Cryptic Genes for Cellobiose Utilization in Natural Isolates of *Escherichia coli*

Barry G. Hall and Paul W. Betts

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

Manuscript received August 8, 1986
Revised copy accepted November 26, 1986

ABSTRACT

The ECOR collection of natural *Escherichia coli* isolates was screened to determine the proportion of strains that carried functional, cryptic and nonfunctional genes for utilization of the three β-glucoside sugars, arbutin, salicin and cellobiose. None of the 71 natural isolates utilized any of the β-glucosides. Each strain was subjected to selection for utilization of each of the sugars. Only five of the isolates were incapable of yielding spontaneous β-glucoside-utilizing mutants. Forty-five strains yielded cellobiose$^+$ mutants, 62 yielded arbutin$^+$ mutants, and 58 strains yielded salicin$^+$ mutants. A subset of the mutants was screen by mRNA hybridization to determine whether they were expressing either the *cel* or the *bgl* β-glucoside utilization operons of *E. coli* K12. Two cellobiose$^+$ and two arbutin$^+$-salicin$^+$ strains failed to express either of these known operons. It is concluded that there are at least four gene clusters specifying β-glucoside utilization functions in *E. coli* populations, and that all of these are normally cryptic. It is estimated that in any random isolate the probability of any particular cluster having been irreversibly inactivated by the accumulation of random mutations is about 0.5.

CRYPTIC genes have been defined as "phenotypically silent DNA sequences not normally expressed during the life cycle of an individual, but capable of activation as a rare event in a few members of a population by mutation, recombination, insertion elements, or other genetic mechanisms" (HALL, YOKOYAMA and CALHOUN 1983). The microbial literature provides a variety of examples of activation of silent genes by mutations selected in the laboratory. These examples include cryptic genes for various biosynthetic pathways in Lactobacillus (MORISHITA *et al.* 1974), for citrate transport in *Escherichia coli* (HALL 1982), and for β-glucoside utilization in *E. coli* and in Salmonella (SCHAEFLER and MALAMY 1969). In addition, a cryptic gene for alcohol dehydrogenase (*Adh*IV) has recently been identified in a eukaryotic organism, yeast (PAQUIN and WILLIAMSON 1986). While there are many other cases of activation of silent genes in bacteria (see HALL, YOKOYAMA and CALHOUN 1983), the above examples share the property that *typical* wild-type isolates obtained from nature lack some function, and that the function in question can be restored by mutation following selection in the laboratory.

Since genes for these functions are not typically expressed in nature, and thus must make no positive contribution to the fitness of the organism, we have considered why those genes are not lost due to the accumulation of inactivating mutations (HALL, YOKOYAMA and CALHOUN 1983). We considered three alternative states of a gene: a functional allele, which is expressed under appropriate circumstances; a cryptic allele, which is not expressed but which can revert to a functional allele by mutation; and a nonfunctional allele, which is irreversibly inactivated. To explain the persistence of cryptic alleles we proposed a model in which functional alleles are advantageous in one environment, while cryptic and nonfunctional alleles are advantageous in another alternate environment. Specifically, we suggested that in the "normal" environment functional alleles would be disadvantageous, whereas in the infrequently encountered "alternative" environment it would be advantageous to possess a functional allele; and that environmental fluctuation could maintain all three kinds of alleles in the population. LI (1984) dealt mathematically with this model, and showed that, with reasonable fitnesses and mutation rates, environmental fluctuations could retain all three kinds of alleles in the population without cryptic alleles having any selective advantage over nonfunctional alleles.

There has been little systematic investigation of cryptic genes *per se*. One of the few systems that has been investigated systematically is the system for the utilization of β-glucoside sugars by members of the bacterial family the Enterobacteriaceae (SCHAEFLER and MALAMY 1969). The system for the catabolism of these sugars has been studied in detail at the genetic (SCHAEFLER and MAAS 1967; PRASAD and SCHAEFLER 1974), physiological (SCHAEFLER 1967; SCHAEFLER and SHIENKEIN 1968; PRASAD and SCHAEFLER 1974) and molecular (REYNOLDS, FELTON and WRIGHT 1981; REYNOLDS *et al.* 1985) levels in *E. coli*. The three naturally occurring β-glucoside sugars of inter-

est are the disaccharide cellobiose, which is the repeating unit of cellulose, and the aryl β-glucosides arbutin and salicin. *Klebsiella* species typically metabolize all three β-glucosides, most *Citrobacter* strains use cellobiose but not arbutin or salicin, while *Proteus vulgaris* strains use the aryl β-glucosides but not cellobiose. *Salmonella* and *E. coli* are described as being unable to utilize any β-glucoside sugars (SCHAEFLER and MINTZER 1959; SCHAEFLER and MALAMY 1969). Both *Salmonella* and *E. coli*, however, mutate easily to β-glucoside positive phenotypes at frequencies as high as 10^{-5} (SCHAEFLER and MINTZER 1959; SCHAEFLER and SCHENKEIN 1968; REYNOLDS, FELTON and WRIGHT 1981). The β-glucoside utilization system has been investigated at the genetic and molecular level only in *E. coli* K12, where it has been shown that the genes for β-glucoside utilization are encoded by the normally silent *bgl* locus at 84 min on the *E. coli* map (SCHAEFLER and MAAS 1967; BACHMANN 1983), and that inducible expression of these genes is activated by the insertion of a mobile DNA element, IS1 or IS5, upstream from the structural genes (Reynolds *et al.* 1985). In addition to the silent *bgl* operon there is a separate, constitutively expressed gene, *bglA*, that encodes an arbutin-specific enzyme, phospho-β-glucosidase A. Constitutive expression of *bglA* is found in a variety of natural *E. coli* strains (SCHAEFLER and MALAMY, 1969), but does not give a arbutin positive phenotype because arbutin is not transported and phosphorylated unless either the *bgl* operon or the recently discovered, silent, *arbT* (KRICKER and HALL 1987) gene is expressed.

Based upon 10 *E. coli* isolates, SCHAEFLER and MALAMY (1969) reported that wild type *E. coli* strains (genotypically $bglR^0$) are unable to utilize any β-glucoside sugars, and that β-glucoside$^+$ mutants (genotypically $bglR^+$) are able to grow only on the aryl β-glucosides arbutin and salicin, not on the disaccharide cellobiose.

We have recently shown that *E. coli* K12 can spontaneously mutate to utilize cellobiose. The mutation that allows cellobiose utilization is in the *cel* operon which is located at 37.8 min on the *E. coli* map, on the opposite side of the chromosome from the *bgl* operon (KRICKER and HALL 1984). Like the *bgl* operon, the *cel* operon specifies a transport protein and a hydrolase protein, both of which must be functional in order for growth on cellobiose to occur (KRICKER and HALL 1987). The *cel* operon is not expressed in wild-type *E. coli* K12 strains, but is expressed constitutively in the decryptified cel^+ mutant strains (KRICKER and HALL 1987). The *cel* operon proteins act not only on cellobiose, but on arbutin and salicin as well, so that strains which express the *cel* operon are phenotypically cellobiose$^+$, arbutin$^+$, and salicin$^+$ (KRICKER and HALL 1984). *E. coli* K12 thus possesses two cryptic operons for the utilization of β-glucoside sugars (KRICKER and HALL, 1984). We have cloned the genes of the *cel* operon (HALL, BETTS and KRICKER 1986) and molecular studies of these genes are in progress.

We are using the cellobiose system of *E. coli* as a model to study the evolution and maintenance of cryptic genes in populations. One key aspect of the model for retention of cryptic genes is that there must be conditions in the normal environment that select against the functional alleles (HALL, YOKOYAMA and CALHOUN 1983). We have recently shown that for the activated *cel* operon of *E. coli* K12 there is such a condition: the presence of both cellobiose and an alternative carbon source such as glycerol in the growth medium strongly selects against cel^+, but not against cryptic (wild-type) alleles (HALL, BETTS and KRICKER 1986).

If these studies are to be of value in understanding the roles of cryptic genes it is necessary to extend the studies to include a variety of natural isolates. If the cellobiose genes of *E. coli* are truely cryptic, in the sense of having retained their information in the face of mutational pressure during a long period of silence, then it should be the case that: (*1*) the majority of natural isolates of *E. coli* are, indeed, cellobiose negative, and 2) a substantial fraction of those natural isolates can give rise to spontaneous cellobiose utilizing mutants. With this in mind we have undertaken a survey of a collection of natural *E. coli* isolates in order to determine the proportion of functional (active), cryptic (inactive, but revertable) and nonfunctional (irreversibly inactivated) alleles in the population. Because the earlier study (SCHAEFLER and MALAMY 1969) had included only a small number of strains, and because it was not known at that time that *E. coli* could mutate to the utilization of cellobiose, our survey has included alleles for the utilization of all three β-glucosides, arbutin, salicin and cellobiose.

MATERIALS AND METHODS

Culture media and conditions: Minimal media were prepared from a phosphate buffered mineral salts solution (HALL and HARTL, 1974) plus a sugar as a carbon source. Plates contained 1.5% agar as a solidifying agent. Glucose and cellobiose were used at a 0.2% (w/v) concentration, while arbutin and salicin were used at a concentration of 4 mM.

MacConkey medium is an indicator medium which allows growth of cells whether or not they ferment the added sugar. If the carbon source is fermented the colonies are red or pink, whereas the colonies are white if they fail to ferment the added sugar. MacConkey medium contained 1% (w/v) of the added sugar, and was prepared as directed by the manufacturer (Difco).

***E. coli* strains:** The ECOR collection of natural isolates (OCHMAN and SELANDER, 1984) was obtained from D. DYKHUIZEN.

Strain MK1 is a $bglR^+$ mutant of *E. coli* K12, and strain

MK2 is a *celR*$^+$ mutant of MK1 (KRICKER and HALL 1984, 1987). Strain JF201 is a *bgl* deletion mutant of *E. coli* K12, and strain BGH2001 is a *celR*$^+$ mutant of JF201 (HALL, BETTS and KRICKER 1986).

Isolation of β-glucoside$^+$ mutants: Each strain from the ECOR collection (OCHMAN and SELANDER 1984) was streaked onto MacConkey cellobiose, MacConkey salicin and MacConkey arbutin medium. All colonies were white, indicative of failure to ferment the included sugar, on these media. Plates were sealed to prevent dehydration and incubated at 30°. At intervals the plates were inspected for the presence of papillae (outgrowths) on the surface of the colonies. When papillae were detected, cells from papillae from up to five different colonies were restreaked onto the same kind of MacConkey plate from which they were obtained. A red (sugar fermenting) colony from each papilla was tested for utilization of all three β-glucoside sugars and for constitutive hydrolysis of *p*-nitrophenyl-β-glucoside (PNPG). Multiple mutants from any given ECOR strain were retained only if they exhibited different phenotypes by these tests. In two cases two mutants with identical phenotypes (as listed in Table 1) were retained because the intensities of sugar fermentation, as judged by colony color on MacConkey medium, differed between the two mutants.

Tests for sugar utilization: Mutants were streaked onto MacConkey plates containing each of the three β-glucoside sugars, and onto minimal medium in which each of the three sugars was the sole carbon source. The mutants were judged to have a positive phenotype only if they formed colonies on the minimal medium within three days. Intensities of the positive phenotypes could often be distinguished by the color of the colonies on MacConkey media.

Test for constitutive PNPG hydrolysis: Colonies grown on glucose minimal medium were overlaid with a solution of 2.5 mM PNPG in minimal medium without a carbon source. Constitutive colonies turn yellow due to the release of *p*-nitrophenol within 10–15 min. Inducible or negative colonies remain white (KRICKER and HALL 1984).

Preparation of RNA: Cultures (50 ml) growing exponentially in the indicated medium, and at a density of approximately 10^8 cells/ml, were harvested by centrifugation, washed once, resuspended in 750 μl of lysis buffer (50 mM glucose, 25 mM Tris-HCl, 10 mM EDTA, pH 8.0), and lysed by the addition of an equal volume of phenol at 60°. The aqueous phase was separated by centrifugation and sequentially extracted with room temperature phenol, phenol: $CHCl_3$ (1:1 ratio), and by $CHCl_3$. The resulting aqueous solution was mixed with 0.5 volume of ice cold 7.5 M ammonium acetate, held 20 min on ice, and centrifuged in a microcentrifuge for 15 min in the cold. The resulting pellet was resuspended in 50 μl of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and the RNA concentration was determined from the absorbance at 260 nm.

Preparation of probe DNA: Probes were prepared from the 4.2-kb HpaI fragment of plasmid pUF564 (HALL, BETTS and KRICKER 1986), which carries the *cel* genes, and from the 3.5-kb *Hin*dIII fragment of plasmid pAR6 (REYNOLDS *et al.* 1985), which carries the *bgl* genes of *E. coli* K12. Plasmids were digested with the appropriate enzyme, subjected to electrophoresis in low melting point agarose gels, and the desired fragments were cut excised from the gels. The excised fragments were labeled by second strand synthesis using ^{32}P-TTP and ^{32}P-dCTP as described by (FEINBERG and VOGELSTEIN 1984).

Preparation of genomic DNA: Ten-milliliter overnight cultures were resuspended in 2 ml 40 mM Tris-50 mM EDTA (pH 8.0) containing 10 mg/ml lysozyme, held 30 min at room temperature, and adjusted to 1% in sodium dodecyl sulfate. The suspensions were adjusted to >1000 units of T1 RNAase (Bio-Rad), held for 5 min at 55°, then extracted twice with phenol and once with chloroform. DNA was precipitated with cold ethanol and spooled onto a glass rod.

Hybridizations: Serial dilutions of known amounts of RNA were applied to Gene Screen Plus (New England Nuclear) according to the instructions of the manufacturer, using a dot blot apparatus. Hybridizations were carried out in 1 M NaCl containing 1% sodium dodecyl sulfate and 100 μg/ml of heat denatured salmon sperm DNA at 60° using approximately 5×10^6 cpm of probe per milliliter of hybridization solution. Filters were washed sequentially in 2 × SSC at room temperature, in 2 × SSC containing 1% SDS at 60°, then subjected to autoradiography at room temperature.

DNA hybridizations of Southern transfers were performed identically except that hybridizations and hot washes were performed at 65°.

RESULTS

Isolation of β-glucoside utilizing mutants: The ECOR collection is a reference collection of *E. coli* consisting of 72 strains that are representative of genotypic diversity as determined by multilocus enzyme electrophoresis (OCHMAN and SELANDER 1984). Although a majority (41) of these isolates were obtained from human hosts, the remainder were drawn from a remarkable diversity of animal hosts; and most of these strains were isolated within the last 12 yr.

We began by determining how many of these isolates could grow on any of the three β-glucoside sugars. Each strain was streaked onto MacConkey cellobiose, MacConkey salicin, and MacConkey arbutin media, and incubated at 30°. MacConkey is an indicator medium with the property that colonies are red or pink if they can ferment the sugar that is present, whereas those that can not utilize the sugar are white. Based upon the appearance of the colonies, no strain from this collection was capable of utilizing any of these sugars. The inability to utilize the sugars was confirmed by the inability of these strains to grow on mineral salts medium in which these sugars were the sole carbon source.

The plates were sealed to prevent dehydration, and incubation was continued for up to 8 weeks. Papillae (outgrowths on the surface of the colonies) began to appear on some colonies after about 1 week. Cells from these papillae were streaked onto arbutin, salicin, and cellobiose MacConkey plates to determine the phenotypes of these spontaneous mutants (see MATERIALS AND METHODS for details of the screening procedure). Mutants with a variety of phenotypes were selected by this procedure, and when more than one mutant phenotype was isolated from a given strain, representatives of each phenotype were purified and retained.

Ninety-eight spontaneous β-glucoside utilizing mutants, designated NIM (natural isolate mutant) strains,

TABLE 1

***β*-Glucoside utilizing mutants of natural isolate strains**

NIM strain[a]	Phenotype[b]	Class[c]	Selected on	ECOR parent strain	NIM strain[a]	Phenotype[b]	Class[c]	Selected on	ECOR parent strain
101	AS	1	Cello	1	483	ACS	2	Arb	48
102	A	1	Arb	1	491	ACS	2	Cello	49
111	AS	1	Sal	11	492	CS	2	Sal	49
121	ASP	1	Arb	12	501	AS	1	Arb	5
122	ACSP	1	Sal	12	511	CS	3	Cello	51
131	ASP	1	Sal	13	512	A	3	Arb	51
132	AS	1	Sal	13	521	A	3	Sal	52
141	AS	1	Arb	14	531	AS	3	Arb	53
151	AP	1	Arb	15	532	ASP	3	Sal	53
152	AS	1	Sal	15	541	C	3	Cello	54
161	AS	1	Sal	16	542	AS	3	Arb	54
162	ASP	1	Arb	16	551	CS	3	Cello	55
181	ASP	1	Arb	18	552	AS	3	Arb	55
191	ACS	1	Arb	19	561	C	3	Cello	56
192	AS	1	Sal	19	562	AS	3	Arb	56
201	AS	1	Arb	2	571	AS	3	Arb	57
211	AS	1	Sal	21	572	S	3	Sal	57
212	AC	1	Cello	21	581	CS	3	Sal	58
221	ASP	1	Sal	22	582	ASP	3	Sal	58
231	ACS	1	Arb	23	591	C	3	Cello	59
241	C	1	Cello	24	592	AS	3	Arb	59
242	ACS	1	Cello	24	601	AC	1	Cello	6
251	AS	2	Sal	25	611	C	3	Cello	61
261	ACSP	2	Cello	26	621	ACS	3	Cello	62
271	ACS	2	Cello	27	622	C	3	Cello	62
281	AS	2	Arb	28	631	ACS	3	Cello	63
301	AS	1	Sal	3	641	C	3	Cello	64
302	S	1	Sal	3	642	AS	3	Arb	64
311	AP	2	Arb	31	651	ACSP	3	Sal	65
312	AS	2	Sal	31	661	C	3	Cello	66
321	ACS	2	Sal	32	662	AS	3	Sal	66
322	CS	2	Sal	32	671	AS	3	Arb	67
331	CS	2	Sal	33	681	ACSP	3	Arb	68
332	AS	2	Arb	33	691	ACSP	3	Arb	69
341	CS	2	Sal	34	701	ACSP	1	Cello	7
351	AC	2	Cello	35	711	CS	3	Sal	71
352	S	2	Sal	35	721	ACS	3	Arb	72
361	AS	2	Sal	36	801	AS	1	Arb	8
371	ACSP	2	Cello	37	901	AP	1	Arb	9
381	S	2	Sal	38	1001	A	1	Arb	10
401	AS	1	Cello	4	1002	AS	1	Cello	10
421	ACS	2	Cello	42	1003	ACS	1	Sal	10
422	ASP	2	Cello	42	3001	C	2	Cello	30
441	ACSP	2	Cello	44	3002	ACSP	2	Arb	30
451	ACSP	2	Cello	45	3003	ACSP	2	Arb	30
461	ACSP	2	Cello	46	4001	ACS	2	Sal	40
471	C	2	Cello	47	5001	ACS	2	Arb	50
481	C	2	Cello	48	6001	ACS	3	Cello	60
482	ACS	2	Sal	48	7001	C	3	Sal	70

[a] NIM strains, column one, are the natural isolate mutant strains derived from the ECOR strains in column five.
[b] Phenotypes: A = arbutin$^+$, S = salicin$^+$, C = cellobiose$^+$, P = consitutive hydrolysis of PNPG.
[c] Class is as described by OCHMAN and SELANDER (1984) based on factor analysis.

were obtained in that experiment (Table 1). Only five of the 71 ECOR strains failed to yield any *β*-Glu$^+$ mutants [ECOR 43 was shown to be identical to ECOR 10 and probably represents an error when the collection was transferred from the Selander laboratory (D. DYKHUIZEN, personal communication)]. These five were ECOR strains 17, 20, 29, 39 and 41. Forty-five strains yielded cellobiose$^+$ mutants, 62 strains yielded arbutin$^+$ mutants, and 58 strains yielded salicin$^+$ mutants. Twenty-eight strains yielded mutants with the *E. coli* K12 *bglR*$^+$ phenotype (arbutin and salicin positive, cellobiose negative, and inducible for the hydrolysis of PNPG {PNPG}), while 12 strains yielded mutants with the *E. coli* K12 *celR*$^+$ phenotype (arbutin, salicin and cellobiose positive and constitutive for the hydrolysis of PNPG). The strains in the

TABLE 2

β-Glucoside$^+$ mutants according to ECOR class

Phenotype[b]	ECOR class[a] 1	2	3	Total
AS	14	5	9	28
A	2	0	2	4
ASP	5	1	2	8
ASCP	2	7	3	12
AP	2	1	0	3
ACS	4	8	4	16
AC	2	1	0	3
C	1	3	8	12
S	1	2	1	4
CS	0	4	4	8
Totals	33	32	33	98

[a] Class of ECOR parent.
[b] Phenotypes: A = arbutin$^+$, S = salicin$^+$, C = cellobiose$^+$, P = consitutive hydrolysis of PNPG.

TABLE 3

β-Glucoside$^+$ mutants of natural isolates according to β-glucoside sugar used for selection

Phenotype[a]	Cellobiose[b]	Arbutin[b]	Salicin[b]	Total
AS	3	14	11	28
A	0	3	1	4
ASP	1	3	4	8
ACSP	6	4	2	12
AP	0	3	0	3
ACS	7	5	4	16
AC	3	0	0	3
C	11	0	1	12
S	0	0	4	4
CS	2	0	6	8
Total	33	32	33	98

[a] Phenotypes: A = arbutin$^+$, S = salicin$^+$, C = cellobiose$^+$, P = consitutive hydrolysis of PNPG.
[b] β-Glucoside sugar used to select the mutants.

ECOR collection are divided into three classes based upon factor analysis of the electrophoretic mobilities of the 11 enzymes assayed (OCHMAN and SELANDER 1984). We asked whether these classes differed in terms of their potential for yielding various mutant phenotypes. Table 2 shows the variety of β-Glu$^+$ mutants obtained arranged according to the class of the parent from which they were obtained. The three classes gave rise to nearly exactly equal numbers of mutants. For only one phenotype (cello$^+$, arb$^-$, sal$^-$) was the distribution among the three classes significantly different from random by a χ^2 test.

Because constitutive hydrolysis of PNPG was considered as part of the phenotype, there are a possible 15 new phenotypes. Since any arbutin$^-$, salicin$^-$, cellobiose$^-$, PNPG constitutive mutants would not have been detected in this screening, there were actually only 14 new phenotypes that could have been obtained. Of those 14 possible phenotypes, 10 were obtained. No arbutin negative PNPG constitutive mutants were obtained, nor were any arbutin positive, cellobiose positive, salicin negative, PNPG constitutive mutants obtained.

Relationship of selective pressure to phenotypes obtained: Papillae are presumed to be the result of spontaneous mutations that permit utilization of the sugar included in the MacConkey plate upon which they arose. The phenotypes of the mutants are thus expected to be strongly correlated with the sugar employed in the selection. This is, in general the case. It is not surprising to find (Table 3) that selection on salicin or arbutin yields mutants that also utilize cellobiose. We know from the *E. coli* K12 system that the proteins of the *cel* operon can act on arbutin and salicin, as well as on cellobiose. All of the mutants selected on arbutin, for instance, utilized arbutin. In contrast, however, four mutants selected on cellobiose utilized arbutin and salicin, but not cellobiose. Similarly two mutants selected on salicin failed to utilize salicin. The most likely explanation for these observations is that the proteins in those mutants have a low level of activity toward the selective sugar, but that level is below the threshold for growth. On the MacConkey plate, where other growth resources are present, a slight activity toward the sugar present might confer a sufficient selective advantage to produce a papilla.

Potential for different β-Glu$^+$ phenotypes from the same strain: Thirty-seven of the 66 strains that produced β-Glu$^+$ mutants yielded only a single phenotype, while the remaining 29 strains produced more than one kind of β-Glu$^+$ mutant. Strain ECOR 24, for instance, yielded a cellobiose$^+$ mutant and a cello$^+$-arb$^+$-sal$^+$ mutant (Table 1). Similar mutants have been obtained in *E. coli* K12 (KRICKER and HALL 1984), and are now believed to be the result of different levels of expression of the *cel* operon (KRICKER and HALL 1987). It is therefore not clear that strain ECOR 24 retains the potential to reactivate both the *cel* and *bgl* operons. Strain ECOR 10, on the other hand, appears to be able to activate genes equivalent to all three of the β-Glucoside utilization operons that have been identified in *E. coli* K12. ECOR 10 produced an arbutin$^+$ mutant, consistent with activation of the *arbT* locus; an arb$^+$ sal$^+$ mutant, consistent with activation of the *bgl* operon, and an arb$^+$ sal$^+$ cello$^+$ mutant, consistent with activation of the *cel* operon. There are several cases in which it seems likely that both the *cel* and the *bgl* operons have been retained in a cryptic state.

Five strains (ECOR 17, 20, 29, 39 and 41) appear to have lost all potential β-glucoside utilization functions.

Several strains yielded only cellobiose utilizing mutants, and these seem likely to have lost both the *arbT* gene and the *bgl* operon. Those strains that yielded

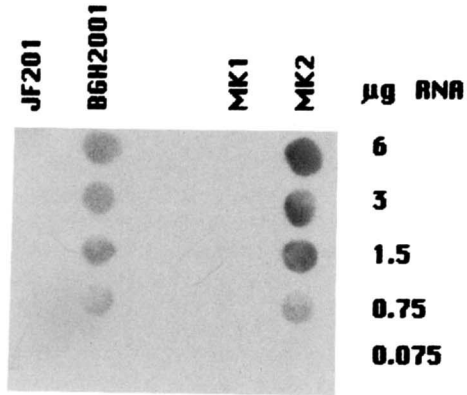

FIGURE 1.—RNA dot blots of *E. coli* K12 strains. Serial dilutions of total RNA from glycerol grown cultures of the indicated strains were probed with *cel* specific DNA. Numbers indicate the total amount of RNA applied to a spot.

only arbutin-utilizing mutants are likely to have lost both *cel* and *bgl* operons, but to have retained *arbT*, while those that yielded only arbutin and salicin-utilizing mutants have probably lost the *cel* operon.

Molecular identification of the β-glucoside utilization genes: The above identifications of the operons being expressed in these mutants are based upon the assumption that the genes present in these natural isolates are quite similar to those we have identified in *E. coli* K12. These assignments must thus be considered quite tentative. It is conceivable that in some natural isolates the proteins specified by the cryptic *bgl* operon are active toward cellobiose, while in others those specified by the cryptic *cel* operon have lost activity toward cellobiose. It is also possible that some of the strains carry cryptic alleles for β-glucoside utilization, other than *arbT*, that are unrelated to either *bgl* or *cel*.

To determine whether the known operons were being expressed in some of the β-glu$^+$ mutants of the ECOR strains, RNA was isolated from seven of the mutants and hybridized to molecular probes specific for the two operons (hybridization under our conditions reveals no homology between the *bgl* and the *cel* operons).

Figure 1 shows that no *cel* specific RNA is produced by the *celR*^{o} *E. coli* K12 strains MK1 and JF201, but that *cel* RNA is produced constitutively by the *celR*$^+$ strains MK2 and BGH2001.

The seven mutant strains were chosen for convenience because they all utilized arbutin and all hydrolyzed PNPG constitutively. Figure 2 shows the hybridization of *cel* and *bgl* ^{32}P-labeled DNA probes to RNA extracted from the mutant strains grown under various conditions. The results from Figure 2 are summarized in Table 4.

The *E. coli* K12 strain MK2 serves as a control in this experiment. MK2 expresses the *cel* operon constitutively and expresses the *bgl* operon when induced by growth on arbutin.

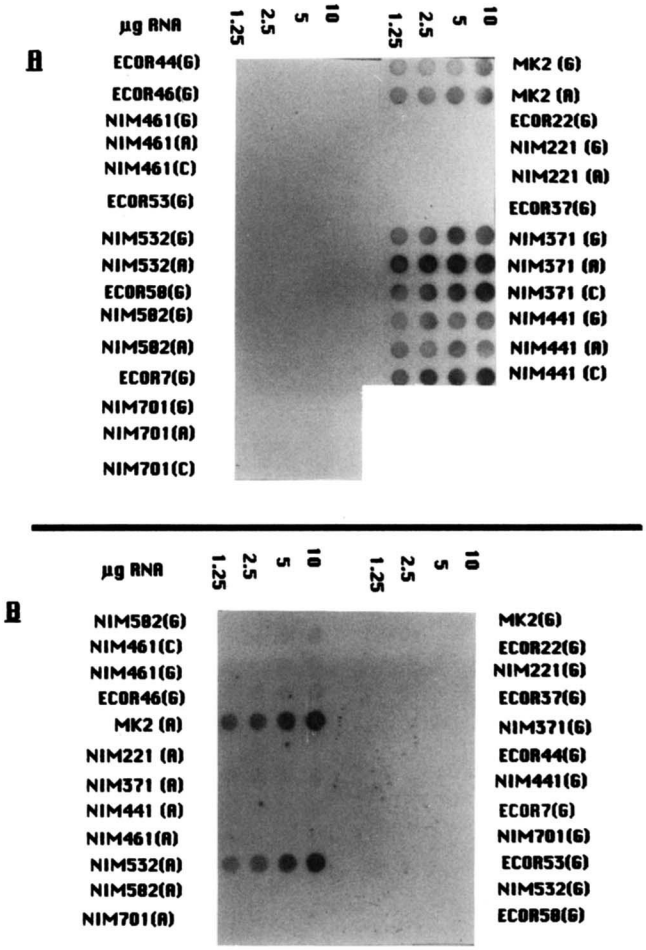

FIGURE 2.—RNA dot blots of natural isolates. Total RNA was probed with either *cel* (panel A) or *bgl* (panel B) specific DNA. Cultures were grown on the sugar indicated in parentheses following the strain name. G, glycerol; A, arbutin; C, cellobiose.

TABLE 4

Regulation of β-glucoside utilization genes in β-glucoside$^+$ mutants

Strain	Phenotype[a]	*bgl* Operon	*cel* Operon
MK2	$A^+S^+C^+P^c$	Inducible by arbutin	Constitutive
NIM 221	$A^+S^+C^-P^c$	Not expressed	Not expressed
NIM 371	$A^+S^+C^+P^c$	Not expressed	Constitutive
NIM 441	$A^+S^+C^+P^c$	Not expressed	Constitutive
NIM 461	$A^+S^+C^+P^c$	Not expressed	Not expressed
NIM 532	$A^+S^+C^-P^c$	Inducible by arbutin	Not expressed
NIM 582	$A^+S^+C^-P^c$	Not expressed	Not expressed
NIM 701	$A^+S^+C^+P^c$	Not expressed	Not expressed

[a] A, arbutin utilization; S, salicin utilization; C, cellobiose utilization; P^c, constitutive hydrolysis of PNPG by intact cells.

None of the parental ECOR strains expressed either operon constitutively.

The cellobiose positive strains NIM 371 and NIM 441 expressed the *cel* operon constitutively, but failed to express the *bgl* operon even when grown on cellobiose or arbutin. These strains appear to be equivalent to *celR*$^+$ mutants of *E. coli* K12.

Two other cellobiose positive strains, NIM 701 and

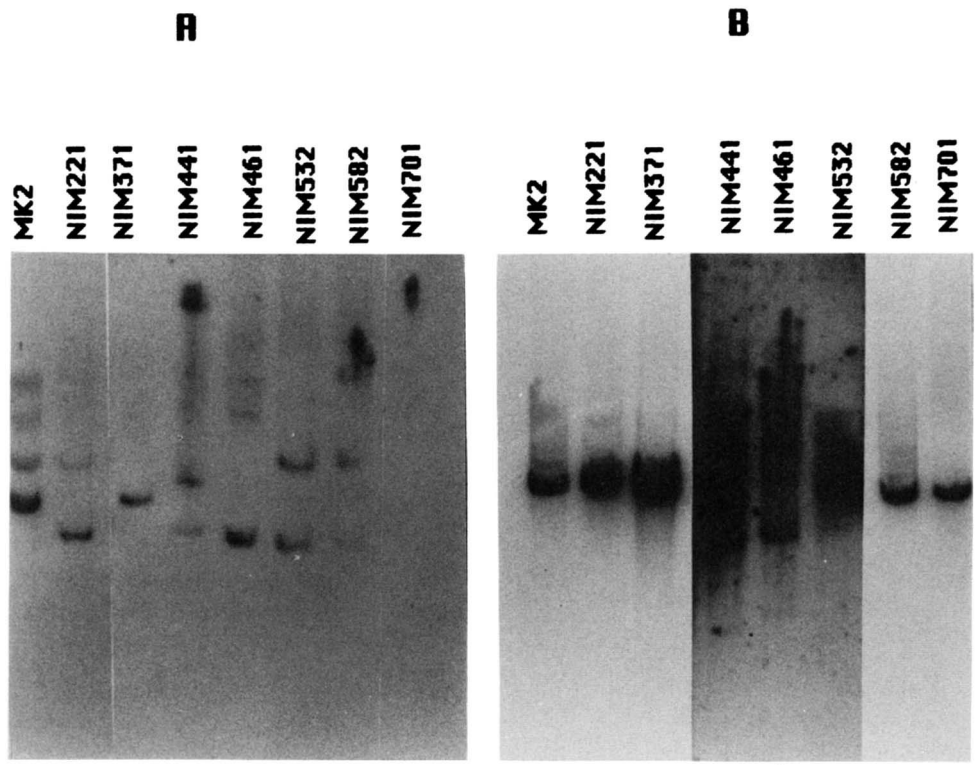

FIGURE 3.—Southern blots of DNA from natural isolate mutants. Genomic DNA was digested with restriction endonuclease EcoRI, subjected to agarose gel electrophoresis, and transfered to Gene Screen Plus blotting medium. Panel A digests were probed with *cel* DNA. Panel B digests were probed with *bgl* DNA. Lanes 4, 5 and 6 in panel B were overexposed to permit easier visualization of the bands.

NIM 461, on the other hand, did not synthesize either *cel* specific or *bgl* specific mRNA at detectable levels under any of the growth conditions tested.

The arbutin, salicin positive, cellobiose negative strain NIM 532 synthesized *bgl* specific mRNA, but only when induced by growth on arbutin. NIM 532 thus appears to be equivalent to a $bglR^+$ mutant of *E. coli* K12. The constitutive hydrolysis of PNPG by this strain is not explained by the inducible $bglR^+$ mutation. NIM221 and NIM582 failed to express either the *cel* or *bgl* operons under any conditions.

The failure to detect *cel* mRNA in cellobiose positive strains, or *bgl* mRNA in two of the three arbutin$^+$, salicin$^+$ strains could possibly result from insufficient homology between the K12 alleles and the alleles present in these strains. To determine whether our probes would hybridize to sequences from these isolates, genomic DNA from each natural isolate mutant was digested with restriction endonuclease EcoRI, the fragments were separated by agarose gel electrophoresis, and transferred to Gene Screen Plus membrane by SOUTHERN (1975) blotting. The genomic blots were hybridized to the same probes as were the RNA dot blots. Figure 3 shows that the *bgl* probe hybridized to sequences present in all of these natural isolates. The *cel* probe failed to hybridize to genomic DNA from strain NIM701. Although this could reflect insufficient homology between the *cel* sequences of *E. coli* K12 and strain NIM701, the observation that the K12 derived *bgl* probe hybridizes strongly to the NIM701 DNA suggests that there has not been such extensive divergence. The most likely interpretation is that the *cel* gene is deleted in strain NIM701. It is therefore reasonable to conclude that strains NIM221, NIM461, NIM582 and NIM701 are not capable of expressing either the *bgl* or the *cel* operon.

CONCLUSIONS

From this survey it is clear that the genes for β-glucoside utilization are generally cryptic in *E. coli*, thus cryptic β-glucoside genes are not simply a laboratory artifact. In only 7% of the strains (5 out of 71) have all of the β-glucoside utilization genes been inactivated. The pattern of phenotypes produced by these strains suggests that in many natural *E. coli* isolates one or more of the β-glucoside utilization genes have been inactivated. This is consistent with the expectation that silent genes will eventually accumulate inactivating mutations. Because of the overlapping specificities of the products of the *bgl* and *cel* operons it is difficult to judge from phenotypes alone which operons have been retained. The limited molecular data presented here emphasizes the danger of assigning genotypes to natural isolate mutants based strictly upon their phenotypes. Additional molecular data will be required to determine what operons are

being expressed in some cases. The observation that two cellobiose utilizing mutants failed to express the *cel* operon strongly suggests that there is at least one additional cryptic operon for cellobiose utilization present in those strains. Similarly, the observation that two arbutin$^+$ salicin$^+$ cellobiose$^-$ strains failed to express either the *cel* or *bgl* operons indicates the existence of still another β-glucoside utilization operon. The existence of another set of genes for transport and hydrolysis of β-glucosides is supported by the observation that strain NIM532 expressed the *bgl* operon *inducibly*, but that it also transports and hydrolyzes the synthetic substrate PNPG *constitutively*. The molecular data must, however, be interpreted with caution. The observation that two of these mutants failed to express the *bgl* operon at a detectable level should not, for instance, be taken as implying that the *bgl* operon has been irreversibly inactivated in those strains.

Until these other genes for β-glucoside utilization are isolated and characterized we can not be absolutely sure that these mutants did not arise from changes in the substrate specificities of other normally expressed enzymes. However, from what is known about the biochemistry and enzymology of β-glucoside utilization in other members of the Enterobacteriaceae, this possibility appears remote.

If we confine our attention to cellobiose utilization, 45 of the 71 ECOR strains retained some cellobiose utilization genes in a potentially functional, but silent, state. In some cases those genes are closely related to the *cel* operon of *E. coli* K12.

These data, taken together with the earlier studies in *E. coli* K12, suggest that the *E. coli* population contains at least four sets of genes for catabolism of β-glucosides, all of which are normally silent. Those four sets are the *bgl* operon, the *cel* operon, and the loci responsible for β-glucoside catabolism in the NIM mutants that do and do not, respectively, catabolize cellobiose. The observation that 7% of the isolates from the ECOR collection were unable to reactivate any of those genes indicates that, as expected, these silent genes are subject to random mutational inactivation. If these four gene clusters are actually present in each strain, and inactivating mutations are randomly distributed among those clusters, then each cluster has roughly a 50% chance of having been inactivated in any random *E. coli* isolate.

This study supports the concept of cryptic genes in general, and these results are entirely consistent with the models for retention of cryptic genes. The major unexpected finding of this study is that there are so many cryptic genes for β-glucoside utilization.

Why should *E. coli* contain such a large set of apparently redundant genes for catabolism of β-glucosides, and why should those genes be silent? This study has provided us with a large set of β-glucoside positive mutants of natural *E. coli* isolates. These mutants are being used for both physiological and molecular studies in attempt to understand the forces that maintain these genes in a silent state.

This study was supported by National Institutes of Health grant AI14766. B. G. H. was the recipient of N.I.H. Research Career Development Award 1K04 AI00366. We are grateful to A. WRIGHT for the gift of plasmid pAR6, and to M. KRICKER for participation in the early portion of this study. We are indebted to ROSIE REDFIELD for teaching us the quick and easy method for genomic DNA preparation.

LITERATURE CITED

BACHMANN, B. J., 1983 Linkage map of *Escherichia coli* K12, edition 6. Microbiol. Rev. **44:** 1–56.

FEINBERG, A. P. and B. VOGELSTEIN, 1984 A technique for radiolabeling DNA restriction endonuclease fragments to a high specific activity. Anal. Biochem. **137:** 266–267.

HALL, B. G., 1982 Chromosomal mutation for citrate utilization in *Escherichia coli* K12. J. Bacteriol. **151:** 269–273.

HALL, B. G. and D. L. HARTL, 1974 Regulation of newly evolved enzymes. I. Selection of a novel lactase regulated by lactose in *Escherichia coli*. Genetics **89:** 453–465.

HALL, B. G., P. W. BETTS and M. KRICKER, 1986 Maintenance of the cellobiose utilization genes of *Escherichia coli* in a cryptic state. Mol. Biol. Evol. **3:** 389–402.

HALL, B. G., S. YOKOYAMA and D. H. CALHOUN, 1983 Role of cryptic genes in microbial evolution. Mol. Biol. Evol. **1:** 109–124.

KRICKER, M. and B. G. HALL, 1984 Directed evolution of cellobiose utilization in *Escherichia coli* K12. Mol. Biol. Evol. **1:** 171–182.

KRICKER, M. and B. G. HALL, 1987 Biochemical genetics of the cryptic gene system for cellobiose utilization in *Escherichia coli* K12. Genetics **115:** 419–429.

LI, W-H., 1984 Retention of cryptic genes in microbial populations. Mol. Biol. Evol. **1:** 213–219.

MORISHITA, T., T. FUKADA, M. SHIROTA and T. YURA, 1974 Genetic basis of nutritional requirements in *Lactobacillus casei*. J. Bacteriol. **120:** 1078–1084.

OCHMAN, H. and R. K. SELANDER, 1984 Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:** 690–693.

PAQUIN, C. E. and V. M. WILLIAMSON, 1986 Ty insertions at two loci account for most of the spontaneous antimycin A resistance mutations during growth at 15°C of *Saccharomyces cerevisiae* strains lacking ADH1. Mol. Cell Biol. **6:** 70–79.

PRASAD, I. and S. SCHAEFLER, 1974 Regulation of the β-glucoside system in *Escherichia coli* K12. J. Bacteriol. **120:** 638–650.

REYNOLDS, A. E., J. FELTON and A. WRIGHT, 1981 Insertion of DNA activates the cryptic *bgl* operon in *E. coli* K12. Nature **293:** 625–629.

REYNOLDS, A. E., S. MAHADEVAN, J. FELTON and A. WRIGHT, 1985 Activation of the cryptic *bgl* operon: insertion sequences, point mutation and changes in superhelicity affect promoter strength. UCLA Symp. Mol. Cell. Biol. **20:** 265–277.

SCHAEFLER, S. and W. MAAS, 1967 Inducible system for the utilization of β-glucosides in *Escherichia coli*. II. Description of mutant types and genetic analysis. J. Bacteriol. **93:** 264–272.

SCHAEFLER, S. and A. MALAMY, 1969 Taxonomic investigations on expressed and cryptic phospho-β-glucosidases in Enterobacteriaceae. J. Bacteriol. **99:** 422–433.

SCHAEFLER, S. and L. MINTZER, 1959 Acquisition of lactose fermenting properties by Salmonellae. I. Interrelationship be-

tween the fermentation of cellobiose and lactose. J. Bacteriol. **78:** 159–163.

SCHAEFLER, S. and I. SCHENKEIN, 1968 β-Glucoside permeases and phospho-β-glucosidases in *Aerobacter aerogenes*: relationship with cryptic phospho-β-glucosidases in Enterobacteriaceae. Proc. Natl. Acad. Sci. USA **59:** 285–292.

SOUTHERN, E., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517.

Communicating editor: W.-H. LI