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.

© 2003 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

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 IS1 and IS5 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...
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-terminal of the putative bglG homologue and the S' 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/MgSO₄ 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 Na₂CO₃. Production of saligenin by cleavage of phospho-salicin was detected by the addition of 0.5 ml of 2 M Na₂CO₃. Production of saligenin by cleavage of phospho-salicin was detected by the addition of 0.5 ml of 2 M Na₂CO₃.
ml of 0.6% 4-amino-antipyrine, followed by the addition of 0.25 ml of \( K_3\text{Fe(CN)}_6 \) 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:

\[
\text{Activity units} = \frac{1000 \times \text{OD}_{509}}{\text{OD}_{600} \times 10^t}
\]

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 (\( 5'-\text{GTTGA ATCTT ATTAG TCGTT CCGCC C-3'} \)) and SM58 (\( 5'-\text{TGGGG ATCCT TAACG CCCCT CTTTT C-3'} \)), which carry \( EcoRI \) and \( BamHI \) 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 (\( 5'-\text{TTCTG ATTAA CGGCT TACGG A-3'} \)) and SM61 (\( 5'-\text{TGGGG ATCCT TAACG CCCCT CTTTT C-3'} \)), which carry \( EcoRI \) and \( BamHI \) sites respectively. The amplification conditions were identical to those described above. The \( K. oxytoca \) \( casA \) sequence was amplified using the primers SM51 (\( 5'-\text{GAGGA ATTCA TGAAA ACATT CCCGA C-3'} \)) and SM52 (\( 5'-\text{TGGGG ATCCT TAACG CCCCT CTTTT C-3'} \)), 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 \( pKaG \) and \( pKpnG \) were constructed by PCR amplification with SM57 and SM58, the \( bglG \) equivalents from \( K. aerogenes \) and \( K. pneumoniae \) respectively, and cloning these at the \( EcoRI \) and \( BamHI \) sites of pUC18. The plasmids \( pKaB \) and \( pKpnB \) were constructed by PCR amplification with SM61 and SM62, the \( bglB \) equivalent from \( K. aerogenes \) and \( K. pneumoniae \) respectively, and cloning these using the \( EcoRI \) and \( BamHI \) 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).

### 3. Results

#### 3.1. Klebsiella sp. show a higher basal level of phospho-\( \beta \)-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-\( \beta \)-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 \( \beta \)-glucosidase activity is not surprising, considering the fact that members of the genus \( Klebsiella \) inhabit niches where \( \beta \)-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...
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 (\(F^\prime\)). 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 \(bglG\) equivalents in \(K. oxytoca\) and \(K. pneumoniae\). 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 LFI-VENGLG motif, indicating that it belongs to the glycosylhydrolyase 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 \(\sim\)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 \(Klebsiella\) species could complement the \(E. coli\) mutant (Fig. 2B), suggesting that these were functional 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 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

3.5. \(K. aerogenes\) \(BglG\) is a member of the \(BglGI/SacY\) 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
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 L-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-
proteins is in turn regulated by the PEP-dependent phospho-
transferases 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 for induction [39]. Cleavage of the β-glucosidic 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 glycosylhydrodolases. 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 is present in a conserved signature sequence LFI-VENGLG 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 anti-termination 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 peremase. Both of these observations strongly suggest that the activity of _K. aerogenes_ BglG is under regulation by _Ka_ BglF, the PTS-dependent peremase, 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**


