Nucleotide Sequence of *bglC*, the Gene Specifying EnzymeII^{bgl} of the PEP:Sugar Phosphotransferase System in *Escherichia coli* K12, and Overexpression of the Gene Product

By HEATHER F. BRAMLEY AND H. L. KORNBERG*

Department of Biochemistry, University of Cambridge, Tennis Court Road,
Cambridge CB2 1QW, UK

(Received 20 August 1986; revised 28 October 1986)

The EnzymeII^{bgl} of the phosphoenolpyruvate- (PEP-) dependent phosphotransferase system catalyses the uptake and concomitant phosphorylation of β -glucosides by *Escherichia coli*; it is specified by the gene *bglC*. The nucleotide sequence of a 3.6 kb *Hind*III restriction fragment spanning *bglC*, cloned on a plasmid, was determined. DNA analysis strongly suggests that the published order of this and other genes involved in β -glucoside utilization, *bglC*, *S*, *B*, is incorrect, and that the regulatory gene *bglS* may be located upstream of the structural genes *bglC* and *bglB*. From the deduced amino acid sequence it is predicted that the membrane protein specified by *bglC* consists of 625 amino acid residues (66.48 kDa). The protein has the hydropathic profile expected of an integral membrane protein (average hydropathy = 0.62). Comparisons between the amino acid sequences deduced for the EnzymeII^{bgl} and for the mannitol-specific EnzymeII^{mtl} show that these proteins are related, and a little direct homology is apparent. A 2.3 kb *Alu*I fragment spanning *bglC* was subcloned into an expression vector which carries the λ p_L promoter and then transformed into a host strain which produces thermolabile cI857 repressor and the anti-terminator N; thermoinduction resulted in the overproduction of a membrane protein and the appearance of Bgl activity.

INTRODUCTION

The PEP-dependent phosphotransferase system (PTS) of *Escherichia coli* K12 and other facultative anaerobic bacteria catalyses the uptake and concomitant phosphorylation of a number of carbohydrates, such as glucose, hexitols, *N*-acetylglucosamine and β -glucosides. These are termed PT sugars (for a recent review, see Postma & Lengeler, 1985). Each PT system comprises the cytoplasmic phosphoproteins EnzymeI and HPr, which are not sugar-specific, and, in addition, one of a range of membrane-associated EnzymesII or EnzymeII/EnzymeIII pairs that are sugar-specific. A phosphate group is transferred from PEP to the incoming sugar via these components as shown in Fig. 1.

The genes encoding the β -glucoside-specific components of the PTS are located at min 83 on the genetic map of *E. coli* K12 (Bachmann, 1983). The *bgl* operon comprises a regulatory region *bglR* and the structural genes *bglC*, *bglS* and *bglB*; of these, *bglC* codes for the EnzymeII^{bgl} (Prasad & Schaefer, 1974). The gene *bglS* specifies a positive regulatory protein that is *trans*-acting in complementation studies (Reynolds *et al.*, 1981), and *bglB* codes for phosphoglucosidase B, which catalyses the hydrolysis of the β -glucoside phosphate, formed from the β -glucoside taken up as shown in Fig. 1. The *bgl* operon is present in an uninducible (cryptic) form in wild-type *E. coli* K12, but the insertion of IS1 or IS5 into *bglR* converts the operon into its inducible form (Reynolds *et al.*, 1981).

Abbreviations: pNP, *p*-nitrophenol; pNPG, *p*-nitrophenyl β -glucoside; ORF, open reading frame; PEP, phosphoenolpyruvate; PTS, phosphotransferase system.

Fig. 1. Representation of the transfer of phosphate groups in the PTS (shown for the β -glucoside PTS) and the linkage of the PTS to intermediary metabolism. No EnzymeIII^{bgl} has yet been identified.

The gene *bglA*, encoding a second phosphoglucosidase A, does not form part of the *bgl* operon, although it occurs at a neighbouring locus. Phosphoglucosidase A catalyses the same reaction as does phosphoglucosidase B, but has slightly different substrate specificities. For example, phosphoglucosidase A does not split salicin phosphate; moreover, BglA is more heat stable than BglB (Wilson & Fox, 1974). The *bglA* gene is expressed constitutively in wild-type *E. coli* K12 (Schaefer & Malamy, 1969).

The 5' transcription initiation site for the *bgl* operon has been located by S1 nuclease mapping and a stretch of the *bglR* region (347 bp long) has been sequenced (Reynolds, 1983).

It is the main aim of this paper to report the DNA sequence of *bglC*, and the amino acid sequence of EnzymeII^{bgl} deduced from it. This sequence is compared with that reported for EnzymeII^{mtl} by Lee & Saier (1983). Our data strongly suggest that the published gene order, *bglR, C, S, B* (Reynolds *et al.*, 1981) is incorrect: it is possible that *bglS* is located upstream of *bglC*. The thermoinducible overproduction of BglC from a λp_L promoter is also described.

METHODS

Bacterial strains and plasmids. The bacterial strains used are derivatives of *E. coli* K12 and are listed in Table 1. Strain JM2137 was made *Srl*⁻ by phage P1-mediated transduction using phage grown on the mutant PB11, which was *recA*, and, highly linked with it, *Tn10::srLA*. Transductants were screened for a *RecA*⁻ phenotype (Maniatis *et al.*, 1982) to give the resultant strain HB12. Plasmid pAR7i is derived from the plasmid pAR7 (kindly given to us by Dr A. Wright, Tufts University, Boston, Mass., USA) by the spontaneous deletion of a 5.5 kb *HpaI* fragment at the site shown in Fig. 2. The *bgl* operon is cryptic in the bacterial strains used. The plasmids carrying the operon in an inducible form are all derived from pAR7 where the *bgl* operon is activated by an IS5 insertion in the *bglR* region (Reynolds, 1983). DNA transformations were performed as described by Hanahan (1983). The construction of strains suitable for overexpression was based on Rosenberg *et al.* (1983). In this case, the 2.3 kb

Table 1. *E. coli* strains and plasmids

Strain or plasmid	Relevant genotype or phenotype	Source
PB11	<i>uvrA srl::Tn10 recA</i>	P. Britton
HB12	<i>lac^Δ (his gnd)^Δ srlA::Tn10 recA</i>	This study from JM2137
AD5230	<i>F⁻ ilv (λc1857 N₇ N₅₃ cro₂₇ pRIS2 [OPQRAJ] galETK) sup^o strA</i>	A. Das
AD5827	<i>F⁻ ilv his proC::Tn10 galOP::IS1 bio (λ Bam N⁺c1857 HI [cro-RAJ-bio-uvrB]) sup^o strA</i>	A. Das
JM2137	<i>(his-gnd)^Δ lac^Δ</i>	M. C. Jones-Mortimer
pAR7	<i>bglR::IS5 bglC⁺ bglS⁺ bglB⁺ Tc^R Ap^R</i>	A. Reynolds
pAR7i	<i>(bglR::IS5)^Δ bglC⁺ bglB⁺ Tc^R Ap^R</i>	Spontaneous <i>HpaI</i> deletion of pAR7
pPM70	<i>lacUV5 Tc^R Ap^R</i>	P. A. Meacock
pHB1	<i>lacUV5-bglB-bglC⁺ Tc^S Ap^R</i>	This study
pHB2	<i>lacUV5-bglC⁺-bglB Tc^S Ap^R</i>	This study
pAD284	<i>pBR322 (BamHI-HindIII)^Δ (λO_LP_L-nutL-N) Ap^R</i>	A. Das
pHB24	<i>pAD284 N⁻ Ap^R (λO_LP_L-nutL-bglC⁺ bglB)</i>	This study
pHB19	<i>pAD284 N⁻ Ap^R (λO_LP_L-nutL-bglB-bglC⁺)</i>	This study

*A*luI fragment (Fig. 2*b*) including all of the DNA represented in Fig. 3 was isolated and ligated into the *HpaI* site of pAD284 to interrupt the *N* antiterminator gene. The plasmid was transformed into AD5230 and the transformants, selected by their ampicillin resistance, were screened for a Gal⁻ phenotype on nutrient agar plates supplemented with 10 mM-galactose and ampicillin (100 µg ml⁻¹) and containing neutral red as indicator. The *gal* genes are situated in this lysogen such that their expression requires the *N* gene product. The plasmids of Gal⁻ strains were mapped with restriction enzymes to determine the orientation of the insert. Suitable plasmids were then transformed into the N⁺ lysogen AD5827, which produces a temperature-sensitive c1857 repressor. Thermoinduction of the transformants resulted in the expression of the insert genes from the λ P_L promoter.

pNP assay. The production of pNP (*p*-nitrophenol) from externally supplied pNPG (*p*-nitrophenyl β -glucoside) was measured spectrophotometrically (Schaefer, 1967) using a modified continuous assay procedure (Elvin, 1983).

Saligenin assay. The assay for saligenin production from salicin (Schaefer, 1967) was modified as follows. Cells were grown to mid-exponential phase in casein hydrolysate (1%, w/v) in the presence or absence of salicin (10 mM) at 37 °C and were then centrifuged (2100 g, 10 min); 2 ml sodium phosphate buffer (pH 7.2) and 0.5 ml 4-aminoantipyrine (0.6%, w/v, in water) were added to 1 ml of the supernatant solution. After 15 min, 0.5 ml potassium ferricyanide (4%, w/v, in water) was added and, after 10 min, the absorbance at 509 nm was measured. Supernatant solutions containing saligenin were red/brown (*A*₅₀₉ approximately 1.6) whereas those not containing saligenin were yellow (*A*₅₀₉ approximately 0.25). Cultures were scored as + or -.

Restriction mapping. Restriction enzymes were obtained from Amersham, and were used according to the manufacturer's instructions, but using a high salt buffer (Maniatis *et al.*, 1982). For *HpaI* digestions, 0.1 M-KCl was also included in the buffer.

DNA preparation and sequencing. The 3.6 kb *HindIII* fragment obtained from pAR7i was purified by electrophoresis through a 1% (w/v) agarose gel and isolated using DE-81 paper (Dretzen *et al.*, 1981). The purified fragment was sonicated to give random fragments, which were cloned into the *SmaI* site of M13mp10 (Bankier & Barrell, 1983). The nucleotide sequence was determined by the dideoxy chain termination method (Sanger *et al.*, 1977). Gradient gels of 0.2–0.6 mm thickness were run in an LKB 2010 Macrophor electrophoresis unit at 55 °C.

Analysis of data. The DNA sequencing data were analysed using the computer programs of Staden (1982*a*). The hydropathy analysis was performed using the SOAP computer program (Kyte & Doolittle, 1982). Amino acid sequence homologies were detected using the DIAGON graphics program (Staden, 1982*b*).

Overexpression. Cultures (50 ml) were grown in Luria both (or casein hydrolysate, 1%, for pNP assays) in the presence of ampicillin (100 µg ml⁻¹). Arbutin (10 mM) was included for the induced PB11(pAR7i) cultures. All cultures were incubated aerobically at 37 °C to mid-exponential phase, when 50 ml fresh medium was added. Thermoinduction was achieved by addition of the fresh medium at 53 °C followed by incubation at 42 °C. Where induction was not required, the fresh medium was at 33 °C and incubation was continued at 33 °C. At the appropriate time after induction, the cultures were harvested (6500 g for 10 min in a Sorvall RC-5B centrifuge), washed in sodium phosphate buffer pH 7.2, and harvested as before.

Membrane and cytoplasm preparations (Henderson & Macpherson, 1986) were resuspended to a final volume of 1 ml and stored at -20 °C.

SDS-PAGE. This was based on the method of Laemmli (1970) with the modifications and SDS dissolving buffer (except with 0.01% bromophenol blue) as described by Macpherson *et al.* (1981). The gels were 15% (w/v)

acrylamide and were 1 mm thick. The following markers were used: bovine serum albumin (M_r 66000), ovalbumin (45000), carbonic anhydrase (29000) and lysozyme (14300). Samples (10–20 μ g) were solubilized at 60 °C for 5 min before loading. Electrophoresis was carried out at 30 mA constant current for 1 h at approx. 18 °C (room temperature).

RESULTS AND DISCUSSION

Construction of strains and plasmids

Plasmid pAR7i was derived from plasmid pAR7 by the spontaneous deletion of a *Hpa*I fragment of approximately 5.5 kb. Restriction mapping showed that the 3.6 kb *Hind*III fragment, covering most of the *bgl* operon (Reynolds, 1983), was unaffected by the deletion (Fig. 2a). In order to show that the 3.6 kb *Hind*III fragment contained the *bgl* structural genes, the fragment was subcloned in both orientations into the *Hind*III site of plasmid pPM70 such that the inserted DNA would be transcribed from the *lac*UV5 promoter of pPM70. The resulting plasmids were transformed into the Bgl⁻ strain HB12. Expression of the *bgl* operon, as measured by the uptake and cleavage of pNPG, is constitutive (Table 2) in HB12(pHB2), which contains the 3.6 kb fragment inserted in the orientation allowing expression of the *bgl* genes from the *lac* promoter (Fig. 2b). This constitutivity is expected as HB12 is not a *lacI*⁻ mutant (Müller-Hill *et al.*, 1968) and therefore does not produce enough of the repressor of the *lac* operon to keep all copies of the multicopy plasmid repressed. However, strain HB12(pHB1), which contains the 3.6 kb fragment inserted in pPM70 in the opposite orientation, also shows a low level of constitutive Bgl activity (Table 2) although the genes cannot be transcribed from the *lac* promoter. Neither of these strains can grow on β -glucosides as sole carbon and energy source. It is possible that some regulatory region of the *bgl* operon is missing from the 3.6 kb fragment. The results of the saligenin assay (Table 2) are consistent with other work (Reynolds, 1983) which suggested that the *Hind*III fragment does not contain a complete *bglB* gene.

Strains AD5827(pHB24) and AD5827(pHB19) were constructed for the overexpression of BglC (see Methods). Restriction mapping showed that plasmid pHB24 contained the 2.3 kb *Alu*I fragment inserted into pAD284 in the orientation suitable for expression of the inserted genes from λ p_L (Fig. 2c) whilst pHB19 contained the insert in the opposite orientation.

The DNA sequence

The 3.6 kb *Hind*III fragment was sequenced with 100% coverage on both strands and with each base covered on average 14 times. Analysis of the DNA sequence revealed the presence of three open reading frames (ORFs). All of the ORFs are on the same strand and ORF2 and 3 are in the same reading frame. The DNA sequence of ORF2 and the preceding intercistronic region, together with the end of ORF1 and the beginning of ORF3, are shown in Fig. 3.

The bglC sequence. The only complete open reading frame (ORF2) codes for a protein of 625 amino acid residues. This is probably the *bglC* gene product (EnzymeII^{bgl}) as this is the only complete ORF in the 3.6 kb *Hind*III fragment known to confer Bgl activity. A possible alternative is that ORF2 may code for some regulatory protein enabling the expression of the cryptic *bglC* on the host cell genome. However this is most unlikely because the polypeptide sequence deduced from the base sequence of ORF2 has properties typical of an integral membrane protein and because it shows some homology with EnzymeII^{mtl} (see later sections) which favours the designation of ORF2 as *bglC*. Seven bases upstream from the initiating ATG there is an AGAGGG sequence which may represent the Shine–Dalgarno ribosomal recognition sequence for *bglC* (Stormo *et al.*, 1982). There are two other initiation codons located near the amino-terminus of the proposed translation product (codons 22 and 51); however, neither of these is preceded by a possible Shine–Dalgarno sequence. The frequency of optimal codon usage index (Ikemura, 1981) is 0.58 whereas that of EnzymeII^{mtl} is 0.80. The low usage of optimal codons in BglC is typical of a protein expressed at a low level in *E. coli*.

A statistical study of positional base preferences (Staden, 1984) within the three phases on this strand indicated that ORF2 is the most likely to be expressed (results not shown) in this region of

Fig. 2. Restriction maps of plasmids pAR7i, pHB2 and pHB24. The sizes of the restriction fragments are shown in kb. (a) pAR7i showing the published gene order of the *bgl* operon and the site of the spontaneous *HpaI* deletion. (b) pHB2. The 3.6 kb *HindIII* fragment is in the correct orientation to allow expression of the *bgl* structural genes from the *lac* promoter. (c) pHB24. The 2.3 kb *AluI* fragment inserted into the *HpaI* site of pAD284. The insert is shown in the correct orientation for expression of the *bgl* genes from the λP_L promoter.

Table 2. Phenotypes of the bacterial strains and plasmids

Results are the means of at least three independent determinations.

Strain	pNP formed [nmol min ⁻¹ (mg dry wt) ⁻¹]		Saligenin assay†	Growth on arbutin or salicin
	+ Inducer*	- Inducer*		
PB11(pBR322)	0	0	—	—
PB11(pAR7i)	19	0	+	+
HB12(pHB1)	3‡	3	—	—
HB12(pHB2)	9‡	9	—	—
HB12	0	0	—	—

* The inducer was arbutin (10 mM).

† Salicin phosphate is hydrolysed only by *BglB* whereas pNPG phosphate can be split by *BglA* or *BglB*. A + result in this assay indicates a functional *BglB*. Salicin is a gratuitous inducer of *bglB* strains (Schaefer & Maas, 1967).

‡ The results were the same with 0.12 mM IPTG as inducer.

Fig. 3. DNA sequence of part of the 3.6 kb *Hind*III fragment spanning *bglC* and showing the deduced amino acid sequence of the gene product. Potential Shine-Dalgarno sequences are boxed. Stop codons are symbolized by three dots. The inverted repeat is underlined.

the DNA. Moreover, the other two phases, and all three phases on the opposite strand, are interrupted frequently by nonsense codons.

The intergenic region. ORF1 is separated from *bglC* by 134 base pairs. Immediately before the start of *bglC* there is a perfect inverted repeat sequence of 15 bases (Fig. 3) of unknown significance.

Fig. 4. Hydropathy plot of the deduced EnzymeII^{bgI} amino acid sequence. The plot was generated using a span of 11 residues plotted every 3 residues. The horizontal line is at -0.4 (the average hydropathy of a soluble protein).

The other ORFs. ORF3 codes for an incomplete protein of 430 amino acids. From the published gene order it would be expected that ORF3 represents *bglS*. However, plasmid pAR7i contains a functional *bglB* (Table 2, saligenin assay) and the complete phosphoglucosidase B is expected to consist of more than 600 amino acids (65 kDa) (Wilson & Fox, 1974). If ORF3 were *bglS* then *bglB* would have to be coded for on the 1.15 kb *EcoRI*–*HindIII* fragment of pAR7i (Fig. 2a), which is impossible if the purified BglB has been correctly sized. If ORF3 were to code for the amino-terminal half of BglB then the complete *bglB* would fit easily onto pAR7i. From previous mapping studies (Reynolds *et al.*, 1981; Prasad & Schaefer, 1974) it was expected that *bglS* (which codes for a positive regulatory protein) would be located between *bglC* and *bglB*. This cannot be correct: there are only 19 base pairs between the termination codon of *bglC* and the initiation codon of ORF3, and there are no suitable ORFs contained within *bglC* or ORF3 in any phase.

ORF1 could code for the carboxy-terminal 80 amino acids of an as yet unidentified protein. The start of transcription for the *bgl* operon has already been located (Reynolds, 1983) upstream of the 3.6 kb *HindIII* fragment. From this it can be calculated that the maximum size of a protein coded by ORF1 is approximately 350 amino acids. This is not large enough to allow ORF1 to represent *bglB*, but it is possible that ORF1 codes for *bglS*. In this case, the intercistronic sequence would separate this regulatory gene from the main structural genes of the operon.

Amino acid sequence of EnzymeII^{bgI}

The amino acid sequence calculated for EnzymeII^{bgI} is shown in Fig. 3. Saier *et al.* (1985) have pointed out that the sugar-specific proteins of the PT systems consist either of an EnzymeII of approximately 65 kDa or of an EnzymeII/III pair with a combined molecular mass of approximately 65 kDa; indeed, the EnzymeII for mannose uptake and its EnzymeIII^{man} have recently (Erni & Zanolari, 1985) been reported to be 27 kDa and 35 kDa proteins respectively. The 66.48 kDa EnzymeII^{bgI} would therefore fall into the former category, along with EnzymeII^{mtl} (68 kDa; Lee & Saier, 1983) and Enzyme II^{N-acetylglucosamine} (65 kDa; Saier, 1985). No EnzymeIII^{bgI} has yet been identified.

The hydropathic profile of EnzymeII^{bgI} is shown in Fig. 4. In this program (Kyte & Doolittle, 1982), the amino acids are scored according to their hydropathy, with isoleucine scoring $+4.5$ and arginine scoring -4.5 . The average hydropathy of EnzymeII^{bgI} is 0.62 , which is well within the range seen for integral membrane proteins (Table 3), and the hydropathic profile reveals alternating hydrophobic and hydrophilic regions. The average hydropathy is fairly constant along the length of the protein; this contrasts with EnzymeII^{mtl}, which has a hydrophobic amino-terminal half, but a carboxy-terminal half typical of a soluble protein (Lee & Saier, 1983).

Table 3. *Average hydropathic index (Kyte & Doolittle, 1982) of several inner-membrane proteins*

Protein	Average hydropathic index	Reference
Bacteriorhodopsin	0.70	Kyte & Doolittle (1982)
EnzymeII ^{mtl}	0.33	Lee & Saier (1983)
EnzymeII ^{bgl}	0.62	This study
MalG	0.83	Dassa & Hofnung (1985)
MalF	0.29	Dassa & Hofnung (1985)

Table 4. *Amino acid sequence homologies between EnzymeII^{bgl} and EnzymeII^{mtl}*

Residue numbers (Bgl²⁶⁰, etc.) locate the homologies in the protein sequences. Exact matches are marked by vertical lines; asterisks denote conservative changes.

A. Bgl ²⁶⁰	Phe	Leu	Leu	Val	Gly	Pro	Leu	Ser	Thr	Trp	Ile	Ser	Glu	Leu	Ile	Ala	Ala	Gly
							*							*	*			
Mtl ₁₄₇	Phe	Leu	Gly	Ile	Gly	Pro	Ile	Val	Glu	Ala	Leu	Ser	Lys	Met	Leu	Ala	Ala	Gly
B. Bgl ¹⁵⁶	Pro	Ile	Ile	Leu	Gly	Tyr	Thr	Ala	Gly	Lys	Arg	Phe	Gly	Gly				
		*	*	*				*										
Mtl ₆₁	Pro	Leu	Leu	Ile	Gly	Tyr	Thr	Gly	Gly	Lys	Leu	Val	Gly	Gly				
C. Bgl ⁶⁰⁸	Pro	His	Gly	Thr														
Mtl ₅₅₃	Pro	His	Gly	Thr														
D. Bgl ³⁵⁶	Ala	Gly	Ser	Ala	Ala	Leu	Thr	Ser	-	Leu	Phe	Gly	Ile	Thr	Glu			
			*			*												
Mtl ₂₄₃	Ala	Gly	Gly	Ala	Ala	Ile	Ile	His	Phe	Leu	Gly	Gly	Ile	His	Glu			

Comparisons of amino acid sequences

The DIAGON algorithm was used to compare the amino acid sequences of EnzymeII^{bgl} and EnzymeII^{mtl}, both with each other and with those of other membrane transport proteins, namely bacteriorhodopsin (Ovchinnikov *et al.*, 1979), AraE and XylE (Maiden *et al.*, 1987) and LacY (Büchel *et al.*, 1980). There was virtually no homology with the sequences of those non-PT proteins. Surprisingly, there appeared also to be little homology between the EnzymesII. There were only three regions that were closely similar in the two proteins. Two of these, labelled A and B in Table 4, are the most extended conserved regions within the sequence. The third, labelled C in Table 4, contains a histidine residue which is of particular interest as, in the PT system, phosphate groups are transferred from one histidine to another (Waygood *et al.*, 1984). This histidine is flanked by a proline and a glycine which could conceivably form a turn in the secondary structure (Chou & Fasman, 1978; Garnier *et al.*, 1978), thus exposing the central histidine.

However, when allowance is made for the effects of small deletions, etc., other similarities also emerge – an example is labelled D in Table 4. The most compelling aspect of these similarities is that they occur in the same pattern throughout the two EnzymesII, including the C-terminal regions (where the hydropathic profiles are quite different), but not in the other sequences tested. This supports the view that the various EnzymesII evolved from a common ancestor (Saier, 1985) and is currently being studied further.

Overexpression of BglC

Thermoinduction of AD5827(pHB24) results in the overproduction of a protein that is located exclusively in the membrane fraction and which is not visible in the membrane preparations of either the uninduced strain AD5827(pHB24) or the induced negative control strains AD5827(pHB19) and AD5230(pAD284) (Fig. 5).

The band of overexpressed protein shows two interesting properties. Firstly, it is diffuse and broad. This property is also shown by the lactose permease (Ehring *et al.*, 1980), but not by all membrane proteins. Secondly, the band of overexpressed protein runs with an electrophoretic mobility equivalent to *M*_r 51 000–58 000. Many membrane proteins such as the lactose permease

Fig. 5. SDS-PAGE of cell fractions (see Methods for details of gel and M_r markers – lanes 1 and 8). Lane 2, induced PB11(pAR7i) membrane; lane 3, uninduced PB11(pAR7i) membrane; lane 4, thermoinduced AD5827(pHB24) membrane (in correct orientation with respect to λp_L promoter); lane 5, uninduced AD5827(pHB24); lane 6, thermoinduced AD5827(pHB19) membrane (in incorrect orientation); lane 7, thermoinduced AD5230(pAD284) membrane (N^+ vector with no insert); lane 9, thermoinduced AD5827(pHB24) cytoplasm; lane 10, uninduced AD5827(pHB24) cytoplasm.

(Ehring *et al.*, 1980), cytochrome oxidase (Bonitz *et al.*, 1980) and EnzymeII^{ml} (Lee & Saier, 1983), have been found to run at positions that indicate anomalously low M_r values. This has been attributed to the increased binding of SDS to the hydrophobic proteins (Rizzolo *et al.*, 1976). The discrepancy of 10000–15000 between the apparent and calculated M_r of *BglC* is a typical difference.

There is no significant difference in the gel pattern of membrane proteins from uninduced and induced cultures of strain PB11(pAR7i). Detection of the *BglC* protein requires its overexpression.

Thermoinduction of the overexpression strain AD5827(pHB24) resulted also in the appearance of *Bgl* activity in the pNP assay [4.8 ± 1.2 nmol pNP formed min^{-1} (mg dry wt) $^{-1}$ mean $\pm$ SD, ($n = 12$)], with no significant difference in activity between cultures harvested at time points from 10 min to 1 h after induction. Prior to thermoinduction the cells gave very little activity in this assay [0.9 ± 0.3 nmol pNP formed min^{-1} (mg dry wt) $^{-1}$ ($n = 5$)]. The thermoinduced activity was less than that of the induced non-overproducing strain PB11(pAR7i) [19 ± 4.3 nmol pNP formed min^{-1} (mg dry wt) $^{-1}$ ($n = 14$)]. One possible explanation for this is that some of the overexpressed protein may be located in the outer membrane (Yatvin *et al.*, 1986). Since Wilson & Fox (1974) have shown that non-phosphorylated pNPG is not a substrate for the phosphoglucosidases *BglB* and *BglA*, the activity detected must be due to *BglC*.

The results of the overexpression experiments provide further evidence that the sequence given as ORF2 represents that of *bglC*.

We thank Drs M. C. Jones-Mortimer, P. J. F. Henderson, T. J. Gibson, D. C. M. Moore, C. J. Howe and E. O. Davis, and Mr M. C. J. Maiden, for much helpful advice. This work was done during the tenure by H. F. B. of a studentship of the Medical Research Council, and was supported in part by Grant no. GR/C/97286 from the Science and Engineering Research Council, and in part by a personal grant-in-aid to H. L. K. from Unilever Research Ltd. DNA sequencing gels were run on an apparatus donated to Dr P. J. F. Henderson by the SmithKline Foundation.

REFERENCES

- BACHMANN, B. J. (1983). Linkage map of *Escherichia coli* K-12, edition 7. *Microbiological Reviews* **47**, 180–230.
- BANKIER, A. T. & BARRELL, B. G. (1983). Shotgun DNA sequencing. *Nucleic Acid Biochemistry* B508. 1–34. In *Techniques in the Life Sciences B5*. Edited by R. A. Flavell. Elsevier Scientific Publishers Ireland Ltd.
- BONITZ, S. G., CORUZZI, G., THALENFELD, B. E., TZAGOLOFF, A. & MACINO, G. (1980). Assembly of the mitochondrial membrane system. Structure and nucleotide sequence of the gene coding for subunit I of yeast cytochrome oxidase. *Journal of Biological Chemistry* **255**, 11927–11941.
- BÜCHEL, D. E., GRONENBORN, B. & MÜLLER-HILL, B. (1980). Sequence of the lactose permease gene. *Nature, London* **283**, 541–545.
- CHOU, P. Y. & FASMAN, G. D. (1978). Empirical predictions of protein conformation. *Annual Review of Biochemistry* **47**, 251–276.
- DASSA, E. & HOFNUNG, M. (1985). Sequence of gene *malE* in *Escherichia coli* K-12: homologies between integral membrane components from binding-protein dependent transport systems. *EMBO Journal* **4**, 2287–2293.
- DRETZEN, G., BELLARD, M., SARSONNE-CORSI, P. & CHAMBON, P. (1981). A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Analytical Biochemistry* **112**, 295–298.
- EHRING, R., BEYREUTHER, K., WRIGHT, J. K. & OVERATH, P. (1980). *In vitro* and *in vivo* products of *Escherichia coli* lactose permease gene are identical. *Nature, London* **283**, 537–540.
- ELVIN, C. M. (1983). *Aspects of carbohydrate transport by Escherichia coli*. PhD Thesis, University of Cambridge, UK.
- ERNI, B. & ZANOLARI, B. (1985). The mannose-permease of the bacterial phosphotransferase system. Gene cloning and purification of the EnzymeII^{Man}/III^{Man} complex of *Escherichia coli*. *Journal of Biological Chemistry* **260**, 15495–15503.
- GARNIER, J., OSGUTHORPE, D. J. & ROBSON, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology* **120**, 97–120.
- HANAHAN, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**, 557–580.
- HENDERSON, P. J. F. & MACPHERSON, A. J. S. (1986). The assay, genetics, proteins and reconstitution of proton-linked galactose, arabinose and xylose transport systems of *Escherichia coli*. *Methods in Enzymology* **125**, 387–429.
- IKEMURA, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optional for the *E. coli* translational system. *Journal of Molecular Biology* **151**, 389–409.
- KYTE, J. & DOOLITTLE, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105–132.
- LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, London* **227**, 680–685.
- LEE, C. A. & SAIER, M. H., JR (1983). Mannitol-specific enzyme II of the bacterial phosphotransferase system. III. The nucleotide sequence of the permease gene. *Journal of Biological Chemistry* **258**, 10761–10767.
- MACPHERSON, A. J. S., JONES-MORTIMER, M. C. & HENDERSON, P. J. F. (1981). Identification of the AraE transport protein of *Escherichia coli*. *Biochemical Journal* **196**, 269–283.
- MAIDEN, M. C. J., DAVIS, E. O., BALDWIN, S. A., MOORE, D. C. M. & HENDERSON, P. J. F. (1987). Mammalian and bacterial sugar transport proteins are homologous. *Nature, London* (in the Press).
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MÜLLER-HILL, B., CRAPO, L. & GILBERT, W. (1968). Mutants that make more *lac* repressor. *Proceedings of the National Academy of Sciences of the United States of America* **59**, 1259–1264.
- OVCHINNIKOV, Y. A., ABDULAEV, N. G., FEIGINA, M. Y., KISELEV, A. V. & LOBANOV, N. A. (1979). The structural basis of the functioning of bacteriorhodopsin: an overview. *FEBS Letters* **100**, 219–224.
- POSTMA, P. W. & LENGELER, J. W. (1985). Phosphoenolpyruvate:carbohydrate phosphotransferase system of bacteria. *Microbiological Reviews* **49**, 232–269.
- PRASAD, I. & SCHAEFLER, S. (1974). Regulation of the β -glucoside system in *Escherichia coli* K-12. *Journal of Bacteriology* **120**, 638–650.
- REYNOLDS, A. E., FELTON, J. & WRIGHT, A. (1981). Insertion of DNA activates the cryptic *bgl* operon in *Escherichia coli* K-12. *Nature, London* **293**, 625–629.
- REYNOLDS, A. E. (1983). *Characterization of mutations which activate the cryptic β -glucoside operon in Escherichia coli K-12*. PhD thesis, Tufts University, Massachusetts.
- RIZZOLO, L. J., LE MAIRE, M., REYNOLDS, J. A. & TANFORD, C. (1976). Molecular weights and hydrophobicity of the polypeptide chain of sarcoplasmic reticulum calcium(II) adenosine triphosphatase and of its primary tryptic fragments. *Biochemistry* **15**, 3433–3437.

- ROSENBERG, M., HO, Y. & SHATZMAN, A. (1983). The use of pKC30 and its derivatives for controlled expression of genes. *Methods in Enzymology* **101**, 123–138.
- SAIER, M. H., JR (1985). *Mechanisms and Regulation of Carbohydrate Transport in Bacteria*. New York: Academic Press.
- SAIER, M. H., GRENIER, F. C., LEE, C. A. & WAYGOOD, E. B. (1985). Evidence for the evolutionary relatedness of the proteins of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *Journal of Cellular Biochemistry* **27**, 43–56.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- SCHAEFLER, S. (1967). Inducible system for the utilization of β -glucosides in *Escherichia coli*. I. Active transport and utilization of β -glucosides. *Journal of Bacteriology* **93**, 254–263.
- SCHAEFLER, S. & MAAS, W. K. (1967). Inducible system for the utilization of β -glucosides in *Escherichia coli*. II. Description of mutant types and genetic analysis. *Journal of Bacteriology* **93**, 264–272.
- SCHAEFLER, S. & MALAMY, A. (1969). Taxonomic investigations on expressed and cryptic phospho- β -glucosidases in *Enterobacteriaceae*. *Journal of Bacteriology* **99**, 422–433.
- STADEN, R. (1982a). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **10**, 4731–4751.
- STADEN, R. (1982b). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Research* **10**, 2951–2961.
- STADEN, R. (1982b). An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Research* **10**, 2951–2961.
- STORMO, G. D., SCHNEIDER, T. D. & GOLD, L. M. (1982). Characterization of translational initiation sites in *Escherichia coli*. *Nucleic Acids Research* **10**, 2971–2996.
- WAYGOOD, E. B., MATTOO, R. L. & PERI, K. G. (1984). Phosphoproteins and the phosphoenolpyruvate:sugar phosphotransferase system in *Salmonella typhimurium* and *Escherichia coli*: evidence for III^{Mannose}, III^{Fructose}, III^{Glucitol}, and the phosphorylation of Enzyme II^{Mannitol} and Enzyme II^{N-acetylglucosamine}. *Journal of Cellular Biochemistry* **25**, 139–159.
- WILSON, G. & FOX, C. F. (1974). The β -glucoside system of *Escherichia coli*. IV. Purification and properties of phospho- β -glucosidases A and B. *Journal of Biological Chemistry* **249**, 5586–5598.
- YATVIN, M. B., SMITH, K. M. & SIEGEL, F. L. (1986). Translocation of nascent non-signal sequence proteins in heated *Escherichia coli*. *Journal of Biological Chemistry* **261**, 8070–8075.