Analysis of the β-glucoside utilization (*bgl*) genes of *Shigella sonnei*: evolutionary implications for their maintenance in a cryptic state

Arun S. Kharat and S. Mahadevan

Author for correspondence: S. Mahadevan. Tel: +91 80 309 2607. Fax: +91 80 360 0683.
e-mail: mahi@serc.iisc.ernet.in

Department of Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bangalore 560012, India

The pattern of expression of the genes involved in the utilization of aryl β-glucosides such as arbutin and salicin is different in the genus *Shigella* compared to *Escherichia coli*. The results presented here indicate that the homologue of the cryptic *bgl* operon of *E. coli* is conserved in *Shigella sonnei* and is the primary system involved in β-glucoside utilization in the organism. The organization of the *bgl* genes in *S. sonnei* is similar to that of *E. coli*; however there are three major differences in terms of their pattern of expression. (i) The *bglB* gene, encoding phospho-β-glucosidase B, is insertionally inactivated in *S. sonnei*. As a result, mutational activation of the silent *bgl* promoter confers an Arbutin-positive (Arb⁺) phenotype to the cells in a single step; however, acquiring a Salicin-positive (Sal⁺) phenotype requires the reversion or suppression of the *bglB* mutation in addition. (ii) Unlike in *E. coli*, a majority of the activating mutations (conferring the Arb⁺ phenotype) map within the unlinked *hns* locus, whereas activation of the *E. coli bgl* operon under the same conditions is predominantly due to insertions within the *bglR* locus. (iii) Although the *bgl* promoter is silent in the wild-type strain of *S. sonnei* (as in the case of *E. coli)*, transcriptional and functional analyses indicated a higher basal level of transcription of the downstream genes. This was correlated with a 1 bp deletion within the putative Rho-independent terminator present in the leader sequence preceding the homologue of the *bglG* gene. The possible evolutionary implications of these differences for the maintenance of the genes in the cryptic state are discussed.

Keywords: cryptic genes, transcriptional activation, insertion elements, antitermination

INTRODUCTION

Members of the family *Enterobacteriaceae* exhibit substantial differences in their ability to utilize β-glucosides such as salicin, arbutin and cellobiose. Though earlier studies have confirmed this at the biochemical level (Schaefler & Malamy, 1969), the genetic basis for this difference is not known.

Among the four different genetic systems involved in β-glucoside utilization in *Escherichia coli* (Prasad & Schaefler, 1974; Kricker & Hall, 1987; Parker & Hall, 1988; Hall & Xu, 1992), the *bgl* operon, encoding the functions necessary for the transport and metabolism of arbutin and salicin, is best characterized. An unusual feature of the *E. coli bgl* operon is its crypticity. In wild-type cells, the structural genes of the operon are maintained in a functional form, but they are neither expressed nor are they inducible. Though wild-type cells are phenotypically Bgl⁻, a single mutational event can lead to the simultaneous acquisition of an Arb⁺ Sal⁺ (arbutin and salicin-positive) phenotype (Schaefler & Malamy, 1969). Mutations that activate the operon map predominantly within the regulatory locus *bglR* and in most cases are caused by insertion of IS elements upstream of the promoter, leading to enhancement of transcription from the *bgl* promoter (Reynolds *et al.*,

The GenBank accession number for the *bglR–bglG* sequence reported in this paper is AF183894.

1981, 1986; Schnetz & Rak, 1992). Transcriptional activation is mediated by the disruption of negative elements located near the promoter (Lopilato & Wright, 1990; Schnetz, 1995; Singh *et al.*, 1995; Mukerji & Mahadevan, 1997; Caramel & Schnetz, 1998). Activation results in the expression of a phospho-*enol*-pyruvate-dependent phosphotransferase (encoded by the *bglF* gene) and a phospho-*β*-glucosidase (encoded by *bglB*) which can cleave salicin and arbutin (Fox & Wilson, 1967; Prasad & Schaefler, 1974). In addition, *E. coli* also constitutively expresses the enzyme phospho-*β*-glucosidase A, specific for arbutin, encoded by the unlinked *bglA* locus (Prasad *et al.*, 1973). In the presence of the phosphotransferase encoded by *bglF*, cells can exhibit an Arb^+ phenotype even when the *bglB* gene of the operon is non-functional.

Regulation of the operon in response to the availability of *β*-glucosides is brought about by a mechanism involving antitermination of transcription, mediated by the *bglG* and *bglF* gene products. BglG and BglF represent a novel class of regulatory systems, with the BglF permease acting as the sensor and the BglG antiterminator functioning as the response regulator (Amster-Choder & Wright, 1993; Mahadevan, 1997; Rutberg, 1997 for a review).

Even though recent studies on the transcriptional activation of the *bgl* operon have provided a wealth of information on the molecular mechanisms involved in its silencing, from an evolutionary point of view, its crypticity remains a puzzle. This is made more enigmatic by the differential capabilities of closely related members of the *Enterobacteriaceae* to utilize *β*-glucosides. Earlier biochemical studies on five different isolates of *Shigella sonnei* showed that they fall into two classes, class I strains that are capable of mutating to Sal^+ in two steps and class II strains that are unable to mutate to Sal^+ even after prolonged incubation (Schaefler & Malamy, 1969). To date, no information at the genetic or molecular level is available that can offer a satisfactory explanation for these differences. The results of our studies reported here provide information about the differences in the organization and expression of the *bgl* genes in *S. sonnei* in comparison to those of *E. coli*. Understanding the divergence of the *bgl* genes at the functional and molecular level in these two closely related organisms is of significance in gaining insights on the reasons for the maintenance of the operon in a cryptic state against evolutionary pressure.

METHODS

Isolation of *S. sonnei* strain AK1. Strain AK1 was isolated from the town sewer line, and was identified and characterized using standard microbiological procedures (Falkow & Mekalanos 1990; Watanabe & Okamura, 1991). The sewage sample was collected and allowed to sediment for 6 h at room temperature. The clear supernatant was streaked on MacConkey lactose medium and incubated at 37 °C for 18 h. Lac^- colonies were checked for growth on deoxycholate/citrate medium and confirmed by growth on *Shigella*/*Salmonella* agar. A single isolate obtained was checked for biochemical, morphological and culture characteristics and was confirmed as *Shigella sonnei*.

Bacterial strains and plasmids. The strains and plasmids used in this study are listed in Table 1.

Isolation of Bgl^+ mutants of AK1. Strain AK1 (Arb^- Sal^-) was streaked on MacConkey arbutin plates, incubated at 37 °C for 96 h and allowed to papillate. One of the papillae showing a strong Arb^+ phenotype was purified on a MacConkey arbutin plate. Strain AK101, derived from the papilla, is Arb^+ but remained Sal^-. In the next step, strain AK101 was streaked on a MacConkey salicin plate. The plates were incubated at 37 °C for 120 h to allow papillation. The Arb^+ Sal^+ strain AK102 was derived from one such papilla. Approximate mutation frequencies were calculated from the papillation frequencies. In an independent experiment, 16 Arb^+ mutants were isolated from papillae on MacConkey arbutin plates for detailed analysis of the spectrum of activating mutations.

DNA manipulations. Plasmid isolation, DNA manipulations and Southern analyses were carried out as described by Sambrook *et al.* (1989). Isolation of genomic DNA was carried out with minor modifications of a published protocol (Owen & Borman, 1987).

PCR analysis. Oligonucleotide primers used to amplify genomic DNA from *Shigella sonnei* and *E. coli* strains were: SM1 (forward, −250), 5′-GTGGATCCATCTTCTACTAC-GTGAAG-3′; SM2 (reverse, +340), 5′-AGGAATTCGAC-TTAAGAGTTCGCTTA-3′. PCR products were cloned into pUC19 at the *Bam*HI and *Eco*RI restriction sites. Plasmids used in *in vivo* transcription assays were obtained by subcloning *Bam*HI–*Eco*RI fragments from the above clones in pBR322.

DNA sequencing. Sequencing of both strands was carried out by the dideoxy chain-termination method using Sequenase, version 2 (US Biochemicals) as described in the manufacturer's protocol. Sequence was analysed manually as well as using the ScanJet 4C sequence-analysis and gel-scanning software (Kodak).

Analysis of the insertion element within *bglB*. Oligonucleotide primers used to amplify the insertion element along with overlapping *bglB* region from *S. sonnei* were: SM49 (forward, +3634), 5′-CACCGTTACCCGGAAGAT-3′ and SM50 (reverse, +3810), 5′-CAGCGGCTTGATCCCCGC-3′.

PCR products were purified and used for sequencing the flanking regions and part of the insertion element using *Taq* DNA polymerase (Gibco BRL) as described in the manufacturer's protocol.

Transcriptional analysis.

S1 nuclease protection assay. Analysis of transcription *in vivo* was carried out using oligonucleotides complementary to *bgl* and *bla* transcripts, as described previously (Singh *et al.*, 1995; Mukerji & Mahadevan, 1997). RNA extraction and S1 nuclease assays were carried out using the same procedure. The probes used in the assay are described below. The protected fragments corresponding to the *bgl* and *bla* transcripts were visualized after autoradiography. The transcripts were quantitated by measuring the radioactivity corresponding to the bands. The data shown are representative of at least three independent experiments in each case.

The oligonucleotide probes used in the *bgl* initiation studies are complementary to *bgl* sequence from −10 to +42. Since

Table 1. Bacterial strains and plasmids

Strain/plasmid	Genotype/description	Source/reference
***E. coli* strains**		
RV	F^- Δ*lac*X74 *thi bgl*R^0 *bglG*$^+$*bglF*$^+$ *bglB*$^+$ (Arb$^-$ Sal$^-$)	A. Wright*
AE328	As RV, *tna*::Tn*10 bgl*R::IS*1* (Arb$^+$ Sal$^+$ Tet^R)	A. Wright*
JF201	F^- Δ*lac*X74 Δ(*bgl–pho*)201 *ara thi gyr*A	Reynolds *et al.* (1981)
DH5α	F′ *endA1 hsdR17* (r_k^- m_k^-) *supE44 thi-1 recA1 gyr*A (NalR) *rel*A1Δ (*lacZYA*−*arg*F) *U169deo*R[ϕ80d*lac*Δ(*lac*Z)M15]	Woodcock *et al.* (1989)
***S. sonnei* strains**		
AK1	*bglR*^{0} *bglG*$^+$ *bglF*$^+$ *bglB* (Arb$^-$ Sal$^-$)	This work
AK101	As AK1, *bglR*::IS*1* (Arb$^+$ Sal$^-$)	This work
AK102	As AK101 (Arb$^+$ Sal$^+$)	This work
AK41-56	As AK1, (Arb$^+$ Sal$^-$)	This work
SSOR	*bglR*^{0} *bglG*$^+$ *bglF*$^+$ *bglB* (Arb$^-$ Sal$^-$) Amp^R Tet^R Kan^R	Kharat & Mahadevan (1999)
CR$^+$	*bglR*^{0} *bglG*$^+$ *bglF*$^+$ *bglB* (Arb$^-$ Sal$^-$) Tet^R Kan^R	R. Roy†
CR$^-$	*bglR*^{0} *bglG*$^+$ *bglF*$^+$ *bglB* (Arb$^-$ Sal$^-$) Tet^R Kan^R Amp^R	R. Roy†
Plasmids		
E. coli plasmids		
pMN300	pBR322 (*Ava*I–*Eco*RI insert); *bglR*^{0} *bglG*$^+$ *bglF*$^+$ *bglB*$^+$	Singh *et al.* (1995)
pMN22AE	pBR322 (*Eco*RI insert); *bglR*::IS*1* *bglG*$^+$ *bglF*$^+$ *bglB*$^+$	Singh *et al.* (1995)
pMN5	pBR322 (*Cla*I–*Hin*dIII insert); *′bglG bglF*$^+$ *bglB′*	Mahadevan *et al.* (1987)
p1H	pBR322 (*Eco*RI–*Hin*dIII insert); *′bglG bglF*$^+$ *bglB*$^+$	Mahadevan *et al.* (1987)
p6J	pBR322 (*Eco*RI insert); *′bglF bglB*$^+$	Mahadevan *et al.* (1987)
pANS4-11	pUC19 (*Eco*RI–*Bam*HI insert); *bglR*^{0} *bglG′*	This work
pHMG409	pLG339 (*Eco*RI–*Stu*I insert); *hns*$^+$	Goransson *et al.* (1990)
pTZUng2	pTZ19R (*Sma*I insert); *ung*$^+$	Varshney *et al.* (1988)
S. sonnei plasmids		
pANS1-3	pUC19 (*Eco*RI–*Bam*HI insert); *bglR*^{0} *bglG′*	This work
pANS5-16	pBR322 (*Eco*RI–*Bam*HI insert); *bglR*::IS*1* *bglG′*	This work

* Tufts University Medical School, Boston, MA, USA.

† Astra–Zeneca, Bangalore, India.

there is a transversion at position +25 in the *S. sonnei* sequence, the probes for *E. coli* (probe I) and *S. sonnei* (probe II) differ at this position, indicated by a bold letter (see below). These probes carry a 10 nt noncomplementary region to allow differentiation of the undigested probe from the signal from the protected fragment.

The oligonucleotide probe to quantitate the *bglG*-specific mRNA (probe III) is complementary to the *bgl*G sequence from +120 to +159. As the *S. sonnei* and *E. coli* nucleotide sequences do not show any differences in this region, the same probe was used in the S1 nuclease protection assay in these two organisms. The 3′ end of the probe carries an 8 nt random sequence that was deliberately introduced to differentiate between the undigested probe and the signal corresponding to the protected fragment. The probe concentration has been standardized previously (Singh *et al.*, 1995) to ensure that it is in molar excess of the mRNA.

The probe I sequence is 5′-CCAGTCATTTATTAATG**G**T-TTTTATAACGAACATCCAGGTTCGGAAATTAAT-3′; the probe II sequence is 5′-CCAGTCATTTATTAATG**T**T-TTTTATAACGAACATCCAGGTTCGGAAATTAAT-3′; probe III sequence 5′-GTTGAGAATTTTGGTGATTTGC-ATGTTCATAGCAAGGACCAACATCGT-3′.

Northern blot analysis. Estimation of *bglB* transcription was carried out by Northern blot analysis as described by Sambrook *et al.* (1989). Total RNA was extracted from the strains JF201 (Δ*bgl*), AK1 (Arb$^-$ Sal$^-$), AK101 (Arb$^+$ Sal$^-$) and AK102 (Arb$^+$ Sal$^+$). RNA (20 μg) was used for blotting. The blots were probed using a probe specific for the *E. coli bglB* gene, obtained by PCR amplification. A DNA fragment carrying the *E. coli ung* gene was used to probe the same blot as an internal control. Transcripts were quantitated by scanning the bands obtained after autoradiography using a Hewlett Packard ScanJet 4C with gel-scanning software from Kodak. The *bglB* transcript was normalized by determining the *bgl*/*ung* ratio.

Assay for *bglB* activity (saliginine assay). Measurement of phospho-β-glucosidase B activity, specific for the *bglB* gene, was carried out using a procedure that was similar to the assay described by Prasad & Schaefler (1974). Cultures were grown in LB at 37 °C with or without addition of 7 mM salicin as inducer. Cells were harvested, washed with 0·85% physiological saline solution and resuspended in 0·1 vol of the same solution. To 0·1 ml cells, an equal volume of 140 mM salicin was added and the tubes were incubated at 37 °C for 30 min. The reaction was stopped by the addition of 0·5 ml 2 M Na_2CO_3 solution. After addition of 0·5 ml 0·6% 4-aminoantipyrine solution, the tubes were incubated at room temperature for another 15 min, followed by the addition of 0·5 ml 4% $K_3[Fe(CN)_6]$ solution. The red colour obtained within 10 min was measured spectrophotometrically at 509 nm. The activity, expressed in units, was calculated as described by Mahadevan & Wright (1987).

RESULTS

Isolation of Bgl$^+$ mutants of *S. sonnei* strain AK1

The wild-type *S. sonnei* strain AK1, a sewage isolate used in this study, cannot utilize the aryl β-glucosides arbutin and salicin. Colonies of the strain remain pale yellow on MacConkey arbutin and MacConkey salicin indicator medium even after 72 h. However, on MacConkey arbutin plates, reddish brown papillae were obtained after 96 h incubation at 37 °C. Strain AK101, isolated from one such papilla, showed an Arb$^+$ phenotype, but remained Sal$^-$ on MacConkey salicin indicator plates. Upon prolonged incubation of the plates, pale red papillae were observed after 108–120 h. Strain AK102, derived from one such papilla, formed red colonies within 16 h on MacConkey salicin plates and is therefore a salicin-positive derivative of AK1, obtained in two steps. Strain AK1 is similar to the three *S. sonnei* strains out of five described by Schaefler & Malamy (1969) that can be activated to give a Sal$^+$ phenotype in two steps. Approximate estimates of mutation frequencies are 10^{-8} (WT to Arb$^+$) and 10^{-12} (Arb$^+$ to Sal$^+$). This property of *S. sonnei* is different from that of *E. coli* strains, which can papillate to Sal$^+$ in a single step within 48 h, suggesting differences in the organization of the *bgl* genes in the two organisms. This is consistent with an earlier observation that three out of five *S. sonnei* strains tested showed a pattern of two-step activation (Schaefler & Malamy, 1969).

Identification of the *bgl* genes of AK1

E. coli has at least three different cryptic systems for the utilization of salicin. Since no reports regarding the nature of the genes involved in β-glucoside utilization in *S. sonnei* are available, we attempted to characterize the genes responsible for conferring a Bgl$^+$ phenotype to strain AK1. The time involved in mutating to Bgl$^+$ suggested the involvement of the homologue of the *E. coli bgl* operon since the activation of the *cel* and *asc* operons requires incubation for longer time periods in *E. coli* strains (Kricker & Hall, 1984; Hall & Xu, 1992). This was confirmed by Southern analysis of genomic DNA from AK1 and its activated derivative AK101 (data not shown). The signal obtained from *S. sonnei* DNA was as strong as that from *E. coli*, suggesting a high degree of homology between the *bgl* sequences in the two organisms.

```
-245 ATCTTCTACTACGTGAAGAGGCAGGATTTCCGTCATGTAGGCGGTGATGA
-195 GCTGGATAAACTGCTGGCGGGAAAAGATAGCGACAAATAATTCACCAGAC
-145 AAATCCCAATAACTTAATTATTGGGATTTGTTATATATAACTTTATAAAT
 -95 TCCTAAAATTACACAAAGTTAATAACTGCGAGCATGGTCATATTTTTATC
                                                  +1
 -45 AATAGCGCATTGCTATTTTCTCTGCACGCAATTAAATTAATTTCCGAACC
   6 TGGATGTTCGTTATAAAAAACATTAATAAATGACTGGATTGTTACTGCAT
  56 TCGCAGGCAAAACCTGACATAACCAGAGAATACTGGTGAAGTCGGTTTTT
 106 TTGTTTATAAAAAAGGTCCTTGCTATGAACATGCAAATCACCAAAATTCT
                    S D
 156 CAACAATAATGTTGTGGTGGTTACTGATGATCAACAGCGGGAAAAAGTC
 206 GTCATGGGGCGAGGAATTGGCTTTCAAAAATGCGCTGGCGAAAGAATTAA
 256 CTCAAGTGGAATAGAAAAAGAGTATGCCTTGAGCAGTCATGAACTGAACG
 306 GGCGATTAAGCAAACTCTTAAGTCATATTCCTCT
```

Fig. 1. Nucleotide sequence of *bglR* region from *S. sonnei* strain AK1. Bases marked with arrows indicate differences seen from the *E. coli* sequence. The inverted repeat within *bglR*, one of the silencer elements characterized in *E. coli*, is indicated by broken arrows. The Rho-independent terminator preceding *bglG* is indicated by solid arrows. The position of the IS*1* insertion is shown by an open inverted triangle and the 9 bp duplication flanking the insertion is underlined. The filled triangle indicates the position of the mutation at +100. The putative promoter elements are boxed.

PCR amplification and nucleotide sequence analysis of the *bglR* region

Since the Southern analysis suggested the presence of a *bgl* homologue in *S. sonnei*, we attempted to further confirm this by amplifying the regulatory region using PCR and characterizing the clones from the wild-type and mutant strains. The *bgl* regulatory region was amplified using primers designed on the basis of the *E. coli bglR* sequence. Analysis of the PCR fragments on agarose gels showed that the sizes of the fragments amplified from genomic DNA of *S. sonnei* are similar to those obtained from *E. coli* controls, confirming the earlier results of the Southern analysis. The size of the fragments obtained from the Arb$^+$ mutant AK101 showed an increase of about 800 bp, consistent with the possibility that the mutant carries an insertion of DNA within the regulatory region.

Clones containing the PCR fragments derived from the wild-type and Bgl$^+$ mutant strains of *S. sonnei* were subjected to nucleotide sequence analysis to determine the extent of homology at the nucleotide level, and the site and nature of the insertion. The sequence of the *bglR* region and a part of the *bglG* gene are presented in Fig. 1. The sequence of the *bglR* region in the case of the wild-type strain AK1 differs only in two positions from the *E. coli* sequence (Schnetz *et al.*, 1987). Interestingly, one of the alterations is within the stem of the putative Rho-independent terminator in the leader sequence preceding the first structural gene. A similar mutation, deletion of a G residue at position +100 generated after

mutagenesis, was earlier shown to result in partially constitutive expression of a reporter gene and increased readthrough of transcription *in vitro* in *E. coli* (Mahadevan & Wright, 1987). The functional significance of this change is discussed below. The second difference is a C to A transversion at position +25 within the leader region. The partial sequence of the first structural gene is identical to that of the *E. coli bglG* gene (with the exception a transversion and two transitions), suggesting that the *S. sonnei bgl* operon may also be regulated by a mechanism involving antitermination of transcription.

Sequence analysis of the PCR fragment from the Arb^+ strain AK101 showed the presence of an extra sequence of DNA within *bglR*. The partial sequence of this element suggests that it is IS*1*, reported previously in *S. sonnei* (Lawrence *et al.*, 1992) (data not shown). The site of insertion (Fig. 1) also suggests that the *S. sonnei bgl* operon may also be activated by a mechanism similar to that of *E. coli*. The 9 bp target duplication brought about by IS*1* is underlined in Fig. 1.

Analysis of the pattern of activation of the *bgl* operon in *S. sonnei*

In *E. coli*, 95–98 % of the mutations that activate the *bgl* operon are linked to the operon and only 2–5 % of the activating mutations are unlinked, involving loci such as *hns*, *gyr*, *bglJ* and *leu*O (Defez & DeFelice, 1981; DiNardo *et al.*, 1982; Giel *et al.*, 1996; Ueguchi *et al.*, 1998). To check whether the pattern of activation is identical in *S. sonnei*, the 16 independently isolated Arb^+ mutants of AK1 (see Methods) were analysed to determine the nature of the activating mutation. Genomic DNA from the mutants was used for PCR amplification and Southern analysis of the *bglR* region. One mutant showed a pattern similar to strain AK101 that carries an IS*1* insertion in *bglR*. The remaining 15 mutant strains showed a pattern similar to the parent strain AK1, indicating the absence of insertions within *bglR*. To examine whether the activating mutations in these mutants are located within any of the known loci, plasmid pHMG409 (Goransson *et al.*, 1990), which carries the wild-type copy of the *hns* gene, was introduced into the mutants by transformation. In 14 mutants, the presence of the wild-type copy of the *hns* gene resulted in an Arb^- phenotype, suggesting that the activating mutations map within the *hns* gene. The location of the mutation in one strain remains unknown. Therefore, although the *bgl* genes share a high degree of homology at the nucleotide level and genetic organization in these two organisms, their pattern of activation is different.

To confirm that the results described above are not confined to strain AK1, similar analyses were carried out using three clinical isolates of *S. sonnei*: SSOR, CR^+ and CR^-. PCR, Southern and nucleotide sequence analysis of the *bglR* region revealed that the organization and the pattern of activation are similar in these strains (data not shown). The only difference is in the case of the class II strain SSOR, in which the presence of an endogenous plasmid suppresses the mutational activation of the *bgl* genes (Kharat & Mahadevan, 1999). Activation in this case requires the loss of the plasmid. Therefore, the differences described above are not confined to a single strain, but are common to *S. sonnei* strains isolated from diverse sources.

Analysis of transcription *in vivo* in *S. sonnei*

The studies described above suggest that the activation of the *bgl* genes in *S. sonnei* is likely to be at the level of transcriptional enhancement, similar to that documented in *E. coli*. The presence of the single base pair deletion within the Rho-independent terminator preceding the *bglG* gene also suggests that the basal level of transcription of *bglG* may be higher in *S. sonnei* than *E. coli*. To test these possibilities, we analysed transcription from the *bgl* promoter in *S. sonnei*.

To monitor transcription from promoters derived from the wild-type and activated *bgl* operon, plasmids carrying *bgl* promoter fragments (−250 to +340) from strains AK1 and AK101 were used to transform the wild-type strain. RNA isolated from the transformants was hybridized to two end-labelled oligonucleotide probes, one complementary to *bglG*, located downstream of the terminator, and the other complementary to *bla* mRNA. After digestion with S1 nuclease, the products were separated by denaturing PAGE. Following autoradiography, the *bgl* transcripts in each case were quantitated and normalized to the *bla* transcript. The results obtained showed that the level of transcription from the promoter carrying an insertion element is 16-fold higher than transcription from the wild-type promoter, indicating that the higher expression of the *bgl* genes in the mutant is brought about by transcriptional enhancement (data not shown).

To test whether the mutation at position +100 has any effect on the basal level of transcription of *bglG*, transcripts from plasmids carrying the promoter fragments from wild-type strains RV (*E. coli*) and AK1 (*S. sonnei*) were quantitated using the oligonucleotides described above. The results, presented in figure 2(c), show that the basal level of transcription downstream of the terminator is significantly higher in *S. sonnei* than *E. coli* (lane 3 compared to lane 2). This is likely to result in a higher basal level of the BglG protein in *S. sonnei* than in *E. coli*. Therefore, there are intrinsic differences in the expression pattern of the *bgl* genes in wild-type strains of *S. sonnei* and *E. coli*. To ensure that this difference in the basal level of transcription is not related to differences at the level of initiation in the two strains, transcription initiation from the *E. coli* and *S. sonnei* wild-type promoters was compared using oligonucleotides complementary to the first 42 nt in each case. These studies show that there is no appreciable difference in the level of transcription initiation from the two promoters (Fig. 2b). Interestingly, the wild-type promoters in the case of *E. coli* and *S. sonnei* show detectable transcription. However, this level of tran-

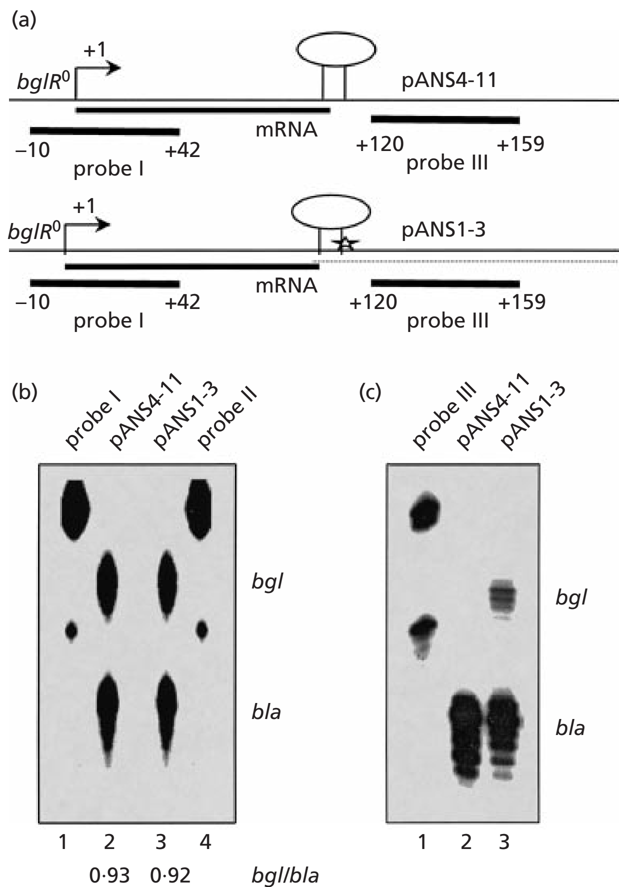

Fig. 2. Transcriptional analysis of the *S. sonnei bgl* operon by S1 nuclease protection. (a) Features of pANS4-11 and pANS1-3 carrying the wild-type promoters from *E. coli* and *S. sonnei*, respectively. The positions of the probes used in the transcriptional analysis are indicated as filled bars. (b) Transcription initiation from wild-type promoters of *E. coli* and *S. sonnei*. Total RNA isolated from the wild-type *E. coli* strain RV, carrying the two plasmids, was used to detect the levels of initiation by the S1 nuclease protection assay (see Methods). Lane 1, probe I; lane 2, transcription initiation from plasmid pANS4-11 (*E. coli* strain RV) detected using probe I; lane 3, transcription initiation from plasmid pANS1-3 (*S. sonnei* strain AK1) detected using probe II; lane 4, probe II. (c) Basal transcription of *bglG* from wild-type promoters of *E. coli* and *S. sonnei* detected using probe III. Lane 1, probe III; lane 2, transcription from pANS4-11 (*E. coli* strain RV); lane 3, transcription from pANS1-3 (*S. sonnei* AK1).

scription is insufficient to confer a Bgl$^+$ phenotype, which requires further enhancement by mutation. Because of efficient termination, no readthrough can be detected downstream in the case of *E. coli*.

To determine the effect of the mutation at position +100 on the level of *bglG* transcription in an activated strain, transcripts from plasmids carrying the promoter region from the Arb$^+$ Sal$^+$ strain AK102 were quantitated in the absence and presence of 7 mM salicin as an inducer, using the oligonucleotide probes described above. The results obtained show that the level of *bglG* transcription downstream of the terminator is similar in both cases, indicating constitutive expression of *bglG*

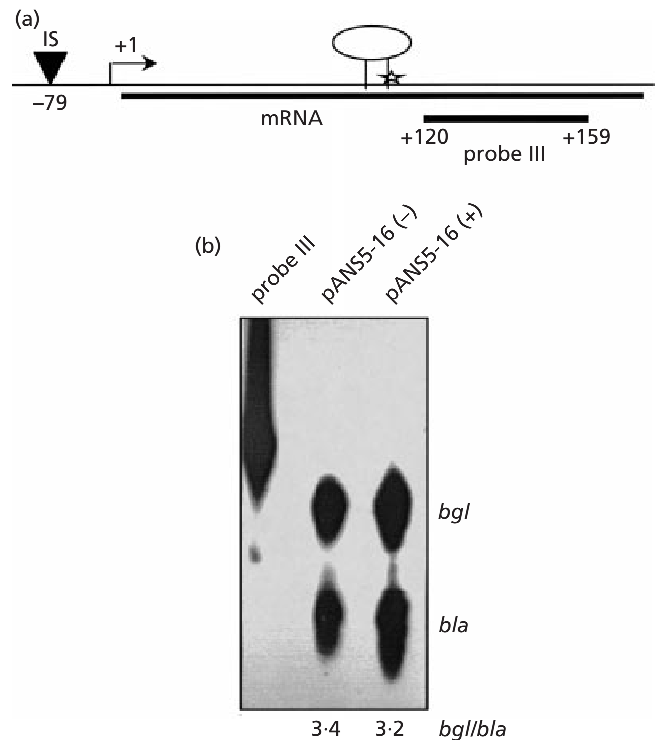

Fig. 3. Transcription from the activated promoter of *S. sonnei*. (a) Features of plasmid pANS5-16 carrying the activated promoter from *S. sonnei* strain AK102. The position of the activating IS element is indicated by a filled triangle. (b) Transcription of *bglG* under inducing and non-inducing conditions. Total RNA was isolated from transformants of the activated strain AK102 carrying the plasmid pANS5-16 (*S. sonnei* AK102), grown in the presence and absence of 7 mM salicin. The RNA was subjected to S1 nuclease analysis as above using probe III. Lane 1, probe III; lane 2, transcription from pANS5-16 in the absence of inducer; lane 3, transcription from the plasmid in the presence of inducer.

(Fig. 3). This is likely to have an impact on the regulation of the operon in *S. sonnei*.

Functional characterization of the *bgl* genes in *S. sonnei*

The two major differences with respect to the activity of the *bgl* genes in *S. sonnei* are the higher basal level of transcription of *bglG* and the two steps involved in activation to give a Sal$^+$ phenotype. To understand the functional significance of these observations, phenotypes of different *S. sonnei* strains were examined under defined genetic conditions.

To test the consequence of the higher basal level of *bglG* in wild-type *S. sonnei* strains, plasmid pMN5 (Mahadevan *et al*., 1987) carrying the *bglF* gene was introduced into the strain. Expression of *bglF* is directed from a vector promoter and is dependent on BglG, as the plasmid contains the Rho-independent terminator preceding the *bglF* gene. The presence of the plasmid in the strain AK1 conferred a weak Arb$^+$ phenotype to the

Table 2. Effect of plasmids carrying different *bgl* genes on the phenotypes of *Shigella* strains

Strain	Plasmid	Phenotype*	
		Arbutin	Salicin
S. sonnei AK1	pBR322	−	−
	pMN5 ($bglF^+$)	±	−
	p1H ($bglG'$ $bglF^+$ $bglB^+$)	ND	±
	pMN22AE ($bglR$::IS1 $bglG^+$ $bglF^+$ $bglB^+$)	+	+
E. coli RV	pBR322	−	−
	pMN5 ($bglF^+$)	−	−
	p1H ($bglG'$ $bglF^+$ $bglB^+$)	−	−
	pMN22AE ($bglR$::IS1 $bglG^+$ $bglF^+$ $bglB^+$)	+	+
S. sonnei AK101	pBR322	+	−
	p1H ($bglG'$ $bglF^+$ $bglB^+$)	+	+
	p6J ($bglB^+$)	+	+
	pMN22AE ($bglR$::IS1 $bglG^+$ $bglF^+$ $bglB^+$)	+	+

* Refers to phenotype on MacConkey arbutin or MacConkey salicin plates. − indicates white colonies, ± indicates pink colonies and + indicates red colonies.

strain confirming the higher basal level of BglG in the strain leading to enhanced expression of the BglF permease in the cell (Table 2). Similarly, when plasmid p1H (Mahadevan *et al*., 1987) carrying the *bglF* and *bglB* genes was introduced, the strain showed a weak Sal^+ phenotype. As in the case of pMN5, the *bglF* and *bglB* genes present on p1H require BglG for optimal expression. Under similar conditions, the wild-type *E. coli* strain RV remained Arb^- Sal^-, indicating that the levels of BglG are lower.

To investigate the reasons for the Sal^- phenotype of strain AK101, plasmid p6J (Mahadevan *et al*., 1987), carrying the *bglB* gene expressed from a vector promoter, was introduced into the strain. This resulted in a Sal^+ phenotype, suggesting that the *bglB* locus in AK101 carries a mutation (Table 2). The emergence of Sal^+ papillae is likely to be the result of reversion or suppression of this mutation subsequent to transcriptional activation of the promoter.

Analysis of the *bglB* locus in AK1, AK101 and AK102

The genetic analysis described above suggests that the *bglB* locus in AK101 carries a mutation and appearance of Sal^+ papillae is the result of reversion or suppression of this mutation. This was verified by carrying out Southern analysis of *S. sonnei* genomic DNA using the *E. coli bglB* gene as a probe. The results, presented in Fig. 4(b), show that the *bglB* locus in AK1 and AK101 carries an additional DNA sequence compared to *E. coli*. The size of this insertion is approximately 1·5 kb and is located at the 5′ end of the *bglB* locus, as evident from the Southern analysis using different restriction enzymes, sites of which are conserved in *S. sonnei*. Interestingly, the Sal^+ strain AK102 shows the same pattern as the wild-type and Arb^+ strains, indicating the retention of the insertion. The strain has apparently regained function of the *bglB* gene as a result of either intragenic or extragenic suppression of the insertion mutation.

Our efforts towards mapping the reversion as intragenic or extragenic, by P1 transduction or by cloning the PCR product carrying the intact insertion, along with flanking *bglB* sequences from the wild-type and mutant strains of *S. sonnei*, have been unsuccessful. This is likely to be related to the instability of the clones that carry the intact DNA element. However, cloning of parts of the insertion along with flanking *bglB* sequences has been possible. Southern analysis of *S. sonnei* genomic DNA using the insertion DNA as a probe has shown the presence of multiple copies of the sequence in the genome (A. S. Kharat & S. Mahadevan, unpublished), suggesting that the insertion is a transposable element.

To identify the point of insertion, PCR amplification of the insertion was followed by cycle sequencing of the 5′ end using primers designed based on the flanking *bglB* sequences. The point of insertion could be identified as position 3761 (coordinate based on data from Schnetz *et al*., 1987) in all *S. sonnei* strains examined. To test whether the element is similar to any of the known insertion sequences in the size range of 1·3–1·5 kb, the PCR product carrying the element was subjected to digestion using different restriction enzymes. These analyses suggest that the element is likely to be a novel one. Nucleotide sequence and additional characterization of the element will be published elsewhere.

Many insertion elements inactivating gene function are known to cause transcriptional polarity. To address this possibility in the case of transcription of the *bglB* locus in *S. sonnei*, *in vivo* transcriptional analysis of the *bglB* gene from the wild-type and salicin-utilizing strains of

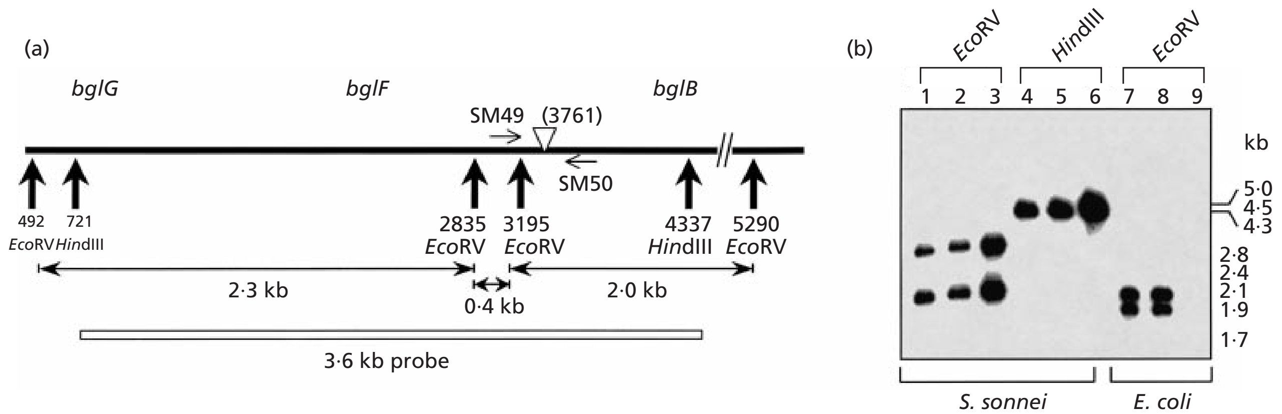

Fig. 4. Comparison of the *bglB* region of AK1, AK101 and AK102 by Southern analysis. (a) Restriction map of the *E. coli bglB* region. The positions of *Eco*RV and *Hin*dIII sites are indicated by vertical arrows. The inverted triangle indicates the position of the insertion element within *bglB*. The 3·6 kb probe used in the Southern analysis is shown as a white box. (b) Southern analysis of the *bglB* region of *S. sonnei* with the 3·6 kb probe. Lanes 1 and 4, *S. sonnei* strain AK1 (WT, Arb$^-$ Sal$^-$); lanes 2 and 5, *S. sonnei* strain AK101 (Arb$^+$ Sal$^-$); lanes 3 and 6, *S. sonnei* strain AK102 (Arb$^+$ Sal$^+$); lane 7, *E. coli* strain RV (WT); lane 8, *E. coli* strain AE328 (Bgl$^+$); lane 9, *E. coli* strain JF201 (Δ*bgl*).

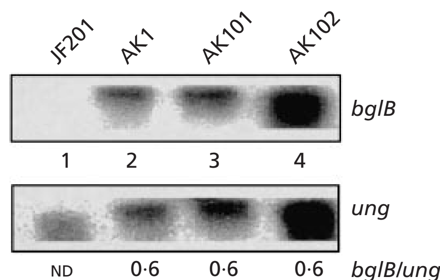

Fig. 5. Northern analysis of transcription of the *bglB* locus in *S. sonnei*. Total RNA was isolated from the strains indicated under inducing conditions and subjected to Northern analysis. The probes used for hybridization are indicated on the right side of each panel. Lane 1, JF201 (Δ*bgl* strain of *E. coli*); lane 2, AK1 (WT, Arb$^-$ Sal$^-$ strain of *S. sonnei*); lane 3, AK101 (Arb$^+$ Sal$^-$, derived from AK1); lane 4, AK102 (Arb$^+$ Sal$^+$, derived from AK101).

Table 3. Measurement of phospho-β-glucosidase B activity in different strains

Values shown are the mean of four experiments.

Strain	Phospho-β glucosidase B activity (units)	
	–	+
E. coli		
RV (Arb$^-$ Sal$^-$)	14 ± 4	12 ± 2
AE328 (Arb$^+$ Sal$^+$)	42 ± 4	253 ± 10
S. sonnei		
AK1 (Arb$^-$ Sal$^-$)	28 ± 4	42 ± 2
AK101 (Arb$^+$ Sal$^-$)	40 ± 4	80 ± 4
AK102 (Arb$^+$ Sal$^+$)	135 ± 7	342 ± 13

S. sonnei was carried out by Northern blot analysis. Blots of total RNA isolated under inducing conditions from strain JF201 (Δ*bgl*) of *E. coli* and strains AK1 (Arb$^-$ Sal$^-$), AK101 (Arb$^+$ Sal$^-$) and AK102 (Arb$^+$ Sal$^+$) of *S. sonnei* were probed with a *bglB*-specific probe. To normalize the levels of transcription, the blot was stripped and reprobed using a probe corresponding to the *E. coli ung* gene as an internal control (Varshney *et al*., 1988). The transcripts were normalized by taking the ratio of counts corresponding to the *bglB* signal to that of *ung*. These results are represented in Fig. 5. There was no appreciable difference in the levels or size of transcript between strains AK101 (Sal$^-$) and AK102 (Sal$^+$), indicating that the insertion within *bglB* is not polar and reversion to the Sal$^+$ state is not by relief of transcriptional polarity. However, these results do not rule out the possibility that the Sal$^+$ state of AK102 is derived by the activation of another cryptic β-glucosidase located elsewhere in the genome.

Measurement of the phospho-β-glucosidase B activity

To confirm the Sal$^+$ phenotype seen in the case of the strain AK102 at the level of *bglB* expression, measurements of phospho-β-glucosidase B activity were carried out in the strain. The results obtained are presented in Table 3. The fully induced activity seen in strain AK102 is comparable to that of *E. coli* strain AE328 that carries an activated copy of the *bgl* operon. Interestingly, the basal level of expression seen in AK102 is fourfold higher than that of AE328, a likely consequence of the mutation in the Rho-independent terminator in the

strain. These results indicate that strain AK102 can cleave salicin and the enzyme is expressed semi-constitutively.

DISCUSSION

The experiments described above were undertaken to address the differences in the pattern of activation of the *bgl* genes in *E. coli* and *S. sonnei*, two closely related organisms, with the objective of understanding the possible evolutionary reasons for the crypticity of the *bgl* genes. The results presented above highlight three major differences in the organization and function of the *bgl* genes in *S. sonnei* compared to *E. coli*. i) The *bglB* gene encoding the phospho-β-glucosidase B enzyme is insertionally inactivated in *S. sonnei*. ii) Despite the loss of function of the major structural gene involved in β-glucoside utilization, the basal level of transcription of the operon is enhanced. iii) Although the promoter structure and organization are conserved, the spectrum of activating mutations obtained under similar conditions is different.

The most striking difference in the expression of the *bgl* genes in *S. sonnei* is the higher basal level of transcription of *bglG*. Interestingly, the higher basal level of expression is not achieved by the activation of the cryptic promoter, but by a mutation in a downstream regulatory element. However, this increase is not sufficient to give an Arb^+ phenotype to the cell. Therefore it is unlikely that the mutation within the regulatory site arose as a direct result of selection for arbutin utilization. The presence of the mutation also makes expression of the operon constitutive once the promoter is activated, which is seen at the level of transcription and enzyme activity.

A majority of the activating mutations (87 %), isolated under nutritionally rich conditions, mapped within the *hns* locus in the case of *S. sonnei*. The complementation analysis using a multicopy plasmid carrying the wild-type copy of the *hns* gene does not rule out the possibility that the mutations map not within the *hns* gene, but at loci under the control of *hns*. Since H-NS is known to regulate *bgl* expression directly, this possibility is unlikely. Activating mutations isolated under identical conditions in *E. coli* map within the *bglR* locus and consist predominantly (98 %) of insertions. Interestingly, late-arising Arb^+ mutants of *E. coli*, isolated after prolonged incubation on minimal arbutin plates, showed a higher incidence (21 %) of lesions in the *hns* gene (Hall, 1998). The basis for this difference between the two strains is not known at this stage. One possibility is that *hns* mutations may occur at the same frequency in *E. coli*, but are not selected because of their weaker Bgl phenotype. Since the basal level of expression of the operon is higher in the *S. sonnei* strains as a result of the regulatory-site mutation, even weaker *hns* mutations may be able to confer a stronger Bgl phenotype in these strains. This is likely to have an impact on the expression of the *bgl* operon in the natural environments of the two organisms.

Activation of the operon by two steps, involving reversion of an insertion mutation in *bglF*, has been reported in the case of a specific strain of *E. coli* (Parker *et al.*, 1988). The difference between this report and our findings is that the insertional inactivation of *bglB* is not confined to a single strain, but is a property of all *S. sonnei* strains examined. Partial characterization of the insertion within the *bglB* locus has indicated that the point of insertion is identical in all strains examined, suggesting that the insertional inactivation of the *bglB* locus occurred early in the evolutionary history of *S. sonnei*.

It is interesting that the insertion element is retained in the Sal^+ revertants. Transcriptional analysis of the *bglB* locus reported here shows that *bglB* transcript is present even in the Sal^- strains. Therefore, the reversion is likely to be at a step involving the translation of the *bglB* message. Efforts to identify the nature of reversion have been hampered by our inability to clone the insertion element in its entirety. Genetic approaches such as P1 transduction, to map the location of the reversion, have also been unsuccessful. However, cloning of partial sequences of the element has been possible. A comparison of the nucleotide sequences of the *bglB* locus in the mutant and wild-type strains is likely to provide information regarding the nature of the reversion. Alternatively, the mutation in AK102 may lie in an unlinked locus and suppress the effect of the insertion. The results of the transcriptional analysis also do not exclude the possibility that the Sal^+ status of AK102 is the result of activation of another gene encoding a cryptic phospho-β-glucosidase specific for salicin.

One possible explanation for the structural gene inactivation is that *S. sonnei* represents a state where the *bgl* operon is in the process of becoming a pseudogene by accumulating mutations. This conclusion has two major limitations. Although the *bglB* gene has accumulated a mutation, the *bglG* and *bglF* genes are intact in the strain, conferring on it the ability to mutate to Arb^+. Secondly, if the operon is in the process of being converted to a pseudogene, it is difficult to explain the mutation in the regulatory site leading to an enhanced basal level of transcription. In the absence of the structural gene function, any selection for an enhanced basal level is unlikely. It is also unlikely that the mutation is selectively neutral since it is conserved in all strains of *Shigella* including *Shigella dysenteriae*, *Shigella flexneri* and *Shigella boydii* (A. S. Kharat & S. Mahadevan, unpublished).

This seemingly paradoxical situation can be addressed by considering the possibility that the regulators of the *bgl* operon, *bglG* and *bglF*, have additional roles to play in regulating cellular functions not directly related to β-glucoside utilization (Mahadevan, 1997). The *bglG*–*bglF* combine represents an efficient regulatory system in which the phosphotransferase component encoded by *bglF* acts as a sensor and the *bglG* gene product acts as a response regulator by mediating antitermination of transcription in response to the availability of β-glucosides. Similar systems that share considerable

functional and sequence similarities have been reported in different organisms. The situation in the case of *Bacillus subtilis* is particularly illuminative because of the presence of several paralogues of the BglG regulator (Stulke *et al*. 1998). It has been shown that the LicT regulator, a homologue of the BglG antiterminator, regulates at least two relatively unrelated functions in *B. subtilis* (Schnetz *et al*., 1996). Therefore it is conceivable that the *bgl* genes of *E. coli* perform regulatory functions elsewhere within the cell and are conserved for this reason. The basal level of the regulators may be sufficient for this putative function. Another possibility is that the mutation in *S. sonnei* that results in an increased basal level was selected under specific conditions that demanded higher levels of the *bgl* regulators. This is consistent with the results of a recent study demonstrating the *in vivo* expression of the wild-type *bgl* operon in a pathogenic strain of *E. coli* recovered from infected mouse liver (Khan & Isaacson, 1998). Such a possibility is also consistent with our results that the *bglR* and *bglG* loci are conserved in pathogenic strains of *Shigella* and *Salmonella* although these strains are incapable of mutating to Bgl^{+} (A. S. Kharat & S. Mahadevan, unpublished). However, until the putative targets for the regulators are identified, the possibility for alternative roles for the *bgl* genes, such as in pathogenicity, remains speculative.

In summary, our studies have helped to highlight the high degree of conservation of the 'cryptic' *bgl* genes in the two closely related micro-organisms, with interesting differences in terms of their organization and function. These results, along with a number of recent observations by different investigators (Keyhani & Roseman, 1997; Khan & Isaacson, 1998), prompt a re-examination of the concept of crypticity of genes in general.

ACKNOWLEDGEMENTS

We thank B. E. Uhlin, C. Ueguchi, J. Lopilato and U. Varshney for the gifts of *hns*, *leu*O, *bglJ* and *ung* constructs, respectively, and R. Roy and V. Ramchandran for the gift of clinical isolates of *S. sonnei*. We wish to thank members of the laboratory for fruitful discussions. A. S. K. acknowledges the help of N. R. Kulkarni in the isolation and identification of the strain AK1. This work was supported by a grant to S. M. from the Council for Scientific and Industrial Research, Government of India. A programme support to the Indian Institute of Science from the Department of Biotechnology is also acknowledged.

REFERENCES

Amster-Choder, O. & Wright, A. (1993). Transcriptional regulation of the *bgl* operon of *Escherichia coli* involves phosphotransferase system-mediated phosphorylation of a transcriptional antiterminator. *J Cell Biochem* **51**, 83–90.

Caramel, A. & Schentz, K. (1998). Lac and lambda repressors relieve silencing of the *Escherichia coli bgl* promoter: activation by alteration of a repressing nucleoprotein complex. *J Mol Biol* **284**, 875–883.

Defez, R. & DeFelice, M. (1981). Cryptic operon for β-glucoside metabolism in *Escherichia coli* K12: genetic evidence for a regulatory protein. *Genetics* **97**, 11–25.

DiNardo, S., Voelkel, K. A., Sternglanz, R., Reynolds, A. E. & Wright, A. (1982). *Escherichia coli* DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. *Cell* **31**, 43–51.

Falkow, S. & Mekalanos, J. (1990). The enteric bacilli and vibrios. In *Microbiology*, 4th edn, pp. 567–587. Edited by B. D. Davis, R. Dulbecco, H. N. Eisen & H. S. Ginsberg. Philadelphia: Lippincott.

Fox, F. C. & Wilson, G. (1967). The role of a phospho-*enol*-pyruvate-dependent kinase system in β-glucoside catabolism in *Escherichia coli*. *Proc Natl Acad Sci USA* **59**, 988–995.

Giel, M., Desnoyer, M. & Lopilato, J. (1996). A mutation in a new gene, *bglJ*, activates the *bgl* operon in *Escherichia coli* K12. *Genetics* **143**, 627–635.

Goransson, M., Sonden, B., Nilsson, P., Dagberg, B., Forsman, K., Emanuelsson, K. & Uhlin, B. E. (1990). Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*. *Nature* **344**, 682–685.

Hall, B. G. (1998). Activation of the *bgl* operon by adaptive mutation. *Mol Biol Evol* **15**, 1–5.

Hall, B. G. & Xu, L. (1992). Nucleotide sequence, function, activation and evolution of the cryptic *asc* operon of *Escherichia coli* K12. *Mol Biol Evol* **9**, 688–706.

Keyhani, N. O. & Roseman, S. (1997). Wild type *Escherichia coli* grows on the chitin disaccharide *N,N′*-diacetylchitobiose by expressing the *cel* operon. *Proc Natl Acad Sci USA* **94**, 14367–14371.

Khan, N. O. & Isaacson, R. E. (1998). *In vivo* expression of the β-glucoside (*bgl*) operon of *Escherichia coli* occurs in mouse liver. *J Bacteriol* **180**, 4746–4749.

Kharat, A. S. & Mahadevan, S. (1999). Plasmid-mediated suppression of the mutational activation of the *bgl* operon of *Shigella sonnei*. *Acta Biochim Polon* **46**, 853–861.

Kricker, M. & Hall, B. G. (1984). Directed evolution of cellobiose utilisation in *Escherichia coli* K-12. *Mol Biol Evol* **1**, 171–182.

Kricker, M. & Hall, B. G. (1987). Biochemical genetics of the cryptic gene system for cellobiose utilisation in *Escherichia coli* K12. *Genetics* **115**, 419–429.

Lawrence, J. G., Ochaman, H. & Hartl, D. L. (1992). The evolution of insertion sequences within enteric bacteria. *Genetics* **131**, 505–513.

Lopilato, J. & Wright, A. (1990). Mechanism of activation of the cryptic *bgl* operon of *Escherichia coli* K12. In *The Bacterial Chromosome*, pp. 435–444. Edited by C. Drlica & M. Riley. Washington, DC: American Society for Microbiology.

Mahadevan, S. (1997). The BglG group of antiterminators: a growing family of bacterial regulators. *J Biosci* **22**, 505–513.

Mahadevan, S. & Wright, A. (1987). A bacterial gene involved in transcription antitermination: regulation at a Rho-independent terminator in the *bgl* operon of *Escherichia coli*. *Cell* **50**, 485–494.

Mahadevan, S., Reynolds, A. E. & Wright, A. (1987). Positive and negative regulation of the *bgl* operon in *Escherichia coli*. *J Bacteriol* **169**, 2570–2578.

Mukerji, M. & Mahadevan, S. (1997). Characterisation of the negative elements involved in silencing the *bgl* operon of *Escherichia coli*: possible roles for DNA gyrase, H-NS, and CRP-cAMP in regulation. *Mol Microbiol* **24**, 617–627.

Owen, R. J. & Borman, P. (1987). A rapid biochemical method for purifying high molecular weight bacterial chromosomal DNA for restriction enzyme analysis. *Nucleic Acid Res* **15**, 3631.

Parker, L. L. & Hall, B. G. (1988). A fourth *Escherichia coli* gene system with the potential to evolve *β*-glucoside utilisation. *Genetics* **119**, 485–490.

Parker, L. L., Betts, P. W. & Hall, B. G. (1988). Activation of a cryptic gene by excision of DNA fragment. *J Bacteriol* **170**, 218–222.

Prasad, I. & Schaefler, S. (1974). Regulation of the *β*-glucoside system in *Escherichia coli* K12. *J Bacteriol* **120**, 638–650.

Prasad, I., Young, B. & Schaefler, S. (1973). Genetic determination of the constitutive biosynthesis of phospho-*β*-glucosidase A in *Escherichia coli* K12. *J Bacteriol* **114**, 909–915.

Reynolds, A. E., Felton, J. & Wright, A. (1981). Insertion of DNA activates the cryptic *bgl* operon in *Escherichia coli* K12. *Nature* **293**, 625–629.

Reynolds, A. E., Mahadevan, S., Le Grice, S. F. J. & Wright, A. (1986). Enhancement of bacterial gene expression by insertion elements or by mutation in a CAP-cAMP binding site. *J Mol Biol* **191**, 85–95.

Rutberg, B. (1997). Antitermination of transcription of catabolic operons. *Mol Microbiol* **23**, 413–421.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schaefler, S. & Malamy, A. (1969). Taxonomic investigations on expressed and cryptic phospho-*β*-glucosidases in *Enterobacteriaceae*. *J Bacteriol* **99**, 422–433.

Schnetz, K. (1995). Silencing of *Escherichia coli bgl* promoter by flanking sequence elements. *EMBO J* **14**, 2545–2550.

Schnetz, K. & Rak, B. (1992). IS5: a mobile enhancer of transcription in *Escherichia coli*. *Proc Natl Acad Sci USA* **89**, 1244–1248.

Schnetz, K., Toloczyki, C. & Rak, B. (1987). *β*-Glucoside (*bgl*) operon of *Escherichia coli* K12: nucleotide sequence, genetic organisation and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. *J Bacteriol* **169**, 2579–2590.

Schnetz, K., Stulke, J., Gertz, S., Kruger, S., Krieg, M., Hecker, M. & Rak, B. (1996). LicT, a *Bacillus subtilis* transcriptional antiterminator protein of the BglG family. *J Bacteriol* **178**, 1971–1978.

Singh, J., Mukerji, M. & Mahadevan, S. (1995). Transcriptional activation of the *Escherichia coli bgl* operon: negative regulation by DNA structural elements near the promoter. *Mol Microbiol* **17**, 1085–1092.

Stulke, J., Arnaud, M., Rapoport, G. & Martin-Verstraete, I. (1998). PRD – a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. *Mol Microbiol* **28**, 865–874.

Ueguchi, C., Ohta, T., Set, C., Suzuki, T. & Mizuno, T. (1998). The *leuO* gene product has a latent ability to relieve *bgl* silencing in *Escherichia coli*. *J Bacteriol* **180**, 190–193.

Varshney, U., Hutcheon, T. & Van de Sande, J. H. (1988). Sequence analysis, expression, and conservation of *Escherichia coli* uracil DNA glycosylase and its gene (*ung*). *J Biol Chem* **263**, 7776–7784.

Watanabe, H. & Okamura, N. (1991). The genus *Shigella*. In *The Prokaryotes*, 2nd edn, vol. III, *A handbook on the biology of bacteria: ecophysiology, isolation, identification and applications*, pp. 2754–2759. Edited by A. Balows & others. New York: Springer-Verlag.

Woodcock, D. M., Crowther, P. J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S. S., Michael, M. Z. & Graham, M. W. (1989). Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acid Res* **17**, 3469–3478.

Received 21 December 1999; revised 28 March 2000; accepted 11 April 2000.