A Fourth *Escherichia coli* Gene System With the Potential to Evolve β-Glucoside Utilization

Laura L. Parker and Barry G. Hall

Department of Molecular and Cell Biology, University of Connecticut U-44, Storrs, Connecticut 06268

Manuscript received December 7, 1987

Revised copy accepted April 1, 1988

ABSTRACT

*Escherichia coli* K12 is being used to study the potential for adaptive evolution that is present in the genome of a single organism. Wild-type *E. coli* K12 do not utilize any of the β-glucoside sugars arbutin, salicin or cellobiose. It has been shown that mutations at three cryptic loci allow utilization of these sugars. Mutations in the *bgl* operon allow inducible growth on arbutin and salicin while *cel* mutations allow constitutive utilization of cellobiose as well as arbutin and salicin. Mutations in a third cryptic locus, *arbT*, allow the transport of arbutin. A salicin⁺ arbutin⁺ cellobiose⁺ mutant has been isolated from a strain which is deleted for both the *bgl* and *cel* operons. Because the mutant utilized salicin and cellobiose as well as arbutin, it is unlikely that it is the result of a mutation in *arbT*. A second step mutant exhibited enhanced growth on salicin and a third step mutant showed better growth on cellobiose. A fourfold level of induction in response to arbutin and a twofold level of induction in response to salicin was observed when these mutants were assayed on the artificial substrate *p*-nitrophenyl-β-D-glucoside. Although growth on cellobiose minimal medium can be detected after prolonged periods of time, these strains are severely inhibited by cellobiose in liquid medium. This system has been cloned and does not hybridize to either *bgl* or *cel* specific probes. We have designated this gene system the *sac* locus. The *sac* locus is a fourth set of genes with the potential for evolving to provide β-glucoside utilization.

LEWONTIN (1974) asked “How much genetic variation is there that can be the basis of adaptive evolution? To answer that question it is not sufficient to measure genetic variation, which we can now do, nor to measure the present variation in fitness associated with that genetic variation, which we have not done but which may be possible with a proper reorientation of theory. We require, further, that we be able to assess the potential for adaptive evolution in genetic variation that may currently be nonadaptive.”

*Escherichia coli* K12 affords us the opportunity to study the potential for adaptive evolution that is present in the genome of a single organism. *E. coli* K12 does not utilize β-glucoside sugars as growth nutrients. By selecting mutants that are capable of growth on β-glucosides as sole carbon and energy sources we can explore the range and diversity of genes that have the potential for evolving to provide a specific new function that permits adaptation to a novel environment in which β-glucosides are the sole carbon source.

*E. coli* K12 possesses a variety of genes for β-glucoside utilization, but these genes are not expressed in the wild-type organism. Mutations at several different loci have been described which activate these genes and thereby allow *E. coli* to utilize the β-glucoside sugars arbutin, salicin and cellobiose. Mutations at the *bgl* locus allow utilization of arbutin and salicin (PRASAD and SCHAEFLER 1974) via the inducible expression of a transport gene *bglF*, a hydrolase gene, *bglB*, and a positive activator of the system, *bglG* (MAHADERVAN, REYNOLDS and WRIGHT 1987). The *cel* operon is expressed constitutively in active strains and includes a gene for β-glucoside transport, *celT*, and a hydrolase gene, *celH*, both of which act on all three sugars (KRICKER and HALL 1987). The product of a third silent locus, *arbT*, allows transport of arbutin (KRICKER and HALL 1987). *bglA* is the only gene for β-glucoside utilization that is expressed constitutively in wild-type *E. coli*. This gene maps outside the *bgl* operon and allows hydrolysis of phosphorylated arbutin (PRASAD, YOUNG and SCHAEFLER 1974). Mutants of *arbT* are able to grow on arbutin because *bglA* specifies the necessary hydrolase.

A recent survey of natural isolates of *E. coli* obtained from farm animals (HALL and FAUNCE 1987) and from African Yellow baboons (ROUTMAN et al. 1985) has shown that while the majority of *E. coli* do not utilize any of the β-glucoside sugars, a few rare β-glucoside positive isolates do exist. Most of the positive strains examined did not express RNA transcripts homologous to either the *bgl* or *cel* operon (HALL and FAUNCE 1987). Our laboratory has also screened the *E. coli* Reference Collection (ECOR) for β-glucoside-positive strains (HALL and BETTS 1987). Although none of the 72 natural isolates was able to utilize any

1 To whom correspondence should be addressed.

β-glucoside sugar, the majority of these strains were able to mutate spontaneously to utilize at least one of these sugars. RNA transcripts were analyzed from a subset of these mutants and it was found that several did not express either the cel or bgl operons. Because many of the mutant strains utilized both cellobiose and salicin it is unlikely that these isolates were simply mutants of arfbT or bglA. We concluded that at least four sets of genes for β-glucoside utilization exist in the population of natural isolates of *E. coli*. We estimated that, on average, a natural isolate will possess two sets of β-glucoside utilization genes in a silent, but potentially active, state and that two sets will have been irreversibly inactivated by the accumulation of random mutations (HALL and BETTS 1987). In other words, our survey of variation in natural populations suggested that the population possesses four loci with the potential for permitting adaptation to growth on β-glucosides.

We now ask whether a specific strain, *E. coli* K12, possesses four loci with this same potential. In other words, how much of the evolutionary potential that is estimated to be extant in the population is possessed by a single member of that population?

**MATERIALS AND METHODS**

**Culture media and conditions:** All cultures were grown at 30° with aeration. Minimal media contained phosphate buffered mineral salts solution (HALL and BETTS 1987) with 1.5% agar as a solidifying agent. Sugars were added at a concentration of 0.1% (w/v) except for glucose which was used at 0.2% (w/v). When required, amino acids were added at a concentration of 100 μg/ml, and ampicillin was used at a concentration of 100 μg/ml. MacConkey indicator plates contained 1% (w/v) of the appropriate sugar and were prepared according to the manufacturer (Difco). If the added sugar was utilized, colonies were pink or red. If it was not used, colonies were white.

**Molecular techniques:** Methods for the preparation of genomic DNA, plasmid DNA, isolation and labeling of probe DNA, and DNA/DNA hybridizations were previously described (HALL and BETTS 1987). The cel specific probe was prepared from a 4.0-kb *HpaI* fragment of plasmid pUF654 (HALL, BETTS and KRICKER 1986) and the bgl specific probe was prepared from a 3.5-kb *HindIII* fragment of plasmid pAR6 (REYNOLDS, FELTON and WRIGHT 1981). Preparation of competent cells and transformations were as previously described (HALL, BETTS and KRICKER 1986).

**In vivo cloning:** The mini-Mu derivative Mud5005 was used as described in GROisman and CASADIAN (1986). The helper phage employed was Mu cts02 pAP5 which carries an ampicillin resistance determinant (AKROYD et al. 1984).

**Other cloning vectors:** The plasmid pBlu+ and pBlu− are 3.0-kb high copy number vectors that carry multiple cloning site cartridges and are the product of Stratagene, Inc., of San Diego, California.

**Growth rates:** Cultures were grown overnight, with constant aeration, in minimal medium containing a limiting (0.1 g/liter) concentration of glucose. Exhaustion of the glucose limits growth to a density of about 10⁸ cells ml⁻¹. Salicin, arbutin, cellobiose or glucose was added to the cultures, and the increase in turbidity was monitored with a Gilford Stasar II spectrophotometer at 600 nm. The growth rates were calculated from the slope of the least squares fit of ln A₆₀₀ vs. time in hours.

**Geneic nomenclature for the bgl operon:** Until recently bglC was used to designate the gene specifying the β-glucoside transport system, and bglS the gene specifying a positive regulatory locus (PRASAD and SCHAEFFER 1974). Two recent papers (MAHDEVAN et al. 1987; SCHNETZ, TOLONCZKI and RAK 1987) reversed those definitions. In order to avoid further confusion, BARBARA J. BACHMANN, curator of the *E. coli* Genetic Stock Center, after consultation with the CGSC Advisory Committee, has renamed these loci (B. BACHMANN, personal communication). In this paper we use the new designations as they will appear in the next edition of the linkage map of *E. coli*. *bgp* now designates the β-glucoside-specific enzyme II of the phosphotransferase system, i.e., the transport protein. *bgG* now designates the gene encoding the positive regulator of expression of the bgl operon.

**E. coli strains:** The strains used in this study are listed in Table 1.

**β-Glucosidase assays:** β-Glucosidase activity was assayed by measuring the hydrolysis of p-nitrophenyl-β-D-glucoside (PNPG) in intact cells. Cultures were grown to late log phase at 30°, and the A₆₀₀ was determined in a Gilford spectrophotometer. The cells were harvested by centrifugation, washed once with mineral salts buffer, and resuspended in mineral salts buffer. One-tenth volume of 50 mM PNPG in phosphate buffer was added to the culture and the assays were incubated at 30° until a yellow color was visible. Reactions were terminated by the addition of an equal volume of 1 M Na₂CO₃. The cells were pelleted by centrifugation and the A₄₁₀ of the supernatant determined in a Gilford spectrophotometer. Four assays were performed for each set of conditions. The activity of the enzyme was determined using a molar extinction coefficient for p-nitrophenol of 16,900 A.U. in a 1-cm lightpath at 410 nm. One unit is the hydrolysis of 1 nanomole of PNPG per minute. Specific activities are reported as units/A₆₀₀. An A₆₀₀ of 1.0 is 10⁹ cells.

**RESULTS**

**Construction of strain LP100:** To determine if *E. coli* K12 possessed a fourth set of genes with the potential for β-glucoside utilization we deleted both the bgl and cel operons.

Plasmid pUF650 contains the cel genes from strain MK2 cloned into the low copy number, temperature sensitive, vector pHSG415 (HASHIMOTO-GOTOH et al. 1981). pUF650 was cut with *HpaI* and self-ligated in order to delete the majority of the cel operon. The resulting plasmid (pUF651) was transformed into the cel−Δbgl strain MK91 and UV irradiated to stimulate recombination. The culture was grown overnight at 42° to select for loss of the plasmid, and plated onto MacConkey cellobiose medium. Several white colonies were selected and screened for the presence of the deletion by DNA/DNA hybridization using the 4.0-kb *HpaI* fragment of the cel operon as a probe. One strain, which failed to hybridize to the probe, was selected and designated LP100. The cel allele present in strain LP100 is designated celΔ (100).
Its parent by forming slightly larger colonies on cellobiose minimal medium. That strain was distinguishable from isolation of another strain which outgrew its siblings, strain LP103. That strain was designated strain LP100. A colony was observed that was significantly larger than strain LP102; it was restreaked onto minimal cellobiose medium. It exhibited a stronger salicin and cellobiose positive result of a separate mutational event. Twenty papillae were streaked onto MacConkey cellobiose medium, and after several days of incubation two strains were isolated even though selection for these phenotypes was never imposed. This suggests that the same system is responsible for utilization of all three sugars.

To rule out the possibility that these mutants were actually celR+ or bglR+ contaminants the DNA from each of these strains was screened for the retention of the deletions by dot blot DNA/DNA hybridization. The results showed that all four strains, LP100, LP101, LP102 and LP103, carry both the bgI and the cel deletions.

We cannot formally rule out the possibility that these strains are actually mutants of arbT and bgIA that allow transport and hydrolysis of salicin and cellobiose in addition to arbutin. That possibility is unlikely, however, because mutations in two different genes would be required. A mutation would be needed to change arbT from its cryptic to its active state, another mutation would be necessary to allow arbT to transport salicin and cellobiose in addition to arbutin, and a third mutation would be necessary to allow bgIA to hydrolyze salicin and cellobiose. While it is not impossible that these three events could occur in a single step, it is highly improbable.

We conclude that a fourth system for β-glucoside utilization exists in E. coli K12 and we designate this system the sac (salicin-arbutin-cellobiose) locus.

### Selection of β-glucoside utilizing mutants: Strain LP100 was plated onto MacConkey cellobiose medium at a density of approximately 50 cells/plate. Colonies were white, indicative of failure to ferment the cellobiose. The plates were sealed and incubated at 30° for approximately 1 month. At the end of this time, papillae (outgrowths) were evident on the surface of the colonies. A single papilla was selected from each colony to ensure that each mutant was the result of a separate mutational event. Twenty papillae were streaked onto MacConkey cellobiose medium, and after several days of incubation two strains exhibited a very light pink color. One of these strains was designated LP101. Strain LP101 also exhibited a strong arbutin- and a weak salicin-positive phenotype on MacConkey media. To select a better cellobiose-utilizing mutant, strain LP101 was streaked onto minimal cellobiose medium, and the plate was sealed and incubated at 30°. After a period of 2 weeks a colony was observed that was significantly larger than its siblings. That colony was designated strain LP102. It exhibited a stronger salicin and cellobiose positive phenotype than did its parent, strain LP101. Strain LP102 was restreaked onto minimal cellobiose medium for a third round of selection. This lead to the isolation of another strain which outgrew its siblings, strain LP103. That strain was distinguishable from its parent by forming slightly larger colonies on cellobiose minimal medium.

It should be noted that each round of selection was performed on cellobiose medium, but after three rounds of selection only a weak cellobiose positive mutant was isolated. Strong arbutin-, salicin-positive strains were isolated even though selection for these phenotypes was never imposed. This suggests that the same system is responsible for utilization of all three sugars.

### Growth rates permitted by expression of the sac locus: Growth rates of strains LP100, LP101, LP102 and LP103 were measured on glucose, arbutin and salicin at 30°. Assays were performed in quadruplicate and the mean and 95% confidence intervals were determined. All four strains grew with the same first order growth rate constant (≈0.6 hr⁻¹) on glucose (Table 2). This suggests that mutations which allow growth on β-glucoside sugars do not impair the central pathway for glucose metabolism. Strain LP101 had a first order growth rate constant of 0.31 hr⁻¹ on arbutin, a value that did not improve significantly upon further selection. This same strain showed slow growth on salicin (0.07 hr⁻¹). Strain LP102 did not show any improvement when grown on arbutin, but the first order growth rate constant doubled to 0.14 hr⁻¹ on salicin (Table 2). Strain LP102 is not distinguishable from strain LP103 on either arbutin or salicin. The distinction between these two strains lies in the ability of strain LP103 to form colonies faster than strain LP102 on solid minimal cellobiose medium.

### Cells expressing the sac genes are inhibited during growth on cellobiose: Strains LP102 and LP103 were isolated on cellobiose minimal medium. Growth was extremely slow when colony size was measured over time. After 5 days of incubation on cellobiose

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP100</td>
<td>F⁻ rpsL trp his argG metA or B ara leu lacY lacZ Δ(A4680) bgI-pho Δ(201) celA(100) sac2²</td>
<td>This study</td>
</tr>
<tr>
<td>LP100R</td>
<td>F⁻ rpsL trp his argG metA or B ara leu lacY lacZ Δ(A4680) bgI-pho Δ(201) celA(100) sac2² recA</td>
<td>This study</td>
</tr>
<tr>
<td>LP101</td>
<td>F⁻ rpsL trp his argG metA or B ara leu lacY lacZ Δ(A4680) bgI-pho Δ(201) celA(100) sac1⁴ Arb⁺ Sal⁺ Cello⁺</td>
<td>This study</td>
</tr>
<tr>
<td>LP102</td>
<td>F⁻ rpsL trp his argG metA or B ara leu lacY lacZ Δ(A4680) bgI-pho Δ(201) celA(100) sac2² Arb⁺ Sal⁺ Cello⁺</td>
<td>This study</td>
</tr>
<tr>
<td>LP103</td>
<td>F⁻ rpsL trp his argG metA or B ara leu lacY lacZ Δ(A4680) bgI-pho Δ(201) celA(100) sac3⁴ Arb⁺ Sal⁺ Cello⁺</td>
<td>This study</td>
</tr>
<tr>
<td>MK91</td>
<td>F⁻ rpsL trp his argG metA or B ara leu lacY lacZ Δ(A4680) bgI-pho Δ(201) celA(100) sac5⁵ Arb⁺ Sal⁺ Cello⁺</td>
<td>Kricke and Hall (1984)</td>
</tr>
</tbody>
</table>

*β-Glucoside Genes in E. coli*
TABLE 2

Growth rates

<table>
<thead>
<tr>
<th>Strain</th>
<th>Glucose</th>
<th>Arbutin</th>
<th>Salicin</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP100</td>
<td>0.600 ± 0.011</td>
<td>0.006 ± 0.002</td>
<td>0.003 ± 0.001</td>
</tr>
<tr>
<td>LP101</td>
<td>0.613 ± 0.014</td>
<td>0.314 ± 0.007</td>
<td>0.072 ± 0.036</td>
</tr>
<tr>
<td>LP102</td>
<td>0.597 ± 0.019</td>
<td>0.340 ± 0.032</td>
<td>0.142 ± 0.003</td>
</tr>
<tr>
<td>LP103</td>
<td>0.613 ± 0.014</td>
<td>0.354 ± 0.021</td>
<td>0.162 ± 0.026</td>
</tr>
</tbody>
</table>

Growth rates are given as the first order rate constant in hr⁻¹ ± the 95% confidence limits.

minimal medium, strain LP101 colonies reached a diameter of 1.0 mm, strain LP102 1.5 mm and strain LP103 2.0 mm. The parent strain, LP100, formed 0.45 mm microcolonies in 5 days. It is clear that strains LP101, LP102 and LP103 are cellobiose positive, even though only weakly so. These growth rates can be contrasted with those on arbutin, where all three strains had colony sizes of 5.0 mm by day 4.

We performed growth rate studies on strains LP100, LP101, LP102 and LP103 in cellobiose liquid medium. The growth curves are shown in Figure 1. Each mutant exhibited an initial growth spurt, then abruptly curtailed growth after a period of 90 min. When monitored over a long period of time, occasional growth spurts were observed. The sporadic nature of these data do not allow us to calculate a first order growth rate constant on cellobiose for any of these strains. This suggests that cells expressing the sac genes are severely inhibited when growing on cellobiose. We suggest that colonies are able to form on plates only because they were incubated for long periods of time. Brief periods of growth over the course of many days may be sufficient for the formation of colonies.

The sac locus is regulated: β-Glucosidase activity specified by the sac genes was measured during differing growth conditions (Table 3) by measuring the hydrolysis of PNPG. It is evident that a basal level of enzyme was synthesized when the cultures were grown on glucose. When grown on arbutin, that level of expression increased fourfold. This suggests that sac is regulated and responds to arbutin. The mean specific activity of strain LP101 grown on glucose and on salicin was the same (3.5), suggesting that strain LP101 is not induced by salicin. Strain LP102 exhibited a mean specific activity of 5.2 when grown on glucose and 9.3 when grown on salicin, suggesting that the additional mutation in strain LP102 allows the sac system to respond to salicin even though there is only twofold induction. No regulatory differences were seen between strains LP102 and LP103 when assayed for PNPG hydrolysis. Cellobiose acts as an anti-inducer in all cases.

It should be noted that strain LP100 exhibited a low level of synthesis of a β-glucosidase even when grown on glucose. This level of expression does not support growth but suggests that sac is not completely silent in its wild-type state. It has been shown that the bgl operon is also expressed at a low level in its cryptic state (Prasad and Schaefer 1974).

Cloning the sac locus: The sac locus was cloned in vivo using the Mini-mu system described by Groisman and Casadaban (1986). Strain LP103 was lysogenized by Mud5005 and Mu cts62 pAp5. Heat induction of the resulting dilysogen produced a lysate that included Mini-mu phage particles carrying random fragments of the LP103 genome. Mud5005 also carries a kanamycin-resistance gene. The lysate was used to infect strain LP100. The infected culture was plated onto salicin minimal medium containing kanamycin. The resulting colonies were expected to carry

FIGURE 1.—Growth on cellobiose minimal medium. Note that the ordinate is linear, not logarithmic.

TABLE 3

Induction of β-glucosidase activity

<table>
<thead>
<tr>
<th>Strain</th>
<th>Glucose*</th>
<th>Cellobiose</th>
<th>Arbutin</th>
<th>Salicin</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP100</td>
<td>0.009 ± 0.008*</td>
<td>0.081 ± 0.005</td>
<td>0.03 ± 0.005</td>
<td>0.04 ± 0.008</td>
</tr>
<tr>
<td>LP101</td>
<td>3.40 ± 0.26</td>
<td>1.50 ± 0.21</td>
<td>12.70 ± 1.91</td>
<td>3.50 ± 0.820</td>
</tr>
<tr>
<td>LP102</td>
<td>5.20 ± 0.690</td>
<td>3.60 ± 0.560</td>
<td>15.90 ± 3.50</td>
<td>9.30 ± 2.00</td>
</tr>
<tr>
<td>LP103</td>
<td>3.10 ± 0.220</td>
<td>1.80 ± 0.160</td>
<td>13.50 ± 0.840</td>
<td>6.70 ± 1.90</td>
</tr>
</tbody>
</table>

* Carbon and energy source during growth.
* Units of β-glucosidase activity/A₄₅₀ = 95% confidence limits.
β-Glucoside Genes in *E. coli*  489

DNA from plasmid pUF721 does not hybridize to probes derived from either the *cel* or the *bgl* operons, confirming that *sac* is indeed a separate and distinct system for the utilization of β-glucoside sugars.

**DISCUSSION**

We have identified a fourth locus for β-glucoside utilization in *E. coli* K12 and designated it the *sac* locus. The *sac* system must specify at least a transport system and a hydrolyase that act on salicin and arbutin and, to a lesser extent, on cellobiose. The *sac* system is inducible by arbutin and a secondary mutation permits some induction by salicin. Induction appears to be inhibited by cellobiose. We have cloned the *sac* genes and shown that when the genes are present at a high copy number cellobiose can be utilized effectively.

Studies on natural isolates of *E. coli* suggested to us that *E. coli* K12 may possess additional systems for the utilization of β-glucoside sugars. It is not known whether the system we have described here is one of the systems that we detected in natural isolates of *E. coli*. It is not known whether natural isolates even possess *arbT* or *bglA*, and it has been found that a large portion of Cel+ natural isolates are actually deleted for the *bgl* operon (Hall 1988). Future studies of β-glucoside utilization by natural isolates of *E. coli* will include analysis of mRNA to determine if *sac* is being expressed in these strains.

Cryptic genes have been defined as genes that are not normally expressed during the lifetime of an organism, but which can be activated by genetic events (Hall, Yokoyama and Calhoun 1983). Should *sac* be considered a cryptic gene system? It is clearly expressed, and its expression is even induced by cellobiose, in the parent strain LP100 (Table 3). If one applies a strict definition of cryptic genes then *sac* is not cryptic. The *cel* operon, in contrast is not expressed at a level that can be detected either by activity measurements or by RNA hybridization (Krick and Hall 1987; Hall and Betts 1987) and can therefore be strictly considered a cryptic locus. Whether or not *sac* should be considered cryptic, the important issue is that the wild-type allele is not expressed at a level that is biologically meaningful in that it permits effective use of β-glucosides, but that it has the potential of mutating to provide that function.

One aspect that distinguishes *sac* from the *cel* and *bgl* loci is that only a single mutation is required to fully active these other loci, whereas three sequential mutations are required before the *sac* locus permits effective use of arbutin, salicin and cellobiose.

The finding that *E. coli* K12 possesses four sets of genes with the potential for evolving β-glucoside utilization is in sharp contrast to results obtained with
the EBG system. If the lac operon of E. coli K12 is deleted the organism has the potential to re-evolve lactose utilization via mutations in the ebg operon (Hall 1982). When the ebg operon was also deleted it proved impossible to obtain lactose utilizing mutants either spontaneously or following chemical mutagenesis (B. G. Hall, unpublished results), suggesting that the potential for evolving lactose utilization is limited to one gene.

Why does E. coli harbor so many seemingly superfluous genes for β-glucoside utilization, when a single operon should be sufficient to meet the needs of a cell? The probability of activating a gene by a random mutation increases with the number of genes present. It may be advantageous for E. coli to have a variety of genes with the potential for β-glucoside utilization because it may increase the probability that a cell would acquire a mutation in a gene which would allow it to be favored in a particular environmental situation.

We are grateful to P. W. Betts for his expert technical assistance and advice. This study was supported by National Institutes of Health grant GM 37110.

LITERATURE CITED


