ELSEVIER

FEMS Microbiology Letters 223 (2003) 267–274

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

The β-glucoside genes of *Klebsiella aerogenes*: conservation and divergence in relation to the cryptic *bgl* genes of *Escherichia coli*

Tirumalai R. Raghunand, S. Mahadevan *

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

Received 27 March 2003; received in revised form 23 April 2003; accepted 1 May 2003

First published online 5 June 2003

Abstract

The ability to metabolize aromatic β-glucosides such as salicin and arbutin varies among members of the Enterobacteriaceae. The ability of *Escherichia coli* to degrade salicin and arbutin appears to be cryptic, subject to activation of the *bgl* genes, whereas many members of the *Klebsiella* genus can metabolize these sugars. We have examined the genetic basis for β-glucoside utilization in *Klebsiella aerogenes*. The *Klebsiella* equivalents of *bglG*, *bglB* and *bglR* have been cloned using the genome sequence database of *Klebsiella pneumoniae*. Nucleotide sequencing shows that the *K. aerogenes bgl* genes show substantial similarities to the *E. coli* counterparts. The *K. aerogenes bgl* genes in multiple copies can also complement *E. coli* mutants deficient in *bglG* encoding the antiterminator and *bglB* encoding the phospho-β-glucosidase, suggesting that they are functional homologues. The regulatory region *bglR* of *K. aerogenes* shows a high degree of similarity of the sequences involved in BglG-mediated regulation. Interestingly, the regions corresponding to the negative elements present in the *E. coli* regulatory region show substantial divergence in *K. aerogenes*. The possible evolutionary implications of the results are discussed.

Keywords: Cryptic genes; β-glucosides; antiterminators; gene silencing

1. Introduction

Many bacterial genomes apparently carry genes that are potentially functional, but never expressed during the normal growth of the organism. The manifestation of the latent phenotype associated with these genes requires genetic processes such as mutation, transposition or recombination. In this respect, they differ from pseudogenes, since many of them can be activated by a single mutational event. Given this, the fact that such genes continue to be maintained in the genome without accumulating mutations is an evolutionary puzzle. The *bgl* operon of *Escherichia coli*, involved in the uptake and degradation of aryl-β-glucosides such as salicin and arbutin [1], is an example of such a cryptic genetic system.

The level of expression of the *bgl* genes in wild-type *E. coli* cells is insufficient to confer growth on β-glucoside substrates. Several classes of mutations have been observed to transcriptionally activate the operon leading to a Bgl^+ phenotype. The majority of activating mutations consist of insertions of IS*1* and IS*5* in a 223-bp target region; some insertions map downstream of the transcription start site [2,3]. In addition to insertions, several unlinked activating mutations have been identified. Mutations in *gyrA* and *gyrB* genes, leading to a reduction in negative supercoiling, have been shown to activate the operon [4]. Mutations that render the histone-like nucleoid structuring protein (H-NS) non-functional also lead to activation of the operon [5,6]. Two new loci, *bglJ* and *leuO*, have been described to be involved in activation of the operon as well [7,8].

The sequences upstream of the promoter have been shown to contain specific elements that have a negative effect on transcription. These include a sequence that can extrude into a cruciform structure under torsional stress [9], and a sequence that interacts with H-NS [9,10]. Activation of the operon is the result of disruption of the negative elements [9,11–14].

Once activated, the operon is subject to induction by β-glucosides. The first two genes of the operon, *bglG* and *bglF*, mediate regulation at this level [15]. BglG acts

* Corresponding author. Tel.: +91 (80) 293 2607; Fax: +91 (80) 360 0683.
E-mail address: mahi@mrdg.iisc.ernet.in (S. Mahadevan).

0378-1097/03/$22.00
doi:10.1016/S0378-1097(03)00393-8

as a positive regulator by mediating antitermination of transcription at two ρ-independent terminators that flank *bglG* [16,17]. It is an RNA binding protein, whose recognition sequence overlaps the terminators [18]. The activity of BglG is negatively regulated by the permease BglF at the level of phosphorylation. In vitro, in the absence of the inducer, BglF phosphorylates BglG [19,20]. The functional form of the protein is a dimer and phosphorylation leads to dissociation of the dimer, thus abrogating recognition of its RNA target [21]. BglF and BglG hence constitute an effective signaling system, the former acting as the membrane-bound sensor of β-glucosides, modulating the activity of the response regulator BglG, which in turn regulates expression of the *bgl* genes. The *bglB* gene encodes the enzyme phospho-β-glucosidase B that is specific to the phosphorylated forms of salicin and arbutin and cannot hydrolyze other β-glucosides such as cellobiose.

Apart from *Erwinia chrysanthemi* [22], the only other phenotypically Bgl$^+$ Gram-negative enteric for which the genes involved in β-glucoside utilization have been isolated is *Klebsiella oxytoca*. The *casRAB* genes from this organism, which encode the permease and phospho-β-glucosidase required for salicin, arbutin and cellobiose metabolism, are significantly similar to the *E. coli bgl* genes [23]. However, the operon sequences reported lack the N-terminus of the putative *bglG* homologue and the 5′ untranslated region, which leaves open the basis for the active state of this operon. Other members of the *Klebsiella* genus such as *K. pneumoniae* and *K. aerogenes* also exhibit a Bgl$^+$ phenotype. However, no information is available regarding the underlying genetic basis for the phenotype. If the genes responsible for the Bgl$^+$ status in *Klebsiella* are homologues of the *E. coli bgl* genes, their conservation in an active state in the organism will be of evolutionary significance.

In this study, we have investigated the genetic basis for the Bgl-positive phenotype of *K. aerogenes*. We find that the equivalents of the *E. coli bgl* genes are conserved in *K. aerogenes* to a high degree. However, there are interesting differences of the upstream regulatory elements that render the genes silent in *E. coli*. The evolutionary implications of these results in relation to the ecological niche occupied by the two organisms are discussed.

2. Materials and methods

2.1. Bacterial growth conditions and transformation

Bacterial strains (Table 1) were grown and maintained in Luria–Bertani liquid medium and agar respectively. For complementation analysis, cells were grown on MacConkey agar plates supplemented with 1% salicin, arbutin or cellobiose. Transformation of plasmid DNA was carried out by the polyethylene glycol/$MgSO_4$ method as described in [24]. For selection of transformants, ampicillin was used at a concentration of 200 μg ml^{-1}.

2.2. DNA manipulations

All DNA manipulations, including DNA sequencing, were carried out as described in [25].

2.3. Saligenin assay

Estimation of BglB (phospho-β-glucosidase B) activity, which measures specifically the cleavage of phospho-salicin, was carried out using a procedure similar to that described in [1]. Cells were grown to mid-exponential phase in M9 minimal medium with 0.4% succinate as the carbon source. 1 ml of cells were harvested, washed in 0.8% saline and resuspended in 0.1 ml saline. 0.1 ml of 4% salicin was added and the mixture incubated at 37°C for 30 min. The reaction was stopped by the addition of 0.5 ml of 2 M Na_2CO_3. Production of saligenin by cleavage of phospho-salicin was detected by the addition of 0.5

Table 1
List of bacterial strains and plasmids used in this study

Strain/plasmid	Vector	Description	Source/reference
Strains			
Klebsiella			
*Ka*AN292		*Klebsiella aerogenes* (sewage isolate) Ap^r, Kan^r	Kharat and Kulkarni (unpublished)
Kpn		*Klebsiella pneumoniae* (lab strain) Ap^r, Kan^r, Tet^r, Cam^r	This work
E. coli			
RV$^+$		Δ*lac*X74 *thi bglR11*(*bglR*::IS*1*) (Bgl$^+$)	A. Wright
AE328		Δ*lac*X74 *thi bglR11*(*bglR*::IS*1*) *tna*::Tn*10* (Bgl$^+$)	A. Wright
R302		*bglB*$^-$ derivative of AE328 (Bgl$^-$)	[43]
RPSM27		*bglG*$^-$ derivative of RV$^+$ *rpoS*::Tn*10* (Bgl$^-$)	[43]
Plasmids			
pMN25	pBR322	*E. coli bglR67 bglG*$^+$*bglF*′ Ap^r	[15]
pANS24-24	pUC18	*E. coli bglB*$^+$ Ap^r	[44]
p*KaG*	pUC18	*K. aerogenes* AN292 *bglG*$^+$ Ap^r	This work
p*KpnG*	pUC18	*K. pneumoniae bglG*$^+$ Ap^r	This work
p*KaB*	pUC18	*K. aerogenes* AN292 *bglB*$^+$ Ap^r	This work
p*KpnB*	pUC18	*K. pneumoniae bglB*$^+$ Ap^r	This work

ml of 0.6% 4-amino-antipyrine, followed by the addition of 0.25 ml of $K_3Fe(CN)_6$ after 15 min at room temperature. A positive reaction, indicated by the appearance of a red color, was quantitated by measuring absorbance at 509 nm. Absorbance at 600 nm was used to normalize values. Units of enzyme activity were determined using the following algorithm:

$$\text{Activity units} = \frac{1000 \times \text{OD}_{509}}{\text{OD}_{600}\ 10vt}$$

where v = volume of concentrated cells used in the assay and t = time of incubation.

2.4. Polymerase chain reaction (PCR) analysis

The *bglG* equivalent from *K. aerogenes* was amplified using *Taq* DNA polymerase with the primers SM57 (5′-GGTGA ATTCT ATTAG TCGTT CCGCC C-3′) and SM58 (5′-TGGGG ATCCT TAACG CCCCT CTTTT C-3′), which carry *Eco*RI and *Bam*HI sites respectively. The amplification conditions were as follows: initial denaturation at 94°C/10 min followed by 30 cycles of denaturation at 94°C/1 min, annealing at 50°C/2 min and extension at 72°C/2 min. A step of extension at 72°C/10 min was carried out at the end. The *bglB* equivalent from *K. aerogenes* was amplified using the primers SM60 (5′-GAGGA ATTCA TGAAA ACATT CCCGA C-3′) and SM61 (5′-TGCGG ATCCT TAGGC TTTCA GGCTG G-3′), which carry *Eco*RI and *Bam*HI sites respectively. The amplification conditions were identical to those described above. The *K. oxytoca casA* sequence was amplified using the primers SM51 (5′-GTGGT ACCGT AAGGA AAAAC AGCAT G-3′) and SM52 (5′-TAGAA TTCTG ATTAA CGGCT TACGG A-3′), with pLOI1906 [23] as the template. The PCR conditions were as described above with the exception that the annealing temperature was 54°C.

2.5. Construction of plasmids

The plasmids p*KaG* and p*KpnG* were constructed by PCR amplification with SM57 and SM58, the *bglG* equivalents from *K. aerogenes* and *K. pneumoniae* respectively, and cloning these at the *Eco*RI and *Bam*HI sites of pUC18. The plasmids p*KaB* and p*KpnB* were constructed by PCR amplification with SM61 and SM62, the *bglB* equivalent from *K. aerogenes* and *K. pneumoniae* respectively, and cloning these using the *Eco*RI and *Bam*HI sites of pUC18.

2.6. Sequence analysis

Pairwise sequence alignments were performed using the BLAST program at NCBI [26]. Multiple sequence alignments were generated using the ClustalW algorithm [27], followed by BOXSHADE analysis (http://www.ch.embnet.org/software/BOX_form.html). RNA secondary structure and free energy values were generated using the mFOLD algorithm [28]. Open reading frames (ORFs) in DNA sequences were detected using the MBS translator (http://pariswater.com/biomol/translator/index.htm).

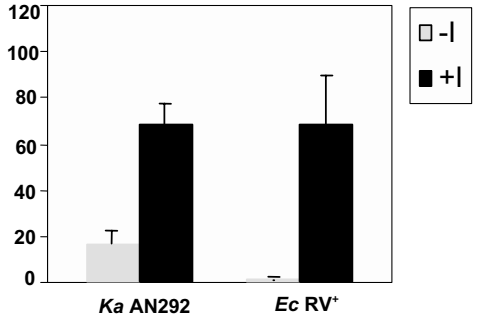


Fig. 1. Induction of β-glucosidase activity in response to salicin. Saligenin assays caried out with *K. aerogenes* AN292 (*Ka*AN292) and *E. coli* RV$^+$ (*Ec* RV$^+$) grown in minimal succinate medium in the presence (+I) or absence (−I) of 7 mM salicin. The *y*-axis represents enzyme activity in arbitrary units (see Section 2).

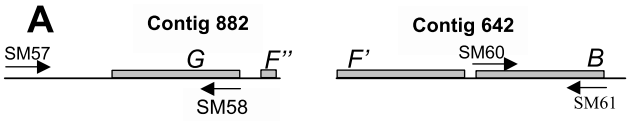
3. Results

3.1. Klebsiella *sp. show a higher basal level of phospho-β-glucosidase B activity compared to* E. coli

Members of the *Klebsiella* genus show a salicin-positive (Sal$^+$) phenotype. When the saligenin assay, which measures salicin-specific phospho-β-glucosidase B activity, was carried out, the basal level of activity in *K. aerogenes* was observed to be 18-fold higher than *E. coli* RV$^+$ (Fig. 1). Upon addition of 7 mM salicin, the induction observed was four-fold. The higher basal level of β-glucosidase activity is not surprising, considering the fact that members of the genus *Klebsiella* inhabit niches where β-glucosides are likely to be present at a low constant level. This would require the requisite enzymes to be present at levels sufficient to catabolize these sugars. The source of the enzyme activity was investigated by looking for homologues of the *bgl* genes of *E. coli*

3.2. K. pneumoniae *contains sequences resembling the* E. coli bgl *operon*

A computer-aided homology analysis was carried out using the recently completed *K. pneumoniae* genome sequence. The sequence available in the public domain is not annotated. It is organized in the form of contigs (http://genome.wustl.edu/gsc/Projects/bacterial/klebsiella/klebsiella.shtml) that are as yet unassembled. Following BLAST and TBLASTN analyses with the individual *E. coli bgl* gene sequences as queries, a putative homologue of this operon was identified spread in two separate contigs (Fig. 2A). Contig 882 was found to contain the



B

R302 (*bglB*⁻) transformed with	Sal Phenotype
pUC18	-
p*KaB*	++
pKpnB	++
pANS24-24 (*E. coli bglB*)	+++

C

RPSM27 (*bglG*⁻) transformed with	Sal Phenotype
pUC18	-
p*Ka*G	+
p*Kpn*G	+
pMN25 (*E. coli bglG*)	+

Fig. 2. Cloning and characterization of the *E. coli bglB* equivalents from *Klebsiella* sp. A: Schematic representation of primers used to amplify the putative *bglB* and *bglG* homologues from *Klebsiella* and their respective contig organizations in the *K. pneumoniae* genome sequence. B: Complementation analysis of *bglB* clones from *K. aerogenes* (p*KaB*) and *K. pneumoniae* (p*KpnB*). The intensity of the red color of colonies on MacConkey salicin plates is indicated by + symbols. C: Complementation analysis of *bglG* clones from *K. aerogenes* (p*Ka*G) and *K. pneumoniae* (p*Kpn*G). The intensity of the red color of colonies on MacConkey salicin plates is indicated by + symbols.

full-length *bglG* equivalent as well as putative upstream regulatory sequences and contig 642 contained the complete *bglB* equivalent. The *bglF* equivalent is split between the two contigs, with a small segment of its N-terminus on the first (*F″*) and the remainder on the second (*F′*). The three genes therefore seem to be organized as an operon as in *E. coli*. The percentage sequence identity observed is in the same range as those reported for the *casRAB* genes from *K. oxytoca* [23].

3.3. *The* bglB-*like genes from* K. pneumoniae *and* K. aerogenes *in multiple copies can complement an* E. coli bglB *mutant*

To clone and characterize the *bgl* equivalent identified in the *K. pneumoniae* database, a PCR-based strategy was employed. Primers (SM60 and SM61), designed to amplify the entire *bglB* (Fig. 2A) gene, were used in PCR reactions with genomic DNA isolated from *K. pneumoniae* and *K. aerogenes*. A 1.4-kb fragment amplified from both species was cloned into pUC18. To determine if these were functional *bglB* equivalents, the clones were transformed into R302, an *E. coli* strain carrying a *bglB* mutation. The ability of the clones to complement the mutation was assessed by the appearance of red colonies on MacConkey agar plates supplemented with salicin. Clones from both *Klebsiella* species could complement the *E. coli* mutant (Fig. 2B), suggesting that these were functional *bglB* homologues. Some caution has to be exercised in this interpretation as the complementation observed is with multiple copies of the *bglB* locus.

The sequence of the entire coding region of the clone carrying the *K. aerogenes bglB* (*Ka* BglB) equivalent was determined (GenBank accession number AY124800). The 1395-bp sequence contained a 464-aa ORF which showed 74% identity (344/464) and 85% similarity (396/464) to *E. coli* BglB. The sequence was then subjected to BLASTP analysis to identify other similar sequences in the database. Apart from showing a high degree of similarity with *E. coli* BglB, significant matches were also obtained with *K. oxytoca* CasB, and *E. chrysanthemi* ArbB. Multiple sequence alignment of the four proteins (data not shown) showed that *Ka* BglB contains the conserved LFIVENGLG motif, indicating that it belongs to the glycosylhydrolase family 1 of enzymes.

3.4. K. aerogenes *BglG can functionally replace* E. coli *BglG*

To functionally characterize the *bglG*-like sequences identified from database analysis, primers (SM57 and SM58) flanking the entire *bglG* sequence including putative regulatory sequences (Fig. 2A) were used to carry out PCR amplification using genomic DNA from *K. aerogenes* and *K. pneumoniae* as templates. The ~1.1-kb fragments amplified from *K. aerogenes* and *K. pneumoniae* were cloned into pUC18. To establish the authenticity of these clones, complementation analysis was carried out with RPSM27, an *E. coli* strain carrying a *bglG* mutation. The ability of the clones to rescue the mutation was assessed by the appearance of red colonies on MacConkey agar plates containing salicin. Clones from both species were able to rescue the Sal$^-$ phenotype of the mutant (Fig. 2C), suggesting that these are functional homologues of *E. coli bglG*.

3.5. K. aerogenes *BglG is a member of the BglG/SacY family of antiterminators*

The sequence of the 1.1-kb insert from the complementing *K. aerogenes* clone was determined (GenBank accession Number AY124799) and was found to be 99% identical to its cognate *K. pneumoniae* sequence identified by database analysis (data not shown). An 834-bp ORF encoding a 277-aa protein was identified, and this showed

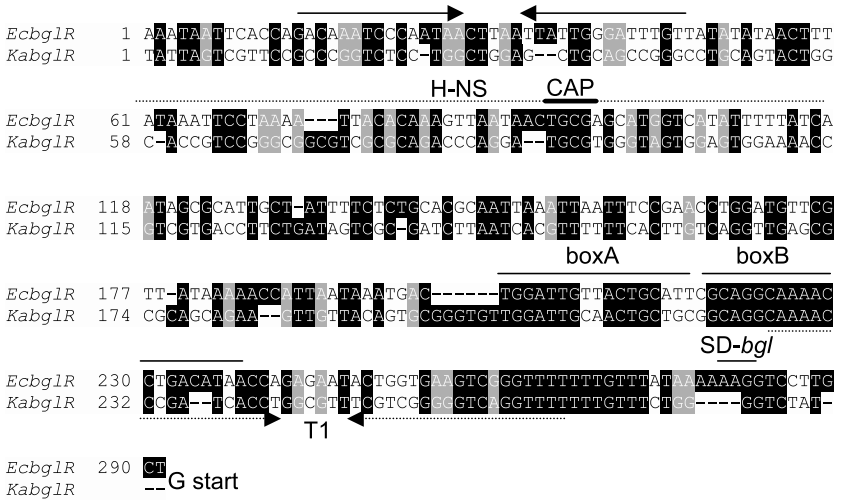


Fig. 3. Sequence alignment of *bgl* regulatory regions from *K. aerogenes* (*Ka bglR*) and *E. coli* (*Ec bglR*). The 284-nt sequence upstream of *Ka bglG* was aligned with the *E. coli bgl* regulatory region. The head to head arrows represent the cruciform. The CAP binding site, RAT sequences (box A, box B), the terminator 1 (T1, broken arrows) and putative Shine–Dalgarno sequence for *E. coli* BglG are all indicated on the sequence. The putative H-NS binding region is indicated with a dotted line.

65% identity (180/275) and 82% similarity (226/275) to *E. coli* BglG. BLAST analysis of this protein indicated significant matches to *E. chrysanthemi* ArbG, *K. oxytoca* CasR, and *Bacillus subtilis* SacY, apart from *E. coli* BglG, all belonging to the BglG/SacY family of antiterminators. Multiple sequence alignments with all these proteins showed the presence of two PTS regulation domain (PRD) motifs [29] in *K. aerogenes* BglG, spanning residues 93–159 and 200–269 (data not shown). This domain, present in several operon-specific transcriptional regulators including antiterminators and activators, is a target for phosphorylation by the PTS. PRD I contains a conserved DH box (residues 100–101), shown to be important for the activity of *E. coli* BglG and *B. subtilis* SacT [30,31]. PRD II contains, among other conserved residues, the His residue corresponding to H208 in *E. coli*, which has been predicted to be the site for phosphorylation by BglF [32]. In addition, sequences associated with β-sheet formation that are believed to be involved in RNA binding/dimerization [33] are partially conserved. All these features suggest that the *K. aerogenes* BglG can be categorized in the BglG/SacY family of antiterminators.

3.6. *The* bgl *regulatory region in* K. aerogenes *lacks the negative elements involved in silencing the* E. coli bgl *promoter*

The *E. coli bgl* promoter is silent because of the presence of negative regulatory elements that interfere with the transcription of the *bgl* genes. To examine the status of these elements in *K. aerogenes*, the 284-bp sequence present upstream of *K. aerogenes bglG* (*Ka bglR*) was aligned with *Ec bglR* (292–575 of the *E. coli bgl* sequence [34]. The sequence alignment (Fig. 3) indicates that the overall similarity between these sequences is low (42%), although significant regions of local homology can be identified. Interestingly, sequence similarity in the regions that comprise the negative elements in *E. coli bglR* (the inverted repeat and the H-NS binding region) is poor, indicating that these elements are compromised in *Ka bglR*. Moreover, the region corresponding to the inverted repeat in *Ka bglR* is GC-rich, which would make cruciform extrusion energetically unfavorable.

E. coli H-NS is known to bind AT-rich tracts, regions that also have the tendency to assume a bent conformation, features essential for H-NS interaction [35,36]. This region in *Ka bglR* is GC-rich as well, which would conceivably reduce or completely exclude binding of H-NS. The absence of these negative elements could be directly related to the active state of the *bgl* genes in *K. aerogenes* compared to *E. coli*. In addition to the negative elements, *Ec bglR* is also known to harbor a binding site for the catabolite activator protein (CAP), a well-characterized transcription activator in bacteria. The *Ec bglR* CAP site shows 15 identities to the proposed 22-base consensus CAP binding site [37], and CAP has been demonstrated to be a positive regulator of *bgl* transcription [2]. In contrast the *Ka bglR* CAP site shows only a 5/22 nucleotide identity to the consensus, suggesting that the role for CAP in *bgl* transcription in *K. aerogenes* is likely to be negligible. This is consistent with the observation that glucose has only a marginal effect in inhibiting the expression of β-glucosidase in *Klebsiella* (data not shown).

3.7. *The* K. aerogenes bgl *regulatory region shows conserved RAT and terminator sequences*

Although the regions corresponding to the negative elements in *Ec bglR* are poorly conserved, *Ka bglR* shows a remarkable conservation of the RNA antiterminator se-

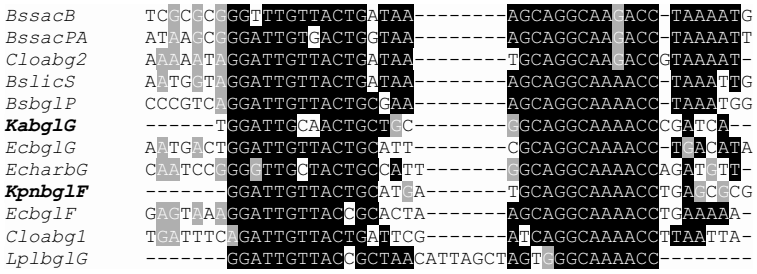


Fig. 4. Multiple sequence alignment of RAT sequences. RAT sequences present upstream of the first and/or second genes of catabolic operons namely *B. subtilis sacB* (*BssacB*), *B. subtilis sacPA* (*BssacPA*), *C. longisporum abg* (*Cloabg1,2*), *B. subtilis licS* (*BsulicS*), *B. subtilis bglPH* (*BsbglP*), *K. aerogenes bglG* (*KabglG*), *E. coli bglG* (*EcbglG*), *E. coli bglF* (*EcbglF*), *E. chrysanthemi arbG* (*EcharbG*), *K. pneumoniae bglF* (*KpnbglF*) and *L. plantarum bglG* (*LplbglG*) were aligned using the ClustalW algorithm.

quence (RAT), the sequence involved in binding BglG [18]. In addition, the RAT overlaps a putative stem loop structure that is also similar to the corresponding sequence in *Ec bglR*, a ρ-independent terminator (T1). Sequence alignment of the *bglG-bglF* inter-cistronic region of *K. pneumoniae* with that in *E. coli* identified a similar region of similarity, constituting the second terminator (data not shown). The free energy of formation of the two identified terminators in *Klebsiella* (KaT1: −19.6 kcal mol^{-1}, KpnT2: −24.8 kcal mol^{-1}) was determined by mFold analysis [28], and found to be comparable to the cognate *E. coli* terminators (EcT1: −19.7 kcal mol^{-1}, EcT2: −26.6 kcal mol^{-1}). To ascertain the degree of conservation in the *Klebsiella* RAT sequences with RAT sequences recognized by the BglG/SacY family of antiterminators, multiple sequence alignment was carried out (Fig. 4). The two query sequences were found to be strikingly similar to the various RAT sequences tested, indicating that *Ka* BglG, in addition to recognizing its cognate RAT sequence, could in principle recognize heterologous RAT sequences as well. The ability of *Ka* BglG to complement an *E. coli bglG* mutant is therefore a consequence of the significant similarity in the recognition sequences of the two antiterminators.

4. Discussion

Several classes of bacteria have been shown to possess β-glucoside utilization systems that are similar in organization and function to the *bgl* operon of *E. coli*. Although the primary function of these genes is to enable the metabolism of aromatic β-glucosides, some organisms have tailored subtle alterations of the genes to suit their physiological requirements. This is exemplified by the ability of the *bvr* locus in *Listeria monocytogenes*, which shares considerable similarities with the *bglG/bglF* genes of *E. coli*, to control the switching of genes involved in pathogenesis [38]. In most cases the regulation of these genes is mediated by the BglG/SacY class of proteins, which antiterminate transcription at ρ-independent terminators, present upstream of the *bgl* structural genes. The activity of these proteins is in turn regulated by the PEP-dependent phosphotransferases via a phosphorylation event coupled to sugar transport, allowing induction of these genes in the presence of β-glucosides. Systems like the *bglPH* operon of *B. subtilis*, which lack an operon-specific antiterminator, recruit paralogues such as *sacY/licT* to bring about induction [39]. Cleavage of the β-glycosidic bond is brought about by phospho-β-glucosidases, which, in certain cases, possess a wide substrate range to include aliphatic sugars such as cellobiose as well, *K. oxytoca* CasB being a prime example. Therefore, although the central theme is maintained, several variations from the *E. coli* paradigm exist in the microbial world.

The *K. aerogenes bglB* equivalent was cloned and shown to be functional by its ability to complement R302, an *E. coli bglB* mutant. Though the complementation was seen with multiple copies of the gene, sequence analysis of this clone, followed by alignment with proteins bearing similarity to *K. aerogenes* BglB, suggested that it belonged to family 1 of the glycosylhydrolases. These proteins also share a high level of identity with phospho-β-galactosidases, β-galactosidases and β-glucosidases originating from Gram-positive as well as Gram-negative bacteria, thermophiles, rats and humans [40]. Although structural information on these proteins is scarce, a glutamate residue present in a conserved signature sequence LFIVENGLG in members of this family (also conserved in *K. aerogenes* BglB) has been shown to be involved in hydrolysis of the β-glucosides in *Agrobacterium* sp. [41]. *K. oxytoca* CasA, which shows significant sequence similarity to *K. aerogenes* BglB, is a phospho-cellobiase and can cleave cellobiose in addition to salicin and arbutin. However, *K. aerogenes* BglB as well as *K. pneumoniae* BglB, expressed in multiple copies, were unable to hydrolyze cellobiose in a co-transformation assay, indicating that unlike their *K. oxytoca* homologue, their substrate range is restricted to the aryl-β-glucosides (data not shown). This observation suggests that *K. aerogenes* has two independent systems for the utilization of the two classes of β-glucosides, with little or no overlapping substrate specificity. However, the possibility that these enzymes possess a very low affinity (or catalytic activity) for cellobiose cannot be ruled out at this stage, though this may not be physiologically significant.

The putative *bglG* homologue of *K. aerogenes* is able to complement RPSM27, a *bglG* mutant of *E. coli* in multiple copies. In contrast, the *E. chrysanthemi arbG* gene was unable to complement an *E. coli* mutant [22], suggesting that even multiple copies of a gene exhibiting a high degree of similarity may not necessarily lead to functional replacement because of changes in specific residues. This indicates that *K. aerogenes* BglG is able to mediate antitermination at the *E. coli bgl* terminators, implying that it is able to recognize *E. coli* RAT sequences. A number of conserved elements were identified on sequence alignment of the predicted 277-aa *Ka* BglG ORF with some of the members of the BglG/SacY family. Significant of these were two PRD domains, known to be involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. These domains encompassed the conserved DH box required for activity, as well as the conserved histidine residue believed to be the site of phosphorylation by the PTS-dependent permeases. Both of these observations strongly suggest that the activity of *K. aerogenes* BglG is under regulation by *Ka* BglF, the PTS-dependent permease, as seen in *E. coli*.

On comparing the regulatory sequences in *K. aerogenes* and *E. coli*, the negative elements implicated in shutting off transcription of the *E. coli bgl* genes were found to be absent, as inferred from the low degree of sequence similarity observed in these regions. In *E. coli bglR*, a sequence of dyad symmetry, with the potential to extrude into a cruciform at enhanced superhelical stress, negatively regulates transcription [9]. In the *K. aerogenes bglR*, this region is GC-rich, which would preclude formation of such a structure, since extrusion in such a context would require significantly more energy than in an AT-rich stretch. A second element that is absent is the sequence that is predicted to bind the nucleoid structuring protein H-NS. This protein is known to have a preference for binding to AT-rich and curved DNA sequences both of which are characteristics of the region bound in *Ec bglR* [42]. This region is also GC-rich in *K. aerogenes bglR*, besides having a lower degree of curvature as determined by theoretical DNA bend analysis (data not shown). In principle therefore, H-NS is unlikely to bind such a sequence of DNA, though this cannot be stated categorically.

The lack of both of these negative elements is consistent with the observation that the *K. aerogenes bgl* genes, unlike their *E. coli* counterparts, are not silent. The active state of the genes in *Klebsiella* may be related to the fact that the primary habitat of *Klebsiella* is soil, where β-glucosides are likely to be encountered more frequently. This may also be related to the high level of basal activity of phospho-β-glucosidase activity detected in *K. aerogenes*. In contrast, *E. coli* resides predominantly in the mammalian intestine where the availability of aryl β-glucosides is uncertain. From the results reported here, it cannot be definitively concluded whether the *bgl* genes of *K. aerogenes* and *E. coli* are homologues. Considering the extensive degree of similarity between the genes and the genetic organization of the two systems, this is a likely possibility. Given this, it is interesting that the major divergence between the two has occurred within the regulatory sequences in response to the ecological niche occupied by the organisms.

Acknowledgements

We thank Dr. Ingram for the gift of the plasmid carrying the *K. oxytoca cas* genes. This work was supported by a grant from the Council for Scientific and Industrial Research (CSIR), Government of India.

References

[1] Prasad, I. and Schaefler, S. (1974) Regulation of the β-glucoside system in *Escherichia coli* K-12. J. Bacteriol. 120, 638–650.
[2] Reynolds, A.E., Mahadevan, S., Le Grice, S.F.J. and 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.
[3] Schnetz, K. and Rak, B. (1992) IS5: a mobile enhancer of transcription in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 89, 1244–1248.
[4] Di Nardo, S., Voelkel, K.A., Sternglanz, R., Reynolds, A.E. and Wright, A. (1982) *Escherichia coli* DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. Cell 31, 43–51.
[5] Defez, R. and De Felice, M. (1981) Cryptic operon for β-glucoside metabolism in *Escherichia coli* K12: genetic evidence for a regulatory protein. Genetics 97, 11–25.
[6] Higgins, C.F., Dorman, C.J., Stirling, D.A., Waddel, L., Broth, I.R., May, G. and Bremer, E. (1988) Physiological role of DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. Cell 52, 569–584.
[7] Giel, M., Desnoyer, M. and Lopilato, J. (1996) A mutation in a new gene, *bglJ*, activates the *bgl* operon in *Escherichia coli* K-12. Genetics 143, 627–635.
[8] Ueguchi, C., Ohta, T., Seto, C., Suzuki, T. and Mizuno, T. (1998) The *leuO* gene product has a latent ability to relieve *bgl* silencing in *Escherichia coli*. J. Bacteriol. 180, 190–193.
[9] Mukerji, M. and Mahadevan, S. (1997) Characterization 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.
[10] Schnetz, K. and Wang, J.C. (1996) Silencing of the *Escherichia coli bgl* promoter: effects of template supercoiling and cell extracts on promoter activity in vitro. Nucleic Acids Res. 24, 2422–2428.
[11] Lopilato, J. and Wright, A. (1990) Mechanisms of activation of the cryptic *bgl* operon of *E. coli* K-12. In: The Bacterial Chromosome (DrLica, K. and Riley, M., Eds.), pp. 435–444. ASM press, Washington, DC.
[12] Schnetz, K. (1995) Silencing of *Escherichia coli bgl* promoter by flanking sequence elements. EMBO J. 14, 2545–2550.
[13] Singh, J., Mukerji, M. and 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.
[14] Caramel, A. and Schnetz, 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.
[15] Mahadevan, S., Reynolds, A.E. and Wright, A. (1987) Positive and negative regulation of the *bgl* operon in *Escherichia coli*. J. Bacteriol. 169, 2570–2578.
[16] Mahadevan, S. and Wright, A. (1987) A bacterial gene involved in transcription antitermination: regulation at a rho-independent terminator in the *bgl* operon of *E. coli*. Cell 50, 485–494.
[17] Schnetz, K. and Rak, B. (1988) Regulation of the *bgl* operon of *Escherichia coli* by transcriptional antitermination. EMBO J. 7, 3271–3277.
[18] Houman, F., Diaz-Torres, M.R. and Wright, A. (1990) Transcriptional antitermination in the *bgl operon* of *E. coli* is modulated by a specific RNA binding protein. Cell 62, 1153–1163.
[19] Amster-Choder, O., Houman, F. and Wright, A. (1989) Protein phosphorylation regulates transcription of the β-glucoside utilisation operon in *E. coli*. Cell 58, 847–855.
[20] Schnetz, K. and Rak, B. (1990) β-Glucoside permease represses the *bgl* operon of *Escherichia coli* by phosphorylation of the antiterminator protein and also interacts with glucose-specific enzyme III, the key element in catabolite control. Proc. Natl. Acad. Sci. USA 87, 5074–5078.
[21] Amster-Choder, O. and Wright, A. (1992) Modulation of the dimerization of a transcriptional antiterminator protein by phosphorylation. Science 257, 1395–1398.
[22] el Hassouni, M., Henrissat, B., Chippaux, M. and Barras, F. (1992) Nucleotide sequences of the *arb* genes, which control β-glucoside utilization in *Erwinia chrysanthemi*: comparison with the *Escherichia coli bgl* operon and evidence for a new β-glycohydrolase family including enzymes from Eubacteria, Archeabacteria, and Humans. J. Bacteriol. 174, 765–777.
[23] Lai, X., Davis, F.C., Hespell, R.B. and Ingram, L.O. (1997) Cloning of cellobiose phosphoenolpyruvate-dependent phosphotransferase genes: functional expression in recombinant *Escherichia coli* and identification of a putative binding region for disaccharides. Appl. Environ. Microbiol. 63, 355–363.
[24] Nishimura, A., Morita, M., Nishimura, Y. and Sugino, Y. (1990) A rapid and highly efficient method for preparation of competent *Escherichia coli* cells. Nucleic Acids Res. 18, 6169.
[25] Sambrook, J., Fritsch, E.F. and Manaitis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[26] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
[27] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
[28] Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 9, 133–148.
[29] Stulke, J., Arnaud, M., Rapoport, G. and 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.
[30] Crutz, A.M., Steinmetz, M., Aymerich, S., Richter, R. and Le Coq, D. (1990) Induction of levansucrase in *Bacillus subtilis*: an antitermination mechanism negatively controlled by the phosphotransferase system. J. Bacteriol. 172, 1043–1050.
[31] Debarbouille, M., Arnaud, M., Fouet, A., Klier, A. and Rapoport, G. (1990) The *sacT* gene regulating the *sacPA* operon in *Bacillus subtilis* shares strong homology with transcriptional antiterminators. J. Bacteriol. 172, 3966–3973.
[32] Chen, Q., Engelberg-Kulka, H. and Amster-Choder, O. (1997) The localization of the phosphorylation site of BglG, the response regulator of the *Escherichia coli bgl* sensory system. J. Biol. Chem. 272, 17263–17268.
[33] van Tilbeurgh, H., Manival, X., Aymerich, S., Lhoste, J.M., Dumas, C. and Kochoyan, M. (1997) Crystal structure of a new RNA-binding domain from the antiterminator protein SacY of *Bacillus subtilis*. EMBO J. 16, 5030–5036.
[34] Schnetz, K., Toloczyki, C. and Rak, B. (1987) β-Glucoside (*bgl*) operon of *Escherichia coli* K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. J. Bacteriol. 169, 2579–2590.
[35] Owen-Hughes, T.A., Pavitt, G.D., Santos, D.S., Sidebotham, J.M., Hulton, C.S., Hinton, J.C. and Higgins, C.F. (1992) The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. Cell 71, 255–265.
[36] Zuber, F., Kotlarz, D., Rimsky, S. and Buc, H. (1994) Modulated expression of promoters containing upstream curved DNA sequences by the *Escherichia coli* nucleoid protein H-NS. Mol. Microbiol. 12, 231–240.
[37] Gunasekera, A., Ebright, Y.W. and Ebright, R.H. (1992) DNA sequence determinants for binding of the *Escherichia coli* catabolite gene activator protein. J. Biol. Chem. 267, 14713–14720.
[38] Brehm, K., Ripio, M.T., Kreft, J. and Vazquez-Boland, J.A. (1999) The *bvr* locus of *Listeria monocytogenes* mediates virulence gene repression by β-glucosides. J. Bacteriol. 181, 5024–5032.
[39] Kruger, S. and Hecker, M. (1995) Regulation of the putative *bglPH* operon for aryl-β-glucoside utilization in *Bacillus subtilis*. J. Bacteriol. 177, 5590–5597.
[40] el Hassouni, M., Chippaux, M. and Barras, F. (1990) Analysis of the *Erwinia chrysanthemi arb* genes, which mediate metabolism of aromatic β-glucosides. J. Bacteriol. 172, 6261–6267.
[41] Withers, S.G., Warren, A.J.R., Street, I.P., Rupitz, K., Kempton, J.B. and Aebersold, R. (1990) Un-equivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a 'retaining' glycosidase. J. Am. Chem. Soc. 112, 5887–5889.
[42] Timchenko, T., Bailone, A. and Devoret, R. (1996) Btcd, a mouse protein that binds to curved DNA, can substitute in *Escherichia coli* for H-NS, a bacterial nucleoid protein. EMBO J. 15, 3986–3992.
[43] Yakkundi, A., Moorthy, S. and Mahadevan, S. (1998) Reversion of an *E. coli* strain carrying an IS*1*-activated *bgl* operon under non selective conditions is predominantly due to deletions within the structural genes. J. Genet. 77, 21–26.
[44] Kharat, A.S. (1999) Molecular and Functional Characterisation of the β-Glucoside Utilisation (*bgl*) Genes in the *Shigella* Group of Organisms. PhD Thesis, Indian Institute of Science, Bangalore.