Directed Evolution of Cellobiose Utilization in *Escherichia coli* K12

Maja Kricker and Barry G. Hall

University of Connecticut

The cellobiose catabolic system of *Escherichia coli* K12 is being used to study the role of cryptic genes in evolution of new functions. *Escherichia coli* does not use β-glucoside sugars; however, mutations in several loci can activate the cryptic *bgl* operon and permit growth on the β-glucoside sugars arbutin and salicin. Such Bgl⁺ mutants do not use cellobiose, which is the most common β-glucoside in nature. We have isolated a Cel⁺ (cellobiose-utilizing) mutant from a Bgl⁺ mutant of *E. coli* K12. The Cel⁻ mutant grows well on cellobiose, arbutin, and salicin. Genes for utilization of these β-glucosides are located at 37.8 min on the *E. coli* map. The genes of the *bgl* operon are not involved in cellobiose utilization. Introduction of a deletion covering *bgl* does not affect the ability to utilize cellobiose, arbutin, or salicin, indicating that the new Cel⁺ genes provide all three functions. Spontaneous cellobiose negative mutants also become arbutin and salicin negative. Analysis of β-glucoside positive revertants of these mutants indicates that there are separate loci for utilization of each of the β-glucoside sugars. The genes are closely linked and may be activated from a single locus. A fourth gene at an unknown location increases the growth rate on cellobiose. The cel genes constitute a second cryptic system for β-glucoside utilization in *E. coli* K12.

Introduction

Bacterial systems frequently provide the most explicit models for studying mechanisms that organisms use to evolve new metabolic functions. They can be manipulated genetically, produce numerous generations in a short time span, and allow comparisons of evolved mutant strains with unevolved antecedents. There is substantial evidence from these systems that one mechanism for the evolution of new functions is the acquisition of mutations altering regulation and catalytic activity of enzymes in existing pathways. Examples include the amidase system of *Pseudomonas putida* (reviewed in Clarke 1978), the utilization of unusual pentoses and pentitols by *Klebsiella pneumoniae* (reviewed in Mortlock 1981), aerobic metabolism of 1,2-propanediol by *E. coli* via the fucose pathway (Cocks et al.

1. Key words: *Escherichia coli*, cellobiose, cryptic genes, β-glucosides.

Address for correspondence and reprints: Barry G. Hall, Biological Sciences Group, U-44, University of Connecticut, Storrs, Connecticut 06268.

© 1984 by The University of Chicago. All rights reserved.
0737-4038/84/0102-0006$02.00
1974), and the evolved β-galactosidase (EBG) system in *Escherichia coli* (reviewed in Hall 1983).

However, organisms need not modify existing systems to increase physiologic versatility. Reactivation of silent sequences by mutation provides a second, though not the only other, means of acquiring new metabolic capabilities. Cryptic genes have been detected in a multitude of bacterial species, including a variety of amino acid synthetic pathways in *Neisseria gonorrhoeae* (Juni and Heym 1980), *Salmonella* sp. (Lederberg 1947), *Pasteurella pestis* (Englesberg and Ingram 1957), and genes for citrate catabolism in *E. coli* (Hall 1982) and for β-glucoside catabolism in *E. coli* (Prasad and Schaefer 1974). Cryptic sequences are not limited only to prokaryotes. Investigators have also identified mutations activating cryptic genes coding for sucrose catabolic functions in *Saccharomyces cerevisiae* (Carlson et al. 1981). These discoveries serve not only to confuse taxonomists (Buchanan and Gibbons 1974) but also to identify potentially important elements of adaptive change. We have chosen the well-characterized β-glucoside system of *E. coli* to examine the role of cryptic genes in adaptive evolution.

Wild-type *E. coli* cannot utilize any β-glucoside sugars as carbon or energy sources. Mutants of *E. coli* have been isolated which utilize the aryl β-glucosides arbutin and salicin, but these do not utilize the most common β-glucoside in nature, the disaccharide cellobiose (Schaefer 1967). Other members of the family Enterobacteriaceae do utilize cellobiose (Schaefer and Malamy 1969), suggesting that it might be possible for *E. coli* exposed to appropriate selective pressures to evolve this capacity. Arbutin and salicin positive (Bgl+) *E. coli* mutants express the *bgl* operon which is normally cryptic in wild-type strains (Schaefer 1967). The *bgl* operon is located at 83 min on the *E. coli* map (Bachmann 1983) and cotransduces with the *ilv* gene cluster at 84 min (Schaefer and Maas 1967). Spontaneous Bgl+ mutations that occur at a high frequency (10⁻⁵ to 10⁻⁷) are caused by insertions of the mobile DNA sequences IS1 or IS5 at a regulatory site *hglR* (Reynolds et al. 1981). The insertions cause inducible expression of two structural genes, *bglB* and *bglC*. The gene *bglC* codes for a transport protein that is a β-glucoside specific enzyme II of the phosphoenolpyruvate-dependent phosphotransferase system (Fox and Wilson 1968); *bglB* specifies phosphoglucosidase B, a hydrolase for phosphorylated β-glucosides. A third gene, *bglS*, codes for a positive regulator of the operon (Prasad and Schaefer 1974).

Mutations outside of the *bgl* operon also allow expression of the *bgl* enzymes. *bglY* mutations map at 28 min and cotransduce with the *trp* operon (DeFez and DeFelice 1981). Mutations in *gyrA* located at 46 min or *gyrB* located at 82 min also activate the *bgl* genes (Dinardo et al. 1982). All of these mutants are inducible for expression of the *bgl* operon.

**Material and Methods**

**Culture Media and Growth Conditions**

Cultures were grown at 37 C with aeration. Minimal media (Hall and Hartl 1974) contained .2% of the appropriate sugar as a carbon source. When required, amino acids were added to a concentration of 100 μg/ml, purines were added to a concentration of 40 μg/ml, spectinomycin or ampicillin was added to a concentration of 50 μg/ml. Solid media contained 1.5% agar.
Indicator Media

All mutants were isolated on either tetrazolium or MacConkey solid indicator medium. Salicin and cellobiose tetrazolium plates were prepared according to Miller (1972, p. 54). Salicin, arbutin, and cellobiose MacConkey media were prepared from MacConkey agar base according to instructions provided by Difco. Colonies unable to ferment the added sugar are white on this medium, whereas fermenting colonies are pink or red.

Matings and Transductions

Matings and transductions were carried out according to Miller (1972). Transductions were mediated by P1 cm clrlO0 (Rosner 1972).

In Vivo Alkaline Phosphatase Determinations

Alkaline phosphatase phenotypes were detected by plate assay (Willsky et al. 1973). Colonies grown on glucose minimal solid medium and glucose solid medium limited for phosphate (5 \times 10^{-5} \text{M}) and containing 0.05 M tris (hydroxymethyl) amino-methane (Tris) buffer pH 8.0, were overlaid with 1 mM p-nitrophenyl phosphate in 0.5 M tris pH 8.0 and 1% agar. Constitutive colonies turn a bright yellow color in 10 min on both types of media. Inducible colonies remain white on glucose minimal medium but turn yellow in a few minutes on limited phosphate medium. Alkaline phosphatase negative mutants remain white on both types of media.

Growth Rates

Cultures were grown in cellobiose or salicin minimal medium to a density of approximately 5 \times 10^8 \text{cells/ml}, washed once, then distributed to flasks containing \beta-glucoside or glucose minimal medium so that the density was approximately 1 \times 10^8 \text{cells/ml}. Turbidity was monitored in a Gilford spectrophotometer at 600 nm for at least two doublings of each culture. The growth rates are reported as the first-order growth rate constant, calculated by the slope of the least squares fit of ln(A_{o0}) versus time (hours), from a minimum of eight points for each of at least three independent growth rate determinations, for each strain on each sugar.

Reversion Frequencies

Strains were grown overnight in glucose minimal medium. Cultures were washed once, concentrated 10-fold, and plated at appropriate dilutions on arbutin, salicin, cellobiose, and glucose minimal plates. Reversion frequencies are reported as the number of revertants per cell plated.

Results

Isolation of Cellobiose Utilizing Mutants

Cellobiose- (Cel+) utilizing mutants were selected in three steps from the wild-type parent 1011A (table 1). In the first step, strain 1011A was streaked on salicin tetrazolium indicator plates (Lederberg 1948), then incubated at 30 C until salicin fermenting papillae appeared on the parent colonies. A salicin positive, arbutin positive (Sal+, Arb+) strain, MK1, was purified from one of these papillae. In the second step, MK1 was streaked onto cellobiose tetrazolium plates, and the same procedure was used to obtain papillae. The papillae formed pink colonies
<table>
<thead>
<tr>
<th>Strains</th>
<th>Relevant Genotype and Phenotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1011A Invent</td>
<td><em>HfrC</em> spc <em>lacZ</em> (deletion W4680) <em>egbA</em> (deletion 11) <em>amp</em></td>
<td>This study</td>
</tr>
<tr>
<td>MK1</td>
<td><em>HfrC</em> spc <em>lacZ</em> (deletion W4680) <em>egbA</em> (deletion 11) <em>amp</em> <em>bgIR</em></td>
<td>This study</td>
</tr>
<tr>
<td>MK2</td>
<td><em>HfrC</em> spc <em>lacZ</em> (deletion W4680) <em>egbA</em> (deletion 11) <em>amp</em> <em>bgIR</em> <em>Cel</em></td>
<td>This study</td>
</tr>
<tr>
<td>CSH75</td>
<td><em>F</em>^-* rpsL ilv ara leu lacY purE trp his argG malA xyl mtl metA or B thi proc gal*</td>
<td>Cold Spring Harbor Laboratory</td>
</tr>
<tr>
<td>SJ30 Invent</td>
<td><em>F</em>^-* rpsL ilv ara leu purE trp his argG metA or B thi malA xyl mtl gal lacZ* (deletion W4680) lacY (deletion W4680) lacY</td>
<td>This study</td>
</tr>
<tr>
<td>SJ30A1</td>
<td><em>bgIr</em>^-*</td>
<td>This study</td>
</tr>
<tr>
<td>AE341 Invent</td>
<td><em>F</em>^-* lac* (deletion x74) tna::Tn10 <em>bgIY supE</em></td>
<td>A. Wright</td>
</tr>
<tr>
<td>JF201 Invent</td>
<td><em>lac</em> (deletion x74) <em>bgI- pho</em> (deletion 201) <em>ara</em> <em>gyrA</em> <em>thi</em></td>
<td>A. Wright</td>
</tr>
<tr>
<td>JF201T</td>
<td><em>bgI- pho</em> (deletion 201) tna::Tn10 transductant of JF201 (donor AE341)</td>
<td>This study</td>
</tr>
<tr>
<td>MK1201</td>
<td><em>ilv</em> <em>bgI- pho</em> (deletion 201) <em>EMS induced Ilv^-</em> mutant of JF201</td>
<td>This study</td>
</tr>
<tr>
<td>MK9</td>
<td><em>F</em>^-* ilv rpsL trp his argG metA or B ara leu lacZ* (deletion W4680) lacY <em>Cel</em>^-*</td>
<td>This study</td>
</tr>
<tr>
<td>MK91 Invent</td>
<td><em>bgI- pho</em> (deletion 201) <em>Cel</em>^-* Transductant of MK9 (donor JF201)</td>
<td>This study</td>
</tr>
<tr>
<td>MK93 Invent</td>
<td><em>ilv</em> <em>bgI- pho</em> (deletion 201) <em>Cel</em>^-* Transductant of MK93 (donor JF201)</td>
<td>This study</td>
</tr>
<tr>
<td>MK94 Invent</td>
<td><em>bgI- Cel</em>^-* <em>Cel</em>^-* Transductant of MK94 (donor JF201)</td>
<td>This study</td>
</tr>
<tr>
<td>MK79 Invent</td>
<td><em>bgI- pho</em> (deletion 201) <em>Cel</em>^-* Transductant of MK79 (donor MK9)</td>
<td>This study</td>
</tr>
<tr>
<td>MK797</td>
<td><em>bgI- pho</em> (deletion 201) <em>Cel</em>^-* Transductant of SJ30 (donor JF201)</td>
<td>This study</td>
</tr>
<tr>
<td>CSH62</td>
<td><em>HfrH thi</em></td>
<td>Miller 1972</td>
</tr>
<tr>
<td>CSH62T</td>
<td><em>HfrH thi bgI- pho</em> (deletion 201) tna::Tn10 <em>Cel</em>^-* Transductant of CSH62 (donor JF201T)</td>
<td>This study</td>
</tr>
<tr>
<td>CSH62TC</td>
<td><em>HfrH thi bgI- pho</em> (deletion 201) tna::Tn10 <em>Cel</em>^-* Transductant of CSH62T (donor MK91)</td>
<td>This study</td>
</tr>
<tr>
<td>GMS343 (CGSC 5496)</td>
<td><em>F</em>^-* aroD6 argE3 lacY1 galK2 man-4 mtl-l rpsL700 tsx-29? supE44? <em>l-</em></td>
<td>B. Bachmann</td>
</tr>
</tbody>
</table>

**NOTE.**—EMS = ethyl methane sulfonate.
when restreaked to MacConkey cellobiose plates (Miller 1972). In the third step, one pink colony was restreaked on cellobiose minimal medium. A single large colony was reisolated and designated MK2; MK2 was red on MacConkey cellobiose medium. Both MK1 and MK2 were, like the parent 1011A, HfrC, Lac-, as well as ampicillin and spectinomycin resistant, and were therefore not contaminants. Strain 1011A did not utilize any β-glucosides. Both MK1 and MK2 utilized arbutin and salicin, but only MK2 utilized cellobiose.

Genetic Analysis

To determine whether cellobiose utilization arose as a consequence of mutations in the \( bgl \) operon, the locus conferring the Cel+ phenotype was mapped. If the \( cel \) mutation was in the \( bgl \) operon it should cotransduce with the \( ilv \) locus at 84 min. When MK1 was used as a donor to transduce MK1201 (\( ilv \) \( bgl \) deletion) to \( Ilv^+ \), 25% of the \( Ilv^+ \) transductants were arbutin and salicin positive. The maximum length of a chromosomal region carried by a \( P1 \) phage transducing particle is equivalent to 2 min on the \( E. coli \) genetic map (Bachmann 1983). The distance between \( ilv \) and the \( β \)-glucoside markers is therefore 1 min by the mapping function of Wu (1966), confirming that MK1 is a \( bglR^+ \) mutant. When a Cel+ strain was used as a donor in the same type of experiment, none of the \( Ilv^+ \) transductants were Cel+, indicating that the \( cel \) locus lay outside of the \( bgl \) operon.

Conjugation experiments indicated that the \( cel \) marker was located in the region near \( aroD \) and \( manA \) (data not shown). To locate the \( cel \) mutation more precisely, CSH62TC (\( bgl \) deletion Cel+) was used as a donor to transduce GMS343 to \( Aro^+ \), and 500 \( Aro^+ \) colonies were scored for mannose and cellobiose utilization (table 2). Among the 17 \( Aro^+ \) Man+ cotransductants, only three were Cel+, showing that \( cel \) does not lie between \( aroD \) and \( manA \). Similarly, the fact that only three of the 144 \( Aro^+ \) Cel+ cotransductants were Man+ shows that \( manA \) does not lie between \( cel \) and \( aroD \). The gene order is thus \( cel, aroD, manA \). The cotransduction frequency of \( aroD \) with \( cel \) is .294; thus the distance between \( cel \) and \( aroD \) is .67 min by the mapping function of Wu (1966). The \( cel \) locus is therefore at 37.8 min on the \( E. coli \) map.

The observation that the gene for cellobiose utilization lies on the opposite side of the map from the \( bgl \) operon does not preclude the possibility that the \( bgl \) operon provides a necessary function for cellobiose metabolism. To explore this possibility we introduced a deletion of the entire \( bgl \) operon into a cellobiose

| Table 2 Mapping of the cel Locus by Transduction |
|----------|----------|---------------|-----------------|-----------------|
| Donor | Recipient | Selected Marker | Recombinant Class | Number of Recombinants |
| CSH62TC | GMS343 | Aro+ | Cel+ Man- | 144 |
| bgl Cel+ | aroD manA | Cel+ Man+ | 3 |
| | | Cel+ Man+ | 14 |
| | | Cel Man+ | 339 |
| | | Total | 500 |
| Cotransduction | Aro+Cel+ | Aro+Man+ | .294 | .034 |
positive strain and asked if the strain could still utilize cellobiose. The donor, JF201, carries a deletion covering the bgl operon and the neighboring phoS and phoT genes and therefore exhibits an alkaline phosphatase constitutive phenotype (Reynolds 1983). When the deletion was introduced into strain MK9 (Cel⁺) by cotransduction with ilv, 34 out of 100 transductants were alkaline phosphatase constitutive and therefore deleted for the bgl operon, yet all alkaline phosphatase constitutive transductants continued to grow on cellobiose. In contrast, when the same deletion was introduced into the ilv⁻ bglR⁻ mutant SJ30A1, 38% of the Ilv⁺ transductants were alkaline phosphatase constitutive, and none of these utilized arbutin and salicin, demonstrating that the deletion did in fact eliminate the Bgl⁺ phenotype. These results showed that the bgl operon did not provide any function necessary for cellobiose catabolism. To determine if the bgl operon could contribute to faster growth on cellobiose, though the operon was not required for cellobiose utilization, the growth rate of the Bgl⁺ Cel⁺ strain MK94 was compared with that of the bgl deletion Cel⁺ strain MK91 on cellobiose (table 3). MK91 grew more slowly than MK94 on cellobiose. However, MK91 grew more slowly than MK94 on glucose as well, indicating that the difference in growth rates is not specific to cellobiose. When normalized to the growth rates on glucose, there was no significant difference in growth rates between the two strains. The bgl operon does not, therefore, contribute in any way to growth on cellobiose.

Both MK91 and MK94 were derived from MK9, an exconjugant of a mating between MK2 and a Cel⁻ recipient (SJ30). MK91 and MK94 thus received a number of loci from MK2 which were not linked to the cel locus at 38 min. However, MK797 was constructed by a transductional cross between a Cel⁺ donor and a bgl deletion Cel⁻ recipient (MK79) and therefore received only those markers that were linked to the cel locus. Both MK91 and MK94 were red on MacConkey cellobiose medium, but Cel⁺ transductants, such as MK797, were pink on this medium, suggesting that the transductants ferment the sugar poorly. To determine if there was a growth rate difference between “red” and “pink” strains, the growth rates on cellobiose and glucose of the “red” strains MK91 and MK94 were compared with that of the “pink” strain MK797. The normalized growth rates of both “red” strains were significantly higher than that of MK797, suggesting that a second mutation enhances the growth rate on cellobiose. The second mutation was present in strain MK2, was transferred to the recipient during the construction of MK9 by conjugation, and was therefore present in the direct descendants of MK9. Because it was not transferred to the recipient during the

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Growth Rates of Representative Strains on Cellobiose and Salicin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain</td>
<td>Cellobiose/</td>
</tr>
<tr>
<td></td>
<td>Glucose</td>
</tr>
<tr>
<td>MK94</td>
<td>0.240 ± 0.029</td>
</tr>
<tr>
<td>MK91</td>
<td>0.197 ± 0.006</td>
</tr>
<tr>
<td>SJ30A1</td>
<td>ND</td>
</tr>
<tr>
<td>MK797</td>
<td>0.126 ± 0.006</td>
</tr>
</tbody>
</table>

Note.—Mean growth rates on the substances above each column are reported as the first-order growth rate constant ±95% confidence limits. Normalized growth rates are the means of the ratios of growth rates on cellobiose or salicin to growth rates on glucose. ND = not determined.
construction of MK797 by transduction, the second mutation (enhancer) is apparently unlinked to the cel locus.

The cel System Specifies Multiple Functions

The ability of strain MK2 to utilize arbutin and salicin did not depend on expression of the bgl operon. MK91, and other Cel+ strains deleted for the bgl operon, remained fully capable of growth on arbutin and salicin. The Cel system thus provides all of the transport and hydrolase functions required for the metabolism of three β-glucoside sugars, arbutin, salicin, and cellobiose.

If MK94 expresses both the bgl operon and the cel locus allowing salicin utilization, then MK94 should grow more rapidly on salicin than either the bglR+ Cel- mutant SJ30A1 or the bgl deletion Cel+ strain MK91. Table 3 (last column) shows that when growth rates on salicin were normalized to glucose, MK94 grew more rapidly on salicin than either MK91 or SJ30A1. The combined growth rates of MK91 and SJ30A1 equaled the growth rate of MK94 on salicin, indicating that the systems are additive and function independently.

Clearly, the cel system specifies three identifiable functions: utilization of arbutin, salicin, and cellobiose. We did not observe any segregation of the cellobiose, arbutin, and salicin phenotypes in a variety of transductions, indicating that either these functions are tightly linked or a single genetic element specifies all three functions. To determine whether these functions were specified by a single genetic element or by three separate genetic elements, spontaneous cellobiose negative mutants were selected. MK91 was streaked onto a cellobiose MacConkey plate and incubated until white papillae appeared on the colonies. Fifteen papillae were picked from different regions of the plate to avoid isolation of siblings. All 15 isolates were cellobiose negative and were also arbutin and salicin negative. The mutants carried the same five auxotrophic markers as MK91, indicating they were indeed derivatives of the parent. In some strains the alkaline phosphatase structural gene was no longer expressed due to unidentified mutations. In those cases presence of the bgl deletion was verified by transducing out the deletion and showing that the recipient had gained the alkaline phosphatase constitutive phenotype.

Isolation of single-step mutants which failed to utilize any of the three β-glucosides suggested that all three functions were specified by a single genetic element. To explore this further, spontaneous cellobiose, arbutin, and salicin positive revertants were selected from various β-glucoside negative mutants. One of these β-glucoside negative mutants, MK912, was streaked on arbutin, salicin, or cellobiose MacConkey plates, and papillae were re-isolated using the same procedure as for isolation of cellobiose negative mutants (fig. 1). Twenty-five revertants were isolated following selection on cellobiose. Although all of these revertants used cellobiose, none utilized arbutin or salicin. From one such revertant, MK9123, several arbutin revertants were isolated. One class of arbutin revertants grew on all three β-glucosides. A second class of revertants grew on arbutin and cellobiose but not salicin. From one Arb+ Cel+ revertant (MK912301), several salicin-utilizing revertants were isolated, and these revertants (e.g., MK9123011) used all three β-glucosides. From the mutant MK912 which utilized no β-glucosides, revertants utilizing all three β-glucosides could be isolated in three sequential steps, indicating that three separate genes are responsible for utilization of the three β-glucoside sugars, arbutin, salicin, and cellobiose.
When 10 single-step arbutin revertants were isolated directly from MK912, two classes were obtained. The first class (MK91243) utilized only arbutin, but not cellobiose or salicin, consistent with a separate arbutin gene. The second class (MK91236) utilized all three β-glucosides. Twenty single-step salicin revertants were similarly isolated from MK912, and all of these belonged to a single class which utilized all three β-glucosides. All functions were regained in a single step, suggesting coordinate control at a single locus of the separate β-glucoside genes. Nine second-step salicin revertants were isolated from MK9123, and again, the same class was obtained, the class which utilized all three β-glucosides. All revertants which utilize salicin utilize the other β-glucosides whether they are selected in one step (MK91226), two steps (MK912302), or three steps (MK9123011), suggesting that salicin utilization is associated with a locus for activation of the cel genes.

Revertants with all three β-glucoside functions could be selected in either a single step (MK91226) or by three sequential steps (MK9123011). If three separate mutational events were required for reversion of the three functions, the reversion frequencies would probably be very low. The spontaneous rate of point mutations in E. coli is typically $10^{-8}$ to $10^{-9}$; however, the highest mutation frequency observed for the single $bglR$ mutation is $10^{-5}$ (Reynolds et al. 1981). If the cel mutations occur at a similar frequency, the expected frequency of three concerted cel mutations is $10^{-15}$. The frequency of revertants as determined by direct plating on minimal medium (see Material and Methods) was approximately $10^{-8}$ whether these were selected on arbutin, salicin, or cellobiose. The overwhelming majority of these utilized all three β-glucosides, further indicating that the cel cluster can be activated by a single mutation. The reversion frequencies do not depend on rec function. Introduction of a recA allele into MK912 did not alter reversion frequencies for any of the β-glucosides, indicating these reversions result from either a point mutation or genetic rearrangement involving illegitimate recombination mechanisms.

**Discussion**

The observation that introduction of a deletion of the bgl operon into Cel+ strains does not prevent growth on any of the three β-glucosides shows that the
Evolution of Cellobiose Utilization

cel system expresses all three functions independently of the bgl operon. Analysis of mutants and revertants indicates that functions for utilization of arbutin, salicin, and cellobiose are specified by separate genetic elements. We designate the arbutin gene celA, the salicin gene celS, and the cellobiose gene celC. Expression of all three functions can be lost by a single mutational step and can be regained at a frequency consistent with activation at a single locus. Nonetheless, components can be activated individually. Because of this, the genetic state of specific cel alleles in any strain cannot be deduced from the phenotype, and gene designations do not indicate allele assignments. These designations are not meant to imply enzyme functions, which are unknown, but to indicate individual genetic components identified by revertant analysis. The lack of recombination between these genes in transductional crosses suggests the genes are tightly linked. The cel genes thus comprise a second cryptic gene cluster for β-glucoside utilization which can be activated in Escherichia coli K12.

Single-step mutants from MK1 were pink on MacConkey cellobiose medium. Two sequential selections were required to obtain a mutant which fermented cellobiose well and exhibited a red phenotype on MacConkey cellobiose medium. Cel+ transductants also exhibit the pink phenotype and have a decreased growth rate on cellobiose compared with "red" strains, indicating that single-step Cel+ mutants from MK1 and the Cel+ transductants probably express only one of two genetic elements involved in cellobiose utilization. One element, within the cel gene cluster, appears to be sufficient for growth on cellobiose. The second element, unlinked to the cel locus and therefore not present in Cel+ transductants, apparently enhances cellobiose utilization. Since Cel+ transductants exhibit the pink phenotype, there is still another cel gene, which we call celM (for Modifier of cellobiose-utilizing activity), at an unknown location, separated from the cel cluster by a distance too large for cotransduction.

Two lines of evidence indicate that the cel mutations do not alter the substrate specificity of a functioning system but do decryptify a silent gene cluster. First, three separable functions can be ascribed to the cel genes, and expression of these functions does not require separate mutational events. It is improbable that genes of another pathway could gain all of the new functions simultaneously. Second, other members of the family Enterobacteriaceae display the same phenotypes as E. coli Cel+ mutants. There is wide variation in the utilization of β-glucosides among Enterobacteriaceae (Schaeffer and Malamy 1974). Wild-type Klebsiella sp. ferment the full range of β-glucosides, arbutin, salicin, and cellobiose. Citrobacter sp. ferments cellobiose well and aromatic β-glucosides poorly or not at all, while wild-type Proteus vulgaris metabolizes aromatic β-glucosides but not cellobiose. Wild-type Salmonella sp. are similar to E. coli in that they do not use any of the β-glucosides. Salmonella mutants have been isolated which grow on cellobiose but not on either arbutin or salicin (Schaeffer and Scheinken 1968). Arbutin-utilizing Salmonella mutants have been obtained in a second step from cellobiose positive strains (Schaeffer and Malamy 1974). From β-glucoside negative mutants we isolated revertants that grew only on cellobiose and that may be genetically equivalent to cellobiose positive Citrobacter and Salmonella strains. Similarly, we isolated second-step revertants which grow on cellobiose and arbutin, parallel Salmonella arbutin- and cellobiose-utilizing mutants. Most of our strains express the same range of functions found in Klebsiella sp. These observations suggest that the cel genes of E. coli did not evolve independently, from an existing
pathway, but are silent homologues of β-glucoside genes found in other Enterobactericiae.

Two functions common to the cel cluster and the bgl operon are the utilization of arbutin and salicin. The cel gene cluster is located nearly 180° from the bgl operon on the E. coli map. A number of functionally related gene pairs involved in central metabolism lie approximately either 90° or 180° apart in E. coli, and it has been proposed that the ancestral chromosome may have undergone two sequential duplications (Zipkas and Riley 1975; Riley and Anilionis 1978). These observations raise the question of whether the two β-glucoside systems arose independently or from an ancient genome duplication.

The maintenance of two cryptic β-glucoside systems in E. coli would appear to have no selective advantage. Yet either can be activated by a single mutation, indicating that the structural genes are intact though not expressed in both operons. Both β-glucoside systems might be activated when aryl β-glucosides are a primary carbon source. Mutants expressing both the cel and bgl genes grow faster on salicin than mutants expressing a single system. Cel+ revertants expressing all β-glucoside utilization functions were isolated in a single step on media containing arbutin or salicin, suggesting that selection for expression of both cryptic systems could occur simultaneously or sequentially on the same carbon source.

There are a number of reports suggesting that unneeded functions reduce fitness (Zamenhoff and Eichorn 1967; Dykhuizen 1978). We obtained spontaneous β-glucoside negative mutants easily by growing Cel+ mutants on rich (MacConkey) medium, indicating that there was considerable selective advantage for cryptification of the cel genes when other carbon sources are available. The bgl operon is inducible, and there should be little advantage to cryptification of genes under transcriptional regulation. However, it has been suggested that since toxic cyanogenic β-glucosides are found in nature, transcriptional control may be inadequate to protect the cell against poisonous compounds if these compounds are inducers (Reynolds et al. 1981). Thus, there are at least two plausible natural conditions that might favor silencing of these systems—the presence of multiple carbon sources or the presence of toxic β-glucosides. Hall et al. (1983) have recently presented a systematic discussion of the role of cryptic genes in microbial evolution. They have suggested that repeated cryptification and decryptification of genes may be a means of long-term regulation of rarely utilized functions and may account for the retention of those genes in the population. Mathematical analyses of this question support their model (Hall et al. 1983; Li 1984). In particular, Li (1984) has shown that a gene that spends an average of 25,000 generations in the cryptic state and only 200 generations in the decryptified state will not be lost owing to irreversible mutational inactivation, even when the rate of decryptification is only $10^{-7}$ per generation. Numerous examples of cryptic systems in microorganisms capable of reactivation under selective conditions (Lederberg 1947; Englesberg and Ingram 1957; Juni and Heym 1980; Hall 1982) support such a model. During this study we have isolated E. coli strains that exhibit every phenotype for β-glucoside utilization found in naturally occurring populations of Enterobactericiae. The cel genes provide a specific system for testing the hypothesis that there is selection for mutations which maintain silent genes for pathways that are unneeded, or even deleterious, under the most prevalent natural conditions.
Acknowledgments

We are grateful to Andrew Wright for the gift of several strains and for many helpful discussions. Barry G. Hall is supported by Research Career Development Award 1 K04 AI00366 from the National Institute of Allergy and Infectious Diseases of the U.S. Public Health Service. This work was supported by grants from the National Institutes of Health and the University of Connecticut Research Foundation.

LITERATURE CITED


WALTER M. FITCH, reviewing editor

Received August 19, 1983; revision received October 24, 1983.