with 45 s cooling periods between). Whole cells and debris were removed by centrifugation at 12000 *g* for 1 min and the supernatant was spun for 1 h at 40000 *g* to pellet the membrane fraction. The supernatant (containing soluble proteins) was evaporated to dryness and redissolved in 30 μ l SDS-PAGE loading buffer. The membrane pellet was resuspended in 1 ml distilled water, recentrifuged at 40000 *g* for 1 h, and the membrane pellet finally dissolved in 30 μ l SDS-PAGE loading buffer.

RESULTS AND DISCUSSION

Cloning the *purB* region

The source of DNA from the 25' region of the *E. coli* chromosome was p124 (Sullivan *et al.*, 1985), a plasmid derived from pLC3-2 (Clarke & Carbon, 1976) by deletion of two *HindIII* fragments. Tn5 insertion mutagenesis has shown that p124 carries at least *asvE* (*trmU*) and *purB* in different transcription units (Sullivan *et al.*, 1985). p124 conferred prototrophy on the *purB* auxotroph MW1047 and a 12-fold increase in S-AMP lyase activity (to 0.216 μ mol min⁻¹ mg⁻¹) compared with the parental prototrophic strain (W3110). Restriction mapping of p124 (Fig. 1) disclosed a third *Sall* site not recorded by Sullivan *et al.* (1985) and suggested that *purB* would be present on a 7.1 kbp *Sall* fragment. This fragment was cloned into

pUC18 using host TG2 to give plasmid pSG108 (Fig. 1). Cell-free extracts of TG2(pSG108) grown without adenine exhibited a 238-fold overexpression of S-AMP lyase (4.284 μ mol min⁻¹ mg⁻¹) compared with TG2(pUC18), which was reduced to half this value by growth under repressing conditions (100 μ g adenine ml⁻¹ added to the culture medium). This suggested that *purB* had been subcloned with its control elements intact. SDS-PAGE of extracts prepared from cells grown without adenine revealed an intense band of a 50 kDa protein (Fig. 2a).

Purification and properties of S-AMP lyase

S-AMP lyase from an extract of overexpressing cells [TG2(pSG108)] was further purified by anion-exchange FPLC. When analysed by SDS-PAGE, the purified enzyme migrated as a band of molecular mass 49500 Da (Fig. 2b) compared with a sequence-derived molecular mass of 51542 Da (see below). The mean native molecular mass, as estimated by non-denaturing PAGE, was 211 kDa (data not shown), which is consistent with the enzyme being a homotetramer. The specific activity of purified S-AMP lyase was 1.7 μ mol S-AMP removed min⁻¹ mg⁻¹. The K_m for S-AMP is 3.7 μ M and the enzyme has a pH optimum of 7.4–7.6. The N-terminal amino acid sequence obtained from purified S-AMP lyase was MELSVLT, which differs from the derived sequence (Fig. 3) only at the fifth residue.

ORFs and their products

The sequence of 2986 bp obtained for the *purB* region was examined for potential reading frames using the FRAMESCAN program (Staden & McLachlan, 1982). Four ORFs are present on one strand; no ORFs of any significant size are present on the complementary strand (Fig. 3). The longest ORF is assigned to *purB* for the following reasons: the N-terminal amino acid sequence of purified S-AMP lyase (see above) corresponds to that derived from the 5' end of the ORF; the ORF encodes a polypeptide of 456 amino acids with a predicted molecular mass of 51542 Da, which is close to that measured for the purified S-AMP lyase subunit by SDS-PAGE (see above); and the amino acid sequence corresponding to the ORF shows a high degree of homology to S-AMP lyase of *Bacillus subtilis* (Ebbole & Zalkin, 1987) and to other enzymes catalysing analogous fumarate elimination reactions (Woods *et al.*, 1988). The most highly conserved

Fig. 2. SDS-PAGE (12% polyacrylamide gels). (a) Protein extracts of TG2 transformants grown without adenine. All lanes contain 5 µg protein. Lanes: 1, TG2(pUC18); 2, TG2(pSG108); 3, protein standards (molecular mass: 20.1, 24, 29, 36, 45 and 66 kDa). (b) S-AMP lyase purified by FPLC. Lanes: 1, 5 µg TG2(pUC18) extract; 2, 5 µg TG2(pSG108) extract; 3 and 4, 5 µg and 1 µg purified enzyme, respectively; 5, protein standards.

region between these sequences has the consensus GSS-MP-K-N (Fig. 3; nt 2255–2284), which probably occurs at the active site of enzymes catalysing fumarate elimination reactions. Based on the mapping data of Sullivan *et al.* (1985), which places *asuE* clockwise of *purB*, the direction of transcription of *purB* is anticlockwise on the *E. coli* chromosome.

The *purB* initiation codon is preceded by two sequences complementary to the 3' end of 16S rRNA which could function as a ribosome-binding site (Gold & Stormo,

1987). Synonymous codon usage (Sharp & Li, 1986) for *purB* is consistent with that of a highly to moderately expressed gene (data not shown) and the proportion of rare codons is low (1.2%).

The second ORF (ORF23; *yefC* in Berlyn *et al.*, 1996) terminates at a TGA codon (position 1370) only 3 nt upstream of the *purB* initiation codon. Potential translation initiation sites for ORF23 are two GTG codons at 725 and 731. The 'Perceptron' algorithm (Stormo *et al.*, 1982) identified ribosome-binding sites for both initiation codons, that for G₇₃₁TG having the higher score. ORF23 encodes a 213 amino acid polypeptide of 22915 Da. A third ORF (ORF15; *yefB* in Berlyn *et al.*, 1996), starting with the ATG codon at 294 and extending for 399 bp, potentially encodes a 133 amino acid polypeptide of 14789 Da.

Strains overexpressing S-AMP lyase also contain a protein of 23.7 kDa showing as a band on SDS-PAGE (Fig. 2a). This protein is possibly the product of ORF23. We demonstrated that ORF23 is indeed encoding a protein by using a system for the high-level expression of cloned genes *in vivo*. A *Bgl*II–*Eco*RI fragment, encompassing ORF15, ORF23 and the 5' end of *purB* (Figs 1 and 3), was cloned into the pBluescript KS(+) vector to yield pTM110. The translational reading frames on the cloned fragment were then selectively expressed from a T7 promoter on the vector by inducing production of T7 RNA polymerase in the presence of rifampicin. Only two proteins (molecular mass approximately 17.5 kDa and 23 kDa) were highly labelled with [³⁵S]methionine during the expression period (Fig. 4a). The smaller protein has the size expected for a fusion of the first 115 codons of *purB* to the *lacZ* antisense strand on the pBluescript vector to yield a polypeptide terminated after 174 amino acid residues. The size of the larger protein (P23) is consistent with its assignment to ORF23. There was no evidence from this experiment for a protein corresponding to ORF15.

The same expression system was used to locate P23 within the cell. P23 and the *purB* fusion protein were selectively expressed from the T7 promoter on pTM110. The whole cells were subsequently treated with chloroform to obtain the periplasmic fraction (Ames *et al.*, 1984). Proteins from the periplasmic fraction and from the remaining cellular fraction were analysed by SDS-PAGE (Fig. 4b). P23 and the *purB* fusion protein were present in the cellular fraction, indicating that both proteins were localized within the membranes or the cytoplasm. Following selective expression from pTM110, whole cells were disrupted by ultrasonication followed by centrifugation to separate the soluble (cytoplasmic plus periplasmic) proteins from the membrane fraction. P23 and the *purB* fusion protein were shown to be localized to the membrane fraction (Fig. 4c). A small amount of P23 was present in the soluble fraction; this may represent protein not yet transported to the membrane. Although S-AMP lyase is a cytoplasmic protein, the fusion protein (*purB'*–*antilacZ*) has two long stretches of hydrophobic residues which may interact non-specifically with the membrane.

10 20 30 40 50 60
CAAGAGCCGCTGCTGCGTGGTCTGGACAGCAATAAAGACCAGAGCTACTTCTTTATAC

70 80 90 100 110 120
GCTCAGCCATGAGCAGATTGCGCAAAGCTGTTCCTCGGTCGGCGAACTGGAAAAACCGCAG

130 140 150 160 170 180
GTGCGTAAGATTGCTGAAGATCTTGGTCTGGTCCACGCGAAGAAAAAGACTCTACCGGC

190 200 210 220 230 240
ATTGCTTCATTGGCGAAGCTAAATTCGCGAGTTCCTGGGCCGTTATCTCCGGCGCAAC

250 260 270 280 290 300
CGGGCAAAATCATTACCGTCGATGGCGATGAAATTTGGTGAGCACCAGGGGCTGATGTATC

310 320 330 340 350 360
ACACTCTCGGTGAGCTAAAGGTCTGGGTATCGGTGGCACCAAGAAAGGTACCGAAGAAC

370 380 390 400 410 420
CGTGTATCTGGTGGACAAAGCGTCGAAACACATCTGGTGTCTGCTCAGGGCCATG

430 440 450 460 470 480
AACACCCGCGCTGATGTCTGCTGGGTGATTGCCAGCAGTTGCACTGGTGCATCGCG

490 500 510 520 530 540
AACCATTACCGGCTATGCGTTCGACGGTAAAAACCGCTATCGCCAGACCGATATCC

550 560 570 580 590 600
CTTGACCGCTCAAGGCGCTGGACGATGATCGATTGAAGTATTTTCGATGAACCGGTG

610 620 630 640 650 660
CCGCGTGACGCGCGGCGAGTCTGCGCTCTTCTATAACGGTGAAGTGTGCTCGGTGGCG

670 680 690 700 710 720
GTATTATTGAGCAGCGTCTGCCGCTGCCGCTGATTATTATCTTAAACAAAGGA

730 740 750 760 770 780
AGCAGTGAACGTGGCAAGAACTTACTATGACATCACCTCGCCCTGGCGGTATTTGCA

790 800 810 820 830 840
GTCGGACCGCTGGTGAACAACTCGCTCACCAGGGGCAATTGTGATCGCGATCGCTACA

850 860 870 880 890 900
CGTCTCACTCAACAGTATTATGATATGAACCCAGCTCGAGCTGGCGGTTTTTGCGCG

910 920 930 940 950 960
TAGCGAAGCCAACTGCGCGTGGGCTGGAAACCTGCTGGCGGTGCTCAATGCCAGCAG

970 980 990 1000 1010 1020
TCGCCAGGGCTTAAACCGCGAATTAACCCGCTACACACTCAGCTTGATGGTGTGAGCG

1030 1040 1050 1060 1070 1080
CAAACCTCTCTCAGCGAAAGCGCGCTCGACACTCTGGGCAACCGGATCAACGGCTGCA

1090 1100 1110 1120 1130 1140
ACGCCAGCTCGAACACTTCGATTACAGTCCGAAACGCTGATGAGCGGATGGCTGTAT

1150 1160 1170 1180 1190 1200
CTATGTTGATGTGATTAGCCCGCTTGGCCGCGCATTCAGGTACCGGTTCCCTGCTGT

1210 1220 1230 1240 1250 1260
ACTGCAAGCCCAAGTGCAGGCGAAGTTCGCGCAACCTGCTGGCAGCATTCGCGC

1270 1280 1290 1300 1310 1320
CGCCGCTCTGCGCAGGTCGCGCGCGGAGCTGCACTGATGTTTTCTCGTAATCG

1330 1340 1350 1360 1370 1380
CTGACCACTCAGGCAAAACAAATCTTGTCTATTAAACCCGGAGTTGATCTATGGA

1390 1400 1410 1420 1430 1440
ATTATCTCTACTGACCGCGTTCCTCTGCTGATGACGCTACGGCGATAAAGTCAGCGC

1450 1460 1470 1480 1490 1500
GCTGCGCGGATTTTCAGCAATATGGTTTCTGAAATTCCTGTACAAGTTGAAGTACG

1510 1520 1530 1540 1550 1560
TTGGTGCAGAAAACCTGGCGCGCAGCGATCAAGGAAGTTCTCTGTTTCTGCTCGCA

1570 1580 1590 1600 1610 1620
CGCAATCGGTTACCTTGATGCAATCGTCGCGAGTTTCAGCGAAGAACTGCGCGCGCAT

1630 1640 1650 1660 1670 1680
CAAACCTATCGAGCGTACCACTAACACGAGCTTAAAGCGGTGAGTATTCTCGTAAAGA

1690 1700 1710 1720 1730 1740
AAAAGTGGCGGAGATCCCGGAAGTGCAGCGGTTTCTGAATTCATCCACTTTGCTGTAC

1750 1760 1770 1780 1790 1800
TTCGGAAGATATCAATACCTCTCCAGCATTAACTGTAAGAAACCGCGGTGATGAAGT

1810 1820 1830 1840 1850 1860
GATCCTGCCATAGTGGCGTCAACTGATTGATGGCATTAAAGATCTCGCGGTTCAAGTATCG

1870 1880 1890 1900 1910 1920
CGATCTCCGCTGCTGCTCGTACCCAGCGTCAGCCAGCCAGCGTCAACCATCGGTAA

1930 1940 1950 1960 1970 1980
AGAGATGGCAAGCTGCGCTACCGTATGGAGCGCCAGTACCGCCAGCTTAAACAGGTGGA

1990 2000 2010 2020 2030 2040
GATCCTCGGCAAAATCAACGCGCGGTGCGTAATAACGCCACATCGCGCGTTACCC

2050 2060 2070 2080 2090 2100
GGAAGTTGACTGGCATCAGTTGAGCAAGAGTTCGTACCTCGCTGGGTATTTCAGTGGAA

2110 2120 2130 2140 2150 2160
CCCGTACACCCAGATCGAAACCGCAGTACATTCGCGAAGTGTGATTGCGGTTGC

2170 2180 2190 2200 2210 2220
GCGCTTCAACATATTCTGATCGACTTGACCGTGACGCTGGGGTTATATCGCCCTTAA

2230 2240 2250 2260 2270 2280
CCACTTCAACAGAAAACCTTGTGGTGGATGGTCTTCCACCATGCCCGCATAAAGT

2290 2300 2310 2320 2330 2340
TAACCGATCGATTCGAAACCTCGAAGGAACTGGGCTTTTCAACGCGGTATTGCA

2350 2360 2370 2380 2390 2400
GCATCTGGCAAGCAAACTGCCGTTTCCCGCTGGCAGCGTACCTGACCGACTTACCGT

2410 2420 2430 2440 2450 2460
GCTGCGTAACCTCGCGGTGGTATCGGTTATGCTTGTGATATCAATCCACCTTGAA

2470 2480 2490 2500 2510 2520
AGCGGTGAGCAAACTGGAAGTGAACCGTGACCATCTGCTGGATGAACCTGGATCACAACTG

2530 2540 2550 2560 2570 2580
GGAAGTGTGCTGGAACCAATCCAGACAGTTATCGCTGCTATGGCATCGAAAAACCGTA

2590 2600 2610 2620 2630 2640
CGAAGAGCTGAAAGAGCTGACTCGCGGTAAAGCGGTGACGCCAAGGCATGAAGCAGTT

2650 2660 2670 2680 2690 2700
TATCGATGGTCTGGCGTTGCCAGAAAGAGAAAGCCCGCTGAAAGCGATGACGCGCGC

2710 2720 2730 2740 2750 2760
TAACTATATTGGTTCGAGCTATCACGATGGTTGATGAGCTGAAATAAACCTCGTATCAGT

2770 2780 2790 2800 2810 2820
CCGGATGCGATGCTGTCGCGCTGCTTATTAAGATTATCCGCTTTTATTTTTCACCTT

2830 2840 2850 2860 2870 2880
TACCTCCCTCCCGCTGGTTTATTAATGTTTACCCCAATAACCATATAATCGCGTTAC

2890 2900 2910 2920 2930 2940
ACTATTTTAATAATTAAGACAGGAGAAATAAAATGCGCGTACTGGTGTGTAAGACAA

2950 2960 2970 2980 2990 3000
TGCGTTGTTACGTACCACTTAAAGTTTCAAGTTCAGATTGAGGATGCTGGT

Fig. 3. For legend see facing page.

The derived amino acid sequences of ORF23 and ORF15 were compared with entries in the SWISS-PROT database (release no. 31) using the FASTA program (Pearson & Lipman, 1988). No significant matches were detected. The protein products of these reading frames can be matched, however, to sequences in the recently published whole-genome sequence of *Haemophilus influenzae* Rd (Fleischmann *et al.*, 1995). P23 has 49.3% identity (68% similarity) to the hypothetical protein product (205 amino acids) of gene HI0638 of *H. influenzae*. This gene is 5' to the *purB* homologue (HI0639), with 22 nt separating the two reading frames. The translation product of ORF15 has similarity to the C-terminal portion of the hypothetical protein product (418 amino acids) of gene HI0174. Two single nucleotide insertions into the *E. coli* sequence upstream of ORF15 (between positions 110/111 and 230/231; Fig. 3) extend the open reading frame to the beginning of the sequence and increase the overlap with HI0174 to 227 amino acids (76.7% identity, 84% similarity). HI0174 is almost 500 000 nt distant from *purB* in *H. influenzae*.

The nucleotide sequence 3' to position 2835 is identical to the sequence for *phoP* previously reported (Groisman *et al.*, 1992; He *et al.*, 1992; Kasahara *et al.*, 1992). A protein in extracts of TG2 carrying pSG108, and running on gels with a molecular mass close to 27 kDa (Fig. 2a), may be the *phoP* product (molecular mass 25.5 kDa; Kasahara *et al.*, 1992).

A nucleotide sequence for the *purB* region has been published by He *et al.* (1992). Their sequence of 2711 bp starts at position 384 in Fig. 3 and includes an additional 78 bp at the 3' end. Their location for the *purB* initiation codon (derived from N-terminal sequencing of S-AMP lyase) agrees with that reported here, but the following differences between the two sequences affect the ORF23 and *purB* coding regions: deletions at positions 871, 1251 and 1261 result in frameshifts within ORF23; inversions at positions 1807/8, 1834/5, 2154/5 and 2342/3 result in amino acid substitutions in S-AMP lyase; and deletion of G at position 2653 results in an S-AMP lyase shorter by 21 amino acids at the C-terminus [ALPEEEKARL(19 amino acid residues)K-COOH becomes RCQKKRKPA-COOH]. These differences eliminate ORF23 (by introducing stop codons) and create a 435 amino acid S-AMP lyase of molecular mass 49 225 Da (2317 Da less than the polypeptide derived from *purB* in Fig. 3). In addition, 'rare' codons (Sharp & Li, 1986) are created within the *purB* coding region. Independent confirmation of the nucleotide sequence from positions 1865 to 2986 reported

here is provided by Kasahara *et al.* (1992). They sequenced 4097 bp from the 25' region of the *E. coli* chromosome that includes the *phoPQ* operon downstream of *purB* and 881 bp of an ORF which they failed to identify as *purB*. The 5' 1122 bp of their sequence exactly matches the 3' end of the sequence reported here, suggesting that differences noted above result from errors within the sequence of He *et al.* (1992).

Transcription signals

The whole sequence was searched for potential promoters using the consensus sequence derived by Harley & Reynolds (1987) and the program PROBE (Giles, 1992). The highest scoring putative σ^{70} promoter occurs 64 nt upstream of the ORF23 initiation codon (Fig. 3). Seven nucleotides downstream of this promoter sequence is the highly conserved CGT, wherein G₈₇₆ is a likely transcription start site (Hawley & McClure, 1983). Only 3 bp separate ORF23 and *purB* and no other promoter-like or termination sequences are found within, or immediately following, ORF23, suggesting that ORF23 and *purB* are co-transcribed. A flagellin-type promoter sequence (Helmann & Chamberlin, 1987) was unexpectedly found between positions 200 and 230 (Fig. 3).

Downstream of the *purB* coding region is a potential rho-independent terminator (Rosenberg & Court, 1979), consisting of a region of hyphenated dyad symmetry capable of forming a stable mRNA stem and loop [ΔG (25 °C) = -26.6 kcal mol⁻¹ (-111.7 kJ mol⁻¹); Tinoco *et al.*, 1973] followed by a T-rich region (Fig. 3).

The extent of the *purB* transcriptional unit was investigated by insertional mutagenesis. The ORF15 and ORF23 reading frames in the 1.75 kbp *Bgl*II fragment from pSG108 were disrupted by insertion of the Ω interposon from plasmid pHP45 Ω . The Ω fragment is a 2 kbp segment of DNA carrying the spectinomycin/streptomycin resistance determinant of R100 flanked by short inverted repeats within which are transcriptional and translational termination signals and polylinkers (Prentki & Krisch, 1984). The Ω fragment was inserted at two positions; either at the unique *Asp*718 site within ORF15 or at the unique *Afl*III site within ORF23 (Fig. 3). The mutagenized sequences were cloned in-frame into an appropriate *lacZ* translational fusion vector (Fig. 5) so that expression of *purB* could be measured from assays of β -galactosidase during exponential growth. Strains carrying the vector plasmids alone (pNM481 and pNM482) had similar very low background levels of β -galactosidase

Fig. 3. Nucleotide sequence of the *purB* region. The nucleotide sequence coordinates are assigned relative to the first nucleotide at the 5' end. In the 5' to 3' direction, 2986 nt of the non-transcribed strand of the *purB* gene and flanking DNA are shown. Two ORFs (ORF15 and ORF23) upstream of *purB* and the downstream N-terminal part of *phoP* are also shown. Amino acid sequences of proposed coding regions are shown below the nucleotide sequences. Start and stop codons are in bold type. Potential ribosome-binding sites are underlined. The -35 and -10 sequences of the promoters for ORF23-*purB* and for *phoP* are underlined. Restriction enzyme sites are indicated, including the sites in ORF15 (*Asp*718) and in ORF23 (*Afl*III) used for insertion of the Ω fragment. Sequences complementary to primers (A-E) used in transcription mapping are overscored; transcription start points identified by transcript mapping are indicated by asterisks. A short sequence of dyad symmetry within ORF23 is indicated by converging arrows. The rho-independent terminator region of hyphenated dyad symmetry is indicated by converging arrows and the T-rich region is in bold type. A *pur* operator and a potential DnaA box in the *purB* coding region are underlined.

Fig. 5. Construction of transcriptional fusion plasmids and insertion mutagenesis of ORF15 and ORF23.

(Fig. 6; only data for pNM481 shown). Enzyme production was greatly increased by the in-frame fusion of the 1.75 kbp DNA fragment carrying ORF15, ORF23 and the 5' end of *purB* (pTM105), indicating the presence of at least one correctly orientated promoter on this fragment. The fragment also carries sequences mediating adenine repression of S-AMP lyase, because β -galactosidase production was reduced by the presence of adenine in the culture medium (Fig. 6). Insertion of Ω into ORF23 (pTM109) reduced enzyme production down to the background level of the control carrying only the vector. Insertion of Ω into ORF15 (pTM106), by contrast,

Fig. 4. Selective expression *in vivo* of cloned plasmid genes from a T7 promoter in strain JM109(DE3). (a) Unfractionated cells. Lanes: 1, plasmid-free strain; 2, with pBluescript KS(+) vector; 3, with pTM110. (b) Fractionated cells. Lanes: 1-3, periplasmic fractions; 4-6, cytoplasmic plus membrane fractions. Lanes: 1 and 4, plasmid-free strain; 2 and 5, with pBluescript KS(+) vector; 3 and 6, with pTM110. (c) Fractionated cells. Lanes: 1-3, membrane protein fractions; 4-6, soluble (periplasmic and cytoplasmic) protein fractions. Lanes: 1 and 4, plasmid-free strain; 2 and 5, with pBluescript KS(+) vector; 3 and 6, with pTM110. Numbers (kDa) on the right of the figures indicate the positions of the protein size markers on the stained gel.

Fig. 6. Detection of *purB* promoter function in *lacZ* transcriptional fusion plasmids before and after insertion mutagenesis of upstream ORFs. Differential synthesis of β -galactosidase [units β -galactosidase (mg protein) $^{-1}$; values given in parentheses after symbol] was measured during exponential growth in strain MC1061 carrying: pNM481 (vector alone) ($\blacktriangle$; 0.01); pTM105 [vector plus 1.75 kbp *Bgl*II fragment (ORF15 ORF23 *purB*)] ($\bullet$; 0.33); pTM109 (Ω inserted into ORF23) ($\blacklozenge$; 0.01); pTM106 (Ω inserted into ORF15) ($\blacksquare$; 0.13); pTM115 (frameshift at the *Asp*718 site in ORF15) ($\blacktriangledown$; 0.36); pTM105 (from a culture grown with 100 μ g adenine ml $^{-1}$ supplementation) ($-\bullet-$; 0.11).

resulted in a reduction in enzyme production to about half that from pTM105 (Fig. 6). The total loss of β -galactosidase by inserting the Ω interposon in ORF23 indicates co-transcription of ORF23 and *purB*, and suggests the presence of at least one promoter between the *Bgl*II site (position 140) and the *A*/III site (position 840) (Fig. 3). The reduced level of enzyme synthesis from pTM106 suggests that a promoter lies between positions 350 and 840, with a second sequence acting as a promoter between 140 and 350. It was noted above that a potential flagellin-type promoter is present in this part of the sequence. The *Asp*718 restriction site within ORF15 of pTM105 was also used to introduce a +4 frameshift into the translational reading frame to yield pTM115 (Fig. 5). This frameshift within ORF15 did not alter the high level of β -galactosidase activity produced by pTM105 (Fig. 6). Correct translation of ORF15 is not, therefore, a prerequisite for the expression of ORF23-*purB*.

Transcription mapping by the primer extension method was used to identify 5' ends of mRNA (Fig. 7). Five primers were prepared to hybridize to, and initiate cDNA synthesis from, mRNA at specific sequences within ORF15, ORF23 and *purB* (Fig. 3). Primers E and A correspond to the two ends of ORF15. No cDNA products that could identify recognizable σ^{70} promoter sequences were produced using these primers (data not shown). Primers B and D correspond to sequences in ORF23 downstream of the putative *purB* promoter identified from the sequence search. Transcription mapping with these primers indicated potential transcriptional start sites in the vicinity of this promoter within ORF15. The longest cDNA probably arises from extension caused by fold-back secondary structure in its GC-rich 3' end; the other cDNAs terminate in the predicted region for transcription initiation, with A₆₇₃ and G₆₇₆ being the most

likely candidates for the +1 nucleotide. Primer C, complementary to the 5' region of *purB*, gave several cDNA products. He *et al.* (1992) also attempted to map the *purB* promoter by primer extension using primers annealing at positions 1456–1475 and 1490–1508 (Fig. 3). These gave reverse transcripts indicating a transcriptional start at A₁₃₃₇ within ORF23. This site is not preceded, however, by a recognizable σ^{70} promoter but is immediately preceded by a sequence of hyphenated dyad symmetry (GCCTGACCACTCAGGCAA) capable of forming a stable mRNA stem and loop (Fig. 3). cDNA synthesis by avian myeloblastosis virus reverse transcriptase can be terminated by stem and loop structures (Tuerk *et al.*, 1988), especially when experimental conditions are not sufficiently denaturing. Two of the products using primer C (Fig. 7) correspond to cDNA terminated at either side of the symmetrical sequence and may be, therefore, artefactual.

Effects of chromosomal mutations in ORF23 and ORF15

ORF23 was mutated *in vitro* by insertion of a Cm^R cassette into the unique *A*/III site within the reading frame. The mutation was then transferred to λ 7F9 (Kohara *et al.*, 1987) by homologous recombination and subsequently to the chromosome by transduction (transduction frequency 4×10^{-7} per phage particle) using selection on nutrient medium containing Cm (see Methods). Twenty transductant colonies showing the required Cm^R Ap^S phenotype were plated to minimal agar medium with or without adenine supplementation. All were viable without adenine supplementation. P23 is therefore not essential for growth of *E. coli*, nor is it required for purine biosynthesis. The chromosomal *purB* and ORF23::Cm^R markers were shown to be closely linked in the constructed mutants. Using phage-P1kc-mediated transduction, a constructed mutant as donor, and strain MW1047 (*purB*747) as recipient, with selection for either Cm^R or purine auxotrophy, there was greater than 95% co-transduction of the non-selected marker in each cross.

The above procedure was repeated using the cassette mutation at the *Asp*718 site in ORF15; however no Cm^R transductants were obtained. The high frequency of transduction observed when introducing ORF23::Cm^R into the chromosome and the absence of transductants for ORF15::Cm^R suggests that ORF15 corresponds to an essential gene.

Other features in the sequence

A sequence differing from the consensus *pur* operator by 2 nt is found in the *purB* coding region 182 nt downstream of the initiation codon (Fig. 3). With the exception of *purR*, all genes regulated by PurR contain repressor-binding sites in their promoter regions (He *et al.*, 1990; Meng *et al.*, 1990). A putative *pur* operator, differing from the consensus sequence at only one position, is also found immediately upstream of the ORF23 homologue (HI0638) in *H. influenzae* (Fleischmann *et al.*, 1995). Autoregulation of *purR* by downstream repressor binding requires two

Fig. 7. Primer extension mapping of the 5' ends of transcripts from the *purB* region. RNA was isolated from strain TG2(pSG108). cDNA synthesis was initiated from three different primers (B, C and D in Fig. 3). (a, b) Lanes: 1, *HinfI* ϕ X174 DNA size markers (lengths given in nucleotides); 2, cDNA products initiated from primers D (a) and C (b) as indicated. (c) cDNA products (lane 3) initiated from primer B run alongside a sequencing ladder using the same primer. The sequence of cDNA is shown on the left.

pur operators (Meng *et al.*, 1990; Rolfes & Zalkin, 1990b) and is assumed to be mediated by transcription termination. The *pur* operator within *purB* has been shown to be a site of transcription termination (He *et al.*, 1992; He & Zalkin, 1992). The *purB* N-terminal coding region also contains a DnaA box at position 1382 (Fig. 3) on the non-transcribed strand which differs from the consensus sequence (TTAT^C_ACA^C_AA) at one position. As it has been demonstrated that GMP biosynthesis is regulated by the DnaA protein binding to a DnaA box on the non-transcribed strand of *guaB* (Tesfa-Selase & Drabble, 1992, 1996), it is possible that *purB* is a target for DnaA regulation of IMP and AMP biosynthesis.

Conclusions

We have shown that *purB* is co-transcribed with an upstream gene, ORF23. The same close coupling of *purB* to ORF23 is also seen in *H. influenzae* (Fleischmann *et al.*, 1995). It is unlikely that the protein product (P23) of ORF23 has any direct role in purine biosynthesis as ORF23 is dispensable, even in the absence of pre-formed

purines. P23 may not even be indirectly involved with purine nucleotide metabolism. The *purF* operon has the arrangement *cvpA-purF-ubiX*, where *cvpA* determines colicin V production (Fath *et al.*, 1989). The nearest upstream mapped gene to *purB* is *asuE*, but *asuE* and *purB* are not co-transcribed (Sullivan *et al.*, 1985). *asuE* cannot, therefore, be assigned to ORF23.

The promoter for ORF23-*purB* lies within the coding region of the upstream gene ORF15. ORF15 is probably the 3' end of a much longer reading frame which extends 5' to the sequence presented here up to *icd*, the next sequenced coding region. This extended reading frame corresponds to HI0174 of *H. influenzae* (Fleischmann *et al.*, 1995) and is a candidate for *asuE*. There are no apparent termination signals to prevent transcription continuing from this reading frame into ORF23-*purB*.

ACKNOWLEDGEMENTS

S.M.G. and T.M. gratefully acknowledge financial support from BBSRC. We thank Russell Jones for assistance with FPLC, and Lawrence Hunt (Institute of Biomolecular Sciences,

Southampton University) for N-terminal analysis. Strains were generously provided by Robert M. Bock and Martin D. Watson.

REFERENCES

- Ames, G. F., Prody, C. & Kustu, S. (1984). Simple, rapid and quantitative release of periplasmic proteins by chloroform. *J Bacteriol* **160**, 1181–1183.
- Berlyn, M. K. B., Brooks Low, K. & Rudd, K. E. (1996). Linkage map of *Escherichia coli* K-12, edition 9. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, pp. 1715–1902. Edited by F. C. Neidhardt and others. Washington, DC: American Society for Microbiology.
- Björk, G. R. (1995). Genetic dissection of synthesis and function of modified nucleosides in bacterial transfer RNA. *Prog Nucleic Acid Res Mol Biol* **50**, 263–338.
- Bradford, M. M. (1976). A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* **72**, 248–254.
- Bryan, J. K. (1977). Molecular weights of protein multimers from polyacrylamide gel electrophoresis. *Anal Biochem* **78**, 513–519.
- Casadaban, M. J. & Cohen, S. N. (1980). Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J Mol Biol* **138**, 179–207.
- Chang, A. C. Y. C. & Cohen, S. N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **134**, 1141–1156.
- Clarke, L. & Carbon, J. (1976). A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *Escherichia coli* genome. *Cell* **9**, 91–99.
- Davies, I. J. & Drabble, W. T. (1996). Stringent and growth-rate-dependent control of the *gua* operon of *Escherichia coli* K-12. *Microbiology* **142**, 2429–2437.
- Ebbole, D. J. & Zalkin, H. (1987). Cloning and characterisation of a 12-gene cluster from *Bacillus subtilis* encoding nine enzymes for *de novo* purine nucleotide synthesis. *J Biol Chem* **262**, 8274–8287.
- Fath, M. J., Mahanty, H. K. & Kolter, R. (1989). Characterisation of a *purF* operon mutation which affects colicin V production. *J Bacteriol* **171**, 3158–3168.
- Fleischmann, R. D., Adams, M. D., White, O. & 37 other authors (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512.
- Giles, I. G. (1992). A computer program to scan DNA sequence databases for the existence of potential probe sequences in DNA. *Biochem Soc Trans* **20**, 292S.
- Gold, L. & Stormo, G. (1987). Translational initiation. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, pp. 1302–1307. Edited by F. C. Neidhardt, J. L. Ingraham, K. Brooks Low, B. Magasanik, M. Schaechter & H. E. Umbarger. Washington, DC: American Society for Microbiology.
- Gots, J. S. & Berberich, M. A. (1965). A structural gene mutation in *Salmonella typhimurium* resulting in repressibility of adenylosuccinase. *Proc Natl Acad Sci USA* **54**, 1254–1261.
- Groisman, E. A., Heffron, F. & Solomon, F. (1992). Molecular genetic analysis of the *Escherichia coli* *phoP* locus. *J Bacteriol* **174**, 486–491.
- Harley, C. B. & Reynolds, R. P. (1987). Analysis of *Escherichia coli* promoters sequences. *Nucleic Acids Res* **15**, 2343–2361.
- Hawley, D. K. & McClure, W. R. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* **11**, 2237–2255.
- He, B. & Zalkin, H. (1992). Repression of *Escherichia coli* *purB* is by a transcriptional roadblock mechanism. *J Bacteriol* **174**, 7121–7127.
- He, B., Shiao, A., Choi, K. Y., Zalkin, H. & Smith, J. M. (1990). Genes of the *Escherichia coli* *pur* regulon are negatively controlled by a repressor-operator interaction. *J Bacteriol* **172**, 4555–4562.
- He, B., Smith, J. M. & Zalkin, H. (1992). *Escherichia coli* *purB* gene: cloning, nucleotide sequence, and regulation by *purR*. *J Bacteriol* **174**, 130–136.
- Helmann, J. D. & Chamberlin, M. J. (1987). DNA sequence analysis suggests that expression of flagellar and chemotaxis genes in *Escherichia coli* and *Salmonella typhimurium* is controlled by an alternative σ factor. *Proc Natl Acad Sci USA* **84**, 6422–6424.
- Kasahara, M., Nakata, A. & Shinagawa, H. (1992). Molecular analysis of the *Escherichia coli* *phoP-phoQ* operon. *J Bacteriol* **174**, 492–498.
- Kohara, Y., Akiyama, K. & Isono, K. (1987). The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**, 495–508.
- Kulakauskas, S., Wilkstrom, P. M. & Berg, D. E. (1991). Efficient introduction of cloned mutant alleles into the *Escherichia coli* chromosome. *J Bacteriol* **173**, 2633–2638.
- Kushner, S. R., Nagaishi, H., Templin, A. & Clark, A. J. (1971). Genetic recombination in *Escherichia coli*: the role of exonuclease I. *Proc Natl Acad Sci USA* **68**, 824–827.
- Mandel, M. & Higa, A. (1970). Calcium-dependent bacteriophage DNA infection. *J Mol Biol* **53**, 159–162.
- Meng, L. M. & Nygaard, P. (1990). Identification of hypoxanthine and guanine as the corepressors for the purine regulon genes of *Escherichia coli*. *Mol Microbiol* **4**, 2187–2192.
- Meng, L. M., Kilstrup, M. & Nygaard, P. (1990). Autoregulation of PurR repressor synthesis and involvement of *purR* in the regulation of *purB*, *purC*, *purL*, *purMN* and *guaBA* expression in *Escherichia coli*. *Eur J Biochem* **187**, 373–379.
- Messing, J. (1991). Cloning in M13 phage or how to use biology at its best. *Gene* **100**, 3–12.
- Miller, S. I. (1991). PhoP/PhoQ: macrophage-specific modulators of *Salmonella* virulence? *Mol Microbiol* **5**, 2073–2078.
- Minton, N. P. (1984). Improved plasmid vectors for the isolation of translational *lac* gene fusions. *Gene* **31**, 269–273.
- Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.
- Prentki, P. & Krisch, H. M. (1984). *In vitro* insertional mutagenesis with a selectable DNA fragment. *Gene* **29**, 303–313.
- Rolfes, R. F. & Zalkin, H. (1988). *Escherichia coli* gene *purR* encoding a repressor protein for purine nucleotide synthesis. Cloning, nucleotide sequence and interaction with the *purF* operator. *J Biol Chem* **263**, 19653–19661.
- Rolfes, R. F. & Zalkin, H. (1990a). Purification of the *Escherichia coli* purine regulon repressor and identification of corepressors. *J Bacteriol* **172**, 5637–5642.
- Rolfes, R. F. & Zalkin, H. (1990b). Autoregulation of *Escherichia coli* *purR* requires two control sites downstream of the promoter. *J Bacteriol* **172**, 5758–5766.
- Rosenberg, M. & Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu Rev Genet* **13**, 319–353.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

- Sasse-Dwight, S. & Gralla, J. D. (1991). Footprinting protein-DNA complexes *in vivo*. *Methods Enzymol* **208**, 146–168.
- Schumacher, M. A., Choi, K. Y., Zalkin, H. & Brennan, R. G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α helices. *Science* **266**, 763–770.
- Sharp, P. M. & Li, W. (1986). Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for ‘rare’ codons. *Nucleic Acids Res* **14**, 7737–7749.
- Staden, R. & McLachlan, A. D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. *Nucleic Acids Res* **10**, 141–156.
- Stormo, G. D., Schneider, T. D., Gold, L. & Ehrenfeucht, A. (1982). Use of the ‘Perceptron’ algorithm to distinguish translational initiation sites in *Escherichia coli*. *Nucleic Acids Res* **10**, 2997–3011.
- Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113–130.
- Sullivan, M. A., Cannon, J. F., Webb, F. H. & Bock, R. (1985). Antisuppressor mutation in *Escherichia coli* defective in the biosynthesis of 5-methylaminomethyl-2-thiouridine. *J Bacteriol* **161**, 368–376.
- Tabor, S. & Richardson, C. C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci USA* **82**, 1074–1078.
- Tesfa-Selase, F. & Drabble, W. T. (1992). Regulation of the *gua* operon of *Escherichia coli* by the DnaA protein. *Mol Gen Genet* **231**, 256–264.
- Tesfa-Selase, F. & Drabble, W. T. (1996). Specific binding of DnaA protein to a DnaA box in the *guaB* gene of *Escherichia coli* K12. *Eur J Biochem* (in press).
- Tinoco, I., Bover, P. N., Denger, B., Levine, M., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nat New Biol* **246**, 40–41.
- Tuerk, C., Gauss, P., Thermes, C., Groebe, D. R., Gayle, M., Guild, N., Stormo, G., D’Aubenton-Carafa, Y., Uhlenbeck, O. C., Tinoco, I., Jr, Brody, E. N. & Gold, L. (1988). CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc Natl Acad Sci USA* **85**, 1364–1368.
- Wilkinson, M. (1991). Purification of RNA. In *Essential Molecular Biology: a Practical Approach*, vol. 1, pp. 77–87. Edited by T. A. Brown. Oxford: Oxford University Press.
- Woods, S. A., Miles, J. S., Roberts, R. E. & Guest, J. R. (1988). Structural and functional relationships between fumarase and aspartase. *FEMS Microbiol Lett* **51**, 181–186.
- Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.

Received 26 March 1996; revised 17 June 1996; accepted 2 July 1996.