Mechanisms of Activation of the Cryptic *cel* Operon of *Escherichia coli* K12

Laura L. Parker[1] and Barry G. Hall[2]

Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut 06268

Manuscript received April 6, 1989
Accepted for publication November 27, 1989

ABSTRACT

The *cel* (cellobiose utilization) operon of *Escherichia coli* K12 is not expressed in the wild-type organism. However, mutants that can express the operon and thereby utilize the β-glucoside sugars cellobiose, arbutin and salicin are easily isolated. Two kinds of mutations are capable of activating the operon. The first involves mutations that allow the repressor to recognize the substrates cellobiose, arbutin and salicin as inducers. We have identified the sequence changes in five different active alleles and found those differences to be single base pair changes at one of two lysine codons in the repressor gene. The second kind of mutation involves the integration of the insertion sequences IS*1*, IS*2* or IS*5* into a 108-bp region 72–180 bp upstream of the start of transcription. Integration occurs at several different sites and in different orientations. Transcription of the *cel* operon begins at the same base pair in all mutants examined. Of 44 independent *cel*$^+$ mutants, 27 were activated by point mutations and 17 were activated by insertion sequences. The preferred mechanism of activation appears to be strain dependent, since one of the parents yielded 94% insertionally activated alleles, while another yielded 100% point mutation activated alleles.

CRYPTIC genes are silent genes that are not able to be expressed in the wild-type organism. *Escherichia coli* K12 possess a variety of genes for β-glucoside utilization, but most of these genes are cryptic in the wild-type organism. Mutations at several different loci that result in the expression of these cryptic genes have been investigated. The best characterized system is the *bgl* operon, a system that specifies all of the functions necessary for fermentation of the aryl-β-glucoside sugars, arbutin and salicin (PRASAD and SCHAEFLER 1974). That operon consists of three structural genes, *bglG*, *bglF* and *bglB* (SCHNETZ, TOLOCZYKI and RAK 1987; PRASAD and SCHAEFLER 1974). The regulatory region of the *bgl* operon contains a 130-bp leader region where termination of transcription occurs in the absence of arbutin and salicin (MAHADEVAN, REYNOLDS and WRIGHT 1987; SCHNETZ and RAK 1988). In the presence of arbutin or salicin the *bglG* gene product acts as a positive regulatory element and prevents termination of transcription. The second gene of the operon, *bglF*, encodes the β-glucoside specific phosphotransferase protein that both transports and phosphorylates the substrates. In the absence of arbutin and salicin, the product of *bglF* interacts with the product of *bglG* to prevent antitermination and thereby represses the operon (MAHADEVAN, REYNOLDS and WRIGHT 1987). The third gene, *bglB*, encodes a phospho-β-glucosidase that is specific for aryl-β-glucosides.

Even if arbutin and salicin are present the wild-type *bgl* operon does not permit fermentation of these sugars because specific changes in the promoter region are required to overcome the cryptic state of the system. It has been shown that insertions of IS*1* or IS*5* into a 47-bp region upstream of a nearby promoter are capable of activating the *bgl* operon (REYNOLDS, FELTON and WRIGHT 1981; REYNOLDS *et al.* 1985, 1986). It is known that the insertions increase the activity of this promoter (SCHNETZ and RAK 1988), however, the exact mechanism by which this happens is not known. It has been shown that ethyl methansulfonate induced single base changes in the cyclic-AMP receptor protein binding site, and particular *gyrA* and *gyrB* mutations, are also capable of activating the *bgl* operon (REYNOLDS *et al.* 1985).

The cryptic *cel* operon has also been studied extensively (KRICKER and HALL 1984, 1987; HALL, BETTS and KRICKER 1986; Parker and HALL 1990). Expression of the *cel* operon allows the utilization of cellobiose, salicin, and arbutin. The operon consists of five genes, *celABCDF* (PARKER and HALL 1990), in which *celB* and *celC* encode gene products that are required for phosphoenolpyruvate-dependent transport and phosphorylation of cellobiose, arbutin and salicin, *celD* encodes a repressor, and *celF* encodes a phospho-β-glucosidase that acts on cellobiose, arbutin and salicin. The function of the *celA* gene product is unknown. In this report we show that activation of the *cel* operon can be accomplished two ways. One mechanism of

[1] Present address: Department of Physiology, Tufts University Medical School, Boston, Massachusetts 02111.
[2] Present address: Department of Biology, University of Rochester, Rochester, New York 14627.

The publication costs of this article were partly defrayed by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

activation involves mutations in *celD* that allow the repressor to recognize cellobiose, arbutin, and salicin as inducers. Another mechanism of activation involves the integration of the insertion sequences IS*1*, IS*2* or IS*5* in a 108-bp region 72–180 bp upstream of the transcriptional initiation site. The insertions increase the basal level of gene expression, suggesting that these elements interfere with binding of the repressor to the operator.

MATERIALS AND METHODS

Culture media and conditions: Minimal media consisted of phosphate-buffered mineral salts solution (HALL and BETTS 1987) with 1.5% agar as a solidifying agent. Sugars were used 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 200 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 not, colonies were white.

Strains and plasmids: *E. coli* strains and plasmids are listed in Table 1.

Transductions: Transductions were mediated by phage P1 cm clr (ROSNER 1972) according to MILLER (1972).

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-kb *Hpa*I fragment of plasmid pUF564 (HALL, BETTS and KRICKER 1986). The IS*2* specific probe was prepared from the 700-bp *Hpa*I/*Hin*dIII fragment of plasmid pBRK10 (KLAER *et al.* 1981), the IS*5* probe from the 3-kb *Pst*I fragment of plasmid pLG2, and the IS*1* probe from the 1.6-kb *Pst*I/*Eco*RI fragment of plasmid pBRG36 (BIEL, ADELT and BERG 1984). Preparation of competent cells and transformations with plasmid DNA were performed according to methods given by MANIATIS, FRITSCH and SAMBROOK (1982).

Isolation of RNA: RNA was isolated as previously described (PARKER and HALL 1990).

***In vivo* cloning:** The mini-Mu derivative Mud5005 was used as described in GROISMAN and CASADABAN (1986). The helper phage employed was Mu cts62 pAp5 which carries an ampicillin resistance determinant (AKROYD *et al.* 1984).

Other cloning vectors: The plasmid pBlu$^+$ (Stratagene, Inc., San Diego, California) is a 3.0-kb high copy number vector that carries a multiple cloning site cartridge and a gene for ampicillin resistance. The plasmid pHSG415 (HASHIMOTO-GOTOH *et al.* 1981) is a low copy number plasmid that confers resistance to ampicillin, kanamycin and chloramphenicol.

Phospho-β-glucosidase assays: Phospho-β-glucosidase activity was determined from the *in vivo* hydrolysis of *p*-nitrophenyl-β-glucopyranoside (PNPG) as previously described (PARKER and HALL 1988).

Amplification of DNA fragments for sequencing: Amplification of fragments for sequencing was performed as previously described (PARKER and HALL 1990).

S1 transcript mapping: Transcriptional start sites were determined as described by BERK and SHARP (1977), with modifications as described by PARKER and HALL (1990).

DNA sequencing: DNA was sequenced directly from double-stranded plasmids, or from single-stranded DNA generated through polymerase chain reaction (PCR) amplification, using the dideoxy chain termination method previously described (SANGER, NICKLEN and COULSON 1977) except that instead of alkaline denaturation of DNA fragments, samples were boiled for 5 min, frozen in a Dry Ice/ethanol bath, and allowed to thaw at room temperature.

RESULTS

The cryptic *cel* operon of *E. coli* K12 can be activated by insertion sequences and point mutations. We initially examined the structure of the *cel* operon of the parent strain MK1 and its Cel$^+$ mutant, strain MK2 by restriction mapping. Genomic DNA from each strain was digested simultaneously with the enzymes *Eco*RI and *Cla*I, separated by agarose gel electrophoresis, and transferred to a nylon membrane. The transferred DNA was hybridized with [^{35}S]dATP-labeled 4.0-kb *Hpa*I fragment of pUF564 (HALL, BETTS and KRICKER 1986) that contained the entire *cel* operon. It was found that the 2.3-kb *Eco*RI/*Cla*I fragment which lies at the 5′ end of the operon was 1.2 kb larger in strain MK2 than in strain MK1, suggesting that foreign DNA had become integrated into this region. Plasmid pUF650 (PARKER and HALL 1990), which carries the entire MK2 *cel* operon, was similarly digested, transferred, and hybridized to appropriate [^{35}S]dATP-labeled fragments derived from plasmids containing the insertion sequences IS*1*, IS*2* or IS*5*. Only the IS*2* probe hybridized to the 3.5-kb *Eco*RI/*Cla*I fragment of plasmid pUF650, indicating that the *cel* allele of strain MK2 was probably activated by IS*2*. Detailed restriction enzyme analysis allowed us to localize the insertion to the 800-bp *Hin*dIII/*Hpa*I fragment that contains *celR* and *celA* (Figure 1), and indicated that IS*2* had inserted in orientation I (GHOSAL, SOMMER and SAEDLER 1979).

We also compared the structure of the *cel* operon from strain JF201 with its five Cel$^+$ mutants, strains BGH2001, BGH2007, BGH2008, BGH2009, and BGH2010 (HALL, BETTS and KRICKER 1986). No differences could be found when DNA from each of these strains was hybridized to fragments from the *cel* operon, suggesting that insertion sequences or other DNA rearrangements were not responsible for activation. Although changes of less than 100 bp would not be detected, it was concluded that the *cel* operons of these strains were probably activated either by point mutations or by suppressor mutations outside of the *cel* operon.

The mechanism of activation is strain dependent: To assess the relative frequency of insertion sequences and point mutations, we isolated an additional 16 mutants from strain MK1, and an additional 16 mutants from strain JF201. All mutants were isolated as papillae on MacConkey cellobiose medium. Fifteen of the 16 mutants from strain MK1 showed an increase in the size of the 2.3-kb *Cla*I/*Eco*RI fragment when

TABLE 1

***E. coli* strains and plasmids**

Strain	Genotype	Source
JF201	*Δlac X74 Δbgl-pho 201 ara gyrA thi*	A. WRIGHT
JF201R	*Δlac X74 Δbgl-pho 201 ara gyrA thi recA*	HALL, BETTS and KRICKER (1986)
BGH2001	*celD1* mutant of JF201 {Cel$^+$}	HALL, BETTS and KRICKER (1986)
BGH2007	*celD7* mutant of JF201 {Cel$^+$}	HALL, BETTS and KRICKER (1986)
BGH2008	*celD8* mutant of JF201 {Cel$^+$}	HALL, BETTS and KRICKER (1986)
BGH2009	*celD9* mutant of JF201 {Cel$^+$}	HALL, BETTS and KRICKER (1986)
BGH2010	*celD10* mutant of JF201 {Cel$^+$}	HALL, BETTS and KRICKER (1986)
MK1	HfrC *spc ΔlacZW4680 ΔebgA1011 bglR$^+$*	KRICKER and HALL (1984)
MK2	IS*2*::*celR1* mutant of MK1 {Cel$^+$}	KRICKER and HALL (1984)
LP100	*F$^-$rpsL trp his argG metA* or *B ara leu lacY lacZ Δ(W4680) bgl-pho Δ(201) celΔ(100)*	PARKER and HALL (1988)
LP201	IS*2*::*celR11* mutant of strain MK1 {Cel$^+$}	This study
LP202	IS*1*::*celR12* mutant of strain MK1 {Cel$^+$}	This study
LP203	IS*1*::*celR13* mutant of strain MK1 {Cel$^+$}	This study
LP204	IS*1*::*celR14* mutant of strain MK1 {Cel$^+$}	This study
LP207	IS*2*::*celR17* mutant of strain MK1 {Cel$^+$}	This study
LP209	IS*2*::*celR19* mutant of strain MK1 {Cel$^+$}	This study
LP216	IS*2*::*celR26* mutant of strain MK1 {Cel$^+$}	This study
GMS343	*F$^-$aroD6 argE3 lacY1 galK2 man4 mtl1 rpsL700 tsx29 supE44*	Coli Genetics Stock Center
LP201T	IS*2*::*celR11* transductant into strain GMS343	This study
LP202T	IS*1*::*celR12* transductant into strain GMS343	This study
LP204T	IS*1*::*celR14* transductant into strain GMS343	This study
CSH62T	*HfrH thi Δbgl-pho tna*::Tn*10*	KRICKER and HALL (1984)
CSH62TG	*gyrA* transductant of strain CSH62T	This study
JM109	*recA1 thi endA1 hsdR17 gyrA96 supE44 relA1 Δlac-proAB (F′ traD36 proAB lacIq ΔZM15)*	YANISH-PERRON, VIERA and MESSING (1985)

Plasmids	Genes	Source
pBRK10	IS2, IS*4*	KLAER *et al.* (1981)
pLG2	IS5	D. HARTL
pBRG36	IS*1*	BIEL, ADELT and BERG (1984)
pUF564	Entire *cel* operon strain MK2	HALL, BETTS and KRICKER (1986)
pUF650	Entire *cel* operon strain MK2	PARKER and HALL (1990)
pUF670	Entire *cel* operon strain BGH2008	PARKER and HALL (1990)
pUF671	Entire *cel* operon strain BGH2008	PARKER and HALL (1990)
pUF673	Entire *cel* operon strain JF201	PARKER and HALL (1990)
pUF681	*celR* and *celA* from strain BGH2008	PARKER and HALL (1990)
pUF687	*celR* and *celA* from strain JF201	PARKER and HALL (1990)

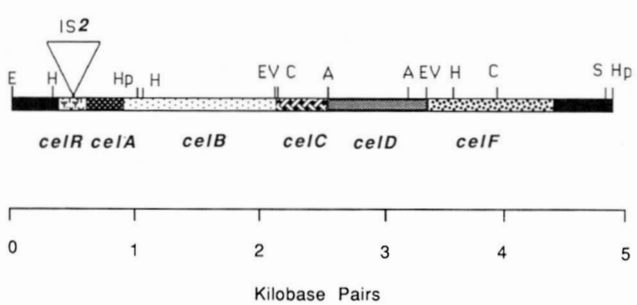

FIGURE 1.—Restriction map of the *cel* operon and flanking regions. The thick black line represents sequences outside of the *cel* operon. Abbreviations: A = *Ava*I, C = *Cla*I, E = *Eco*RI, EV = *Eco*RV, H = *Hin*dIII, Hp = *Hpa*I, S = *Sal*I.

DNA from these mutants was probed with *cel* DNA, indicating the presence of insertion sequences in the 5′ region of the *cel* operon (Figure 2). One mutant showed no structural change in the *cel* operon, indicating that it was probably activated by a point mutation. Further restriction analysis and hybridization experiments showed that the *cel* operons of four of the mutants contained IS*2*, all four of these elements integrated in orientation I (GHOSAL, SOMMER and SAEDLER 1979). Ten mutants harbored IS*1*; four integrated with the right end proximal to the *cel* operon, six integrated with the left end proximal to the *cel* operon (OHTSUBO and OHTSUBO 1978). The *cel* operon of one mutant contained IS*5*. Each insertion was localized to the same 800-bp *Hin*dIII/*Hpa*I fragment where the IS*2* of strain MK2 had integrated. When the structure of the *cel* operons from 16 mutants of strain JF201 were examined, the results were strikingly different. None of 16 mutants showed structural differences in the *cel* operon when they were compared with their parent wild-type strain. These mutants were probably activated either by point muta-

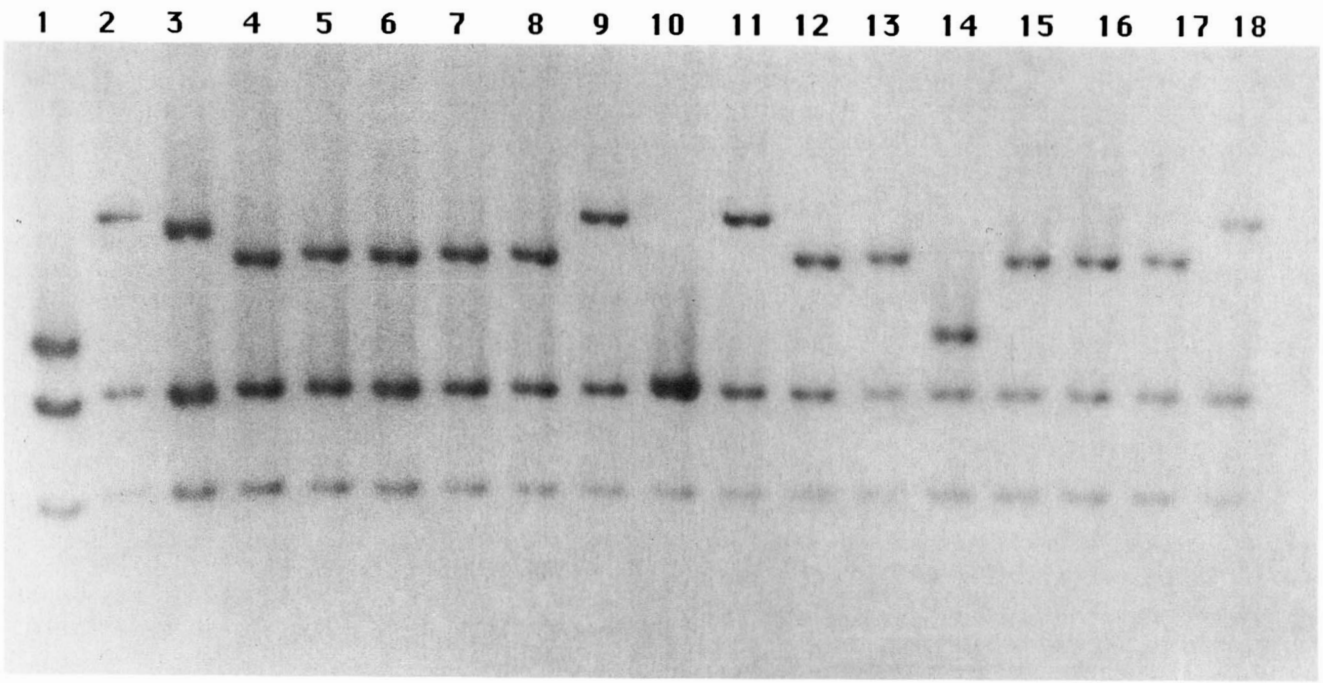

FIGURE 2.—Mutants of strain MK1 are activated by three different insertion sequences. Lane 1 shows wild-type strain MK1, lane 2 shows the IS*2* activated strain MK2. Lanes 3 to 18 show 16 additional mutants of MK1. Lanes 4, 5, 6, 7, 8, 12, 13, 15, 16 and 17 show IS*1* activated mutants. Lanes 3, 9, 11 and 18 show mutants activated by IS2. Lane 10 shows an IS*5* activated mutant, and lane 14 shows a mutant activated by a point mutation. DNA in all lanes was cut simultaneously with *Eco*RI and *Cla*I and probed with a 4-kb *Hpa*I fragment containing the entire *cel* operon. Note that no size increase is shown in the lane carrying the IS*5* mutant because *Eco*RI cuts within the element, creating a 1.9-kb doublet.

tions, or by suppressor mutations outside of the *cel* operon. Overall, 94% of the Cel$^+$ mutants of strain MK1 contained insertion sequences in the *cel* operon (PARKER and HALL 1990), while none of the Cel$^+$ mutants of strain JF201 contained insertion sequences.

To be sure that each of the activating events did, in fact, occur within the *cel* operon, each of the 16 mutants of strain MK1 was transduced into strain GMS343. *aroD*$^+$ transductants were selected and scored for the ability to utilize cellobiose. The Cel$^+$ phenotype cotransduced 25–35% of the time with *aroD*$^+$, consistent with the Cel$^+$ phenotype resulting from mutations in the *cel* operon which cotransduces with *aroD* 30% of the time (KRICKER and HALL 1984).

All 32 mutants were selected as papillae on MacConkey cellobiose medium, therefore it is unlikely that the method of selection influenced the mechanism of activation. It is more plausible that the different genetic backgrounds of strains MK1 and JF201 contributed to which mechanism of activation acted most frequently.

Strain JF201 has a mutation in the gene for DNA gyrase (*gyrA201*), suggesting that the level of genomic supercoiling might influence the mechanism of activation either by favoring point mutations, or by decreasing the frequency of transposition. This would be consistent with the observation that some mutations in *gyrA* have been shown to activate the *bgl* operon (DINARDO *et al.* 1982). To address this issue, we isolated four Cel$^+$ mutants of strain CSH62TG (*gyrA 201*) and two Cel$^+$ mutants of the isogenic *gyrA*$^+$ strain CSH62T. One of the four mutants of strain CSH62TG was activated by the insertion of IS2, making it unlikely that the *gyrA* mutation in strain JF201 was responsible for the failure to detect any insertionally activated *cel*$^+$ alleles in that strain. IS*2* integrated into the same 800-bp *Hin*dIII/*Hpa*I fragment, but this time, integration was in orientation II (GHOSAL, SOMMER and SAEDLER 1979). Neither Cel$^+$ mutant of strain CSH62T contained an insertion sequence in the *cel* operon. We conclude that although the preferred mechanism of activation is somewhat strain dependent, it probably does not depend upon the mutation at *gyrA*.

We also screened two natural isolates that express the *cel* operon (HALL and FAUNCE 1987). Genomic DNA was isolated, digested simultaneously with the restriction enzymes *Eco*RI and *Cla*I, separated on an agarose gel, and transferred to a filter. The DNA was hybridized, as above, with a 4.0-kb [^{35}S]dATP-labeled *Hpa*I fragment (HALL, BETTS and KRICKER 1986) containing the *cel* operon. Autoradiography indicated that one of these strains was activated by an insertion sequence and the other by a point mutation. We conclude that both mechanisms probably occur in

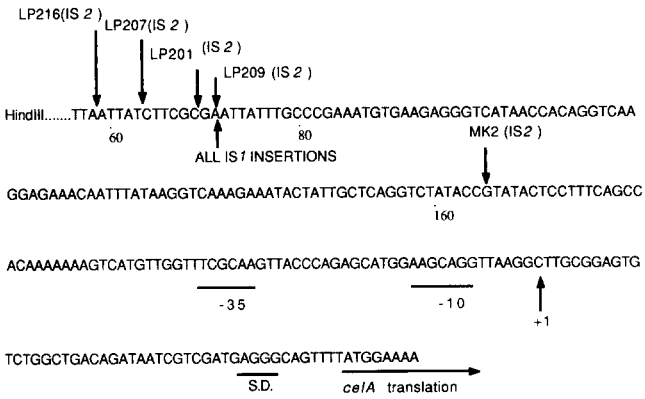

FIGURE 3.—The activating insertions fall within a 108-bp region 72–180 bp upstream of the transcriptional start site. The transcriptional start site is designated +1. The −10 and −35 promoter regions are underlined, as is the ribosome binding site of *celA*.

nature, and that neither is an artifact of studying the laboratory strain *E. coli* K12.

To determine whether the copy number of each element could account for the sharp contrasts in the preferred mechanism of activation, we examined the numbers of each insertion sequence in strains MK1 and JF201. DNA from these strains was digested with enzymes that did not cut within the insertion sequences, the fragments were separated by agarose gel electrophoresis, transferred to three separate filters, and hybridized with [^{35}S]dATP- labeled probes derived from each IS. Strain MK1 has 7 copies of IS*1* while strain JF201 has 6, MK1 has 9 copies of IS2 while JF201 has 6, and strain MK1 has 13 copies of IS*5* while strain JF201 has 12. Strain MK1 does have more copies of each element than does strain JF201, however it is unlikely that these small differences in copy number could account for such a large difference in the favored mechanism of activation in strains MK1 and JF201. These data suggest that unknown genetic factors influence which mechanism of activation occurs most frequently in a particular strain.

Insertion sequences integrate into a 108-bp region of the *cel* operon: Oligonucleotide primers corresponding to the complement of bp 902 to 921 of the *cel* operon (PARKER and HALL 1990) and bp 75 to 55 of IS*1* (OHTSUBO and OHTSUBO 1978) were used in the polymerase chain reaction (PCR) to amplify a region of DNA containing the junction between IS*1* and *cel* DNA for each of the six strains in which IS*1* was inserted with the left end proximal to *cel* DNA. The amplified DNA from each strain was sequenced using the oligonucleotide primer corresponding to IS*1* DNA. All of the IS*1* insertions occurred at the same site, between bp 71 and 72 of the *cel* operon (Figure 3). The experiment was also conducted with the same *cel* specific primer and an oligonucleotide primer corresponding to bases 70 to 51 of IS*2* (GHOSAL, SOMMER and SAEDLER 1979). The five IS*2* insertions in orientation I were sequenced, and a variety of insertion sites were found (Figure 3). The location of each of these elements when inserted in the opposite orientation was not determined, nor was the insertion site of IS*5* determined. Ten of the integration events were within 13 bp of each other, and only one strain, MK2, contained IS2 inserted 120 bp downstream.

Activating point mutations are not found in *celR*. It seemed likely that the point mutations that activated the *cel* operon would occur in the same regulatory region where the IS elements inserted. We therefore cloned a 6-kb *Eco*RI/*Bam*HI fragment containing the entire *cel* operon from the Cel$^+$ strains BGH2008 and BGH2009 into the *Eco*RI/*Bam*HI sites of the low copy number vector pHSG415 (HASHIMOTO-GHOTO *et al.* 1981). The resulting plasmids were designated pUF670 (BGH2008) and pUF675 (BGH2009). As expected, pUF670 and pUF675 conferred a Cel$^+$ phenotype on its host, strain JF201. In contrast, the wild type *cel* operon did not confer a Cel$^+$ phenotype even when cloned into a high copy number vector (pUF673). These data indicate that the activating mutations must lie within the 6 kb of cloned DNA, and rule out suppressor mutations, or other mutations outside the *cel* operon, as mechanisms of activation.

To examine the regulatory regions of pUF670 and pUF675, we subcloned the 800-bp *Hin*dIII/*Hpa*I fragment containing *celR* and *celA* from each plasmid into the vector pBlu$^+$. The insert DNA of each plasmid was sequenced using primers specific for the vector, for the complement of bp 300–324 of the *cel* operon (PARKER and HALL 1990), and for the complement of bp 519–535 of the *cel* operon (PARKER and HALL 1990). No differences between these sequences and the wild type sequence (PARKER and HALL 1990) were found, indicating that the activating point mutations did not lie in *celR* or *celA*.

Activating point mutations are in *celD*: It has been determined (PARKER and HALL 1990) that *celD* encodes a repressor of the *cel* operon, therefore, changes in this protein would be expected to activate the *cel* operon. The wild-type *celD* gene from plasmid pUF673 (PARKER and HALL 1990) was sequenced using four oligonucleotide primers specific for regions of *celD*. No differences were found between the wild-type *celD* sequence and the sequence of the *celD* gene of the IS*2*- activated allele from strain MK2 (PARKER and HALL 1990). Similarly, the *celD* gene from plasmid pUF671, which contains the entire *cel* operon from the Cel$^+$ point mutation activated strain BGH2008 (PARKER and HALL 1990), was sequenced. Only one sequence difference was found, a change at bp 2996 that resulted in a substitution of a threonine for a lysine (Table 2).

The *celD* gene was amplified by PCR from four additional strains in which the *cel* operon had been activated by point mutations, strains BGH2001,

TABLE 2

Base pair substitutions that activate the *cel* operon

Strain	Wild type		Mutant		Base pair
	Codon	Amino acid	Codon	Amino acid	
BGH2001	AAA	Lysine	GAA	Glutamic acid	2995
BGH2007	AAA	Lysine	GAA	Glutamic acid	2995
BGH2008	AAA	Lysine	ACA	Threonine	2996
BGH2009	AAG	Lysine	GAG	Glutamic acid	2983
BGH2010	AAG	Lysine	GAG	Glutamic acid	2983

BGH2007, BGH2009 and BGH2010, and sequenced. In each case, the sequence of the mutant *celD* allele differed from the wild-type sequence by a single base pair substitution in one of two lysine codons (Table 2). A total of three mutant sites in *celD* were thus identified.

The single base pair substitutions allow the repressor to recognize β-glucoside sugars as inducers. Phospho-β-glucosidase activity specified by the *cel* operon was measured under a variety of growth conditions by measuring the rate of *in vivo* hydrolysis of PNPG (Table 3). The amount of phospho-β-glucosidase produced by the wild-type organism (strain JF201) was too low to detect. However, when the wild-type *cel* operon was cloned onto the high copy number vector pBlu$^+$ (pUF673), a basal level was detectable and we were able to determine that the wild-type *cel* allele is induced 3-fold by cellobiose. Assuming a copy number of 50 for pUF673, this extrapolates to a basal level of about 0.0007 unit of phospho-β-glucosidase activity per A_{600}, and an induced level of about 0.002 unit per A_{600} for the chromosomal wild-type *cel* operon.

Mutations in *celD* increased the basal level of phospho-β-glucosidase, particularly noticeable in strains BGH2009 and BGH2010. A more significant change in the phospho-β-glucosidase level was found when each strain was grown in the presence of cellobiose: strains BGH2001 and BGH2007 showed approximately an 80-fold increase in the level of phospho-β-glucosidase production over the basal level, BGH2008, a 20-fold increase, and strains BGH2009 and BGH2010, a 30-fold increase. We conclude that the mutations in *celD* result in repressor that is able to better recognize cellobiose as an inducer. When grown on arbutin, BGH2001 and BGH2007 showed a 30-fold increase in phospho-β-glucosidase production, BGH2008 showed less than a 2-fold increase, and BGH2009 and BGH2010 showed an approximate 4-fold increase. These results indicate that substitution of either of the lysines with a glutamic acid results in the ability of the repressor protein to respond effectively to arbutin as an inducer, but the change to threonine permits only a marginal increase in the ability of the repressor protein to recognize arbutin as an inducer. Only strains BGH2009 and BGH2010 produced significantly more phospho-β-glucosidase than the wild-type strain when grown on salicin, indicating that only the substitution found in those two strains allows inducibility by salicin. None of these strains grow on salicin minimal medium, suggesting that a 5-fold increase in the amount of enzyme present is not enough to support growth.

When strains BGH2001, BGH2007, BGH2008, BGH2009 and BGH2010 were transformed with pUF688, a plasmid carrying the wild-type *celD* gene, the level of phospho-β-glucosidase activity in cells grown in the presence of arbutin, salicin, or cellobiose dropped to essentially zero indicating that the *celD* mutations are recessive and can be complemented by a high level of wild-type repressor protein (PARKER and HALL 1990). The observation that inducible strains become uninducible when a wild-type *celD* gene is present, adds additional evidence indicating that the mutations in *celD* are the activating mutations and strongly supports the notion that the mutations enable the repressor to recognize β-glucoside sugars as inducers. (If the wild-type repressor recognized these sugars as inducers, then the plasmid borne wild-type *celD* gene would not affect inducibility.)

Integration of insertion sequences increases the basal level of β-glucosidase production: Insertion sequences have been shown to increase gene expression by integrating into promoter regions and providing more efficient promoters (JAURIN and NORMAK 1983; PRENTKI *et al.* 1986; KLECKNER *et al.* 1978, KLECKNER, CHAN and TYE 1975). If the insertion sequences in the *cel* operon were functioning in this manner, it is expected that transcription would initiate at different sites in different insertionally activated strains. Therefore, we mapped the transcriptional start sites in four insertionally activated strains, LP201, LP202, LP203, MK91, and in the point mutation-activated strain BGH2008. All five stains have the same transcription initiation site (Figure 4). These data suggest that the insertion sequences do not function by providing outward reading promoters or by creating hybrid promoters.

The *cel* operon is regulated by a repressor, suggesting that the insertion sequences may function by disrupting the operator. If this were the case, it is expected that the *cel* operon would be expressed constitutively in strains that are insertionally activated. We measured the *cel* specified phospho-β-glucosidase activity of several insertionally activated strains, and found three distinct phenotypes. Strain MK91 harbors the *cel* allele from strain MK2 (KRICKER and HALL 1984), and constitutively synthesizes phospho-β-glucosidase at a very high level (Table 3) suggesting that the integration of IS2 between bp 166 and 167 may disrupt an operator site. The amount of phospho-β-

TABLE 3

Phospho-β-glucosidase activity

Strain	Activation	Inducer[a]			
		None	Cellobiose	Arbutin	Salicin
JF201	Wild type	<0.01[b]	<0.01	<0.01	<0.01
LP100/pUF673	Wild type	0.036 ± 0.010	0.107 ± 0.006	ND	ND
JF201 (calculated)[c]	Wild type	0.0007	0.002		
BGH2001	pt mutation	0.010 ± 0.002	0.738 ± 0.019	0.297 ± 0.063	0.020 ± 0.020
BGH2007	pt mutation	0.018 ± 0.003	1.07 ± 0.090	0.231 ± 0.043	0.036 ± 0.003
BGH2008	pt mutation	0.023 ± 0.004	0.469 ± 0.080	0.035 ± 0.007	0.020 ± 0.010
BGH2009	pt mutation	0.036 ± 0.012	1.07 ± 0.010	0.151 ± 0.018	0.151 ± 0.015
BGH2010	pt mutation	0.067 ± 0.009	1.28 ± 0.003	0.187 ± 0.030	0.150 ± 0.021
MK91	IS*2* bp 166	2.40 ± 0.160	1.32 ± 0.410	2.73 ± 0.200	2.85 ± 0.450
MK91/pUF688[d]		0.551 ± 0.130	2.57 ± 0.050	ND	ND
LP201	IS*2* bp 70	0.092 ± 0.004	0.885 ± 0.126	ND	ND
LP201/pUF688		0.033 ± 0.007	ND	ND	ND
LP202	IS*1* bp 72	0.068 ± 0.008	1.03 ± 0.145	ND	ND
LP202/pUF688		0.016 ± 0.001	ND	ND	ND
LP204	IS*1* bp 72	0.055 ± 0.016	0.148 ± 0.014	ND	ND
LP204/pUF688		0.016 ± 0.007	ND	ND	ND

[a] Carbon and energy source during growth was glucose. Cellobiose, arbutin and salicin were added as inducers.
[b] Units of β-glucosidase activity are expressed as nmol PNPG hydrolyzed min^{-1} per A_{600} ± 95% confidence limits.
[c] Calculation is based on an assumed copy number of 50 for pUF673. See text.
[d] pUF688 contains the wild-type allele of *celD* expressed from the *lac* promoter of pBlu$^+$.

glucosidase synthesized when strain MK91 is grown on cellobiose is somewhat lower than the level when the cells are grown on glucose. This reduction probably reflects the physiological state of the organism when grown on cellobiose, since cellobiose has been shown to function as both a nutrient and an inhibitor (HALL, BETTS and KRICKER 1986). If IS*2* does interrupt an operator site, it is expected that this mutation would be cis-dominant, and that additional copies of the wild-type repressor gene would not affect the amount of phospho-β-glucosidase synthesized. When strain MK91 was transformed with pUF688, which carries the wild-type repressor gene, the amount of phospho-β-glucosidase decreased 4 fold, indicating that the insertion does not completely prevent the binding of the repressor to a hypothetical operator.

The *cel* alleles from strains LP201 (*celR11*::IS*2*, *bglR*$^+$) LP202 (*celR12*::IS*1*, *bglR*$^+$) and LP204 (*celR14*::IS*1*, *bglR*$^+$) were transduced into strain GMS343 (which contains a wild-type (*i.e.*, cryptic) *bgl* operon) to permit the *in vivo* determination of *cel*-specified phospho-β-glucosidase activity. The transductants were designated LP201T, LP202T and LP204T.

Strain LP204T (*celR14*::IS*1*) contains IS*1* inserted 100 bp upstream of the site of the IS*2* insertion in strain MK91. The basal level of phospho-β-glucosidase synthesis in strain LP204T is higher than that of the wild-type strain, and in fact corresponds well to the level conferred by the wild-type allele when it is present on the high copy number plasmid pUF673 (Table 3). We estimate that the copy number of plasmid pUF673 is 50 to 100 copies per chromosome. If both the single copy allele in strain LP204T and the 50 copy wild-type allele on plasmid pUF673 permit the same level of phospho-β-glucosidase synthesis, then the basal level of expression of the *cel* operon in strain LP204T is about 50-fold higher than the wild-type basal level. Like the wild-type *cel* operon, the LP204T operon is induced about 3-fold by cellobiose. The level at which the LP204T allele is expressed can be contrasted with the level permitted by the MK2 allele, which is expressed at about 2500 times the wild-type level. It seems unlikely that two insertions that are separated by over 100 bp can both be within the same operator site.

Strains LP201T (*celR*::IS*2*) and LP202T (*celR*::IS*1*, orientation opposite to that of LP204T) contain insertion elements which have inserted at about the same site as strain LP204T, but their phenotypes are very different. The level of phospho-β-glucosidase activity produced when these two strains were grown on glucose was about the same as found in strain LP204T. However, the amount of phospho-β-glucosidase synthesized increased an additional 9-fold when they were grown on cellobiose, indicating that strains LP201T and LP202T are inducible. The level of induction of these two strains was significantly higher than the wild-type level of induction, suggesting that these strains may have mutations in *celD* as well as insertions into *celR*. When strains LP201T and LP202T were transformed with pUF688, carrying the wild-type *celD* gene, the level of phospho-β-glucosidase synthesis dropped 4-fold, indicating that, like

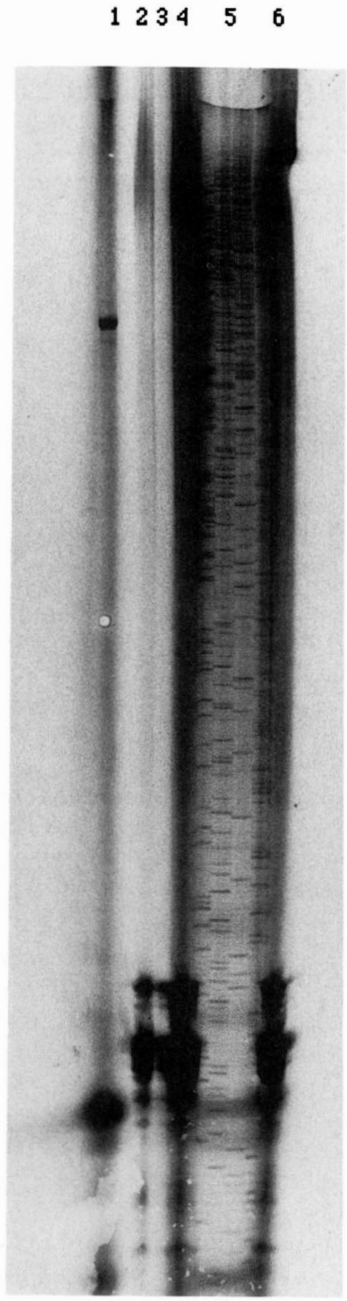

FIGURE 4.—S1 transcription start site analysis. Lane 1 shows probe DNA, lanes 2, 3 and 4 show protection of probe DNA by RNA from strains BHG2008 (Cel$^+$), MK1 (WT), and MK2 (Cel$^+$), respectively, lane 5 shows a sequencing reaction used as a size standard, and lane 6 shows protection of probe DNA by RNA from strain LP201.

strain MK91, these strains can be repressed by additional copies of the wild-type repressor gene.

From these results it cannot be concluded that the insertion sequences disrupt an operator. Strain MK91 carries IS*2* inserted the closest to the transcriptional start site, and although this strain does produce β-glucosidase constitutively, it is repressed by additional copies of the wild-type repressor. However, the meaning of this result is difficult to ascertain because the plasmid pBlu$^+$ is a very high copy number plasmid and the repressor may bind at a weak second operator when it is present at this high level. Strains LP201T, LP202T and LP204T show inducible expression of β-glucosidase, and repression in the presence of added copies of the repressor gene, indicating that these elements do not prevent repressor function, but instead simply increase the basal level of β-glucosidase production. It is suggested that the insertion sequences somewhat reduce binding of the repressor to its operator, but the mechanism is not known.

DISCUSSION

We have found that two different mechanisms are capable of activating the cryptic *cel* operon of *E. coli*. One mechanism involves the integration of insertion sequences in a 108-bp region upstream of the transcriptional start site. This finding is not especially surprising because it is well documented that insertion sequences are capable of controlling gene expression. These elements were first identified as the cause of strong polar mutations in genes such as the *gal* (JORDAN, SAEDLER and STARLINGER 1968; SHAPIRO 1969) operon. Several instances have been reported where insertion sequences increase expression of a nearby gene by fusing endogenous promoter sequences with promoter sequences carried on these insertion elements. Both IS*1* and IS*2* have been shown to function in this manner, providing a −35 region which, when combined with the endogenous −10 region, provides a promoter that increases transcription over that allowed by the endogenous promoter (JAURIN and NORMAK 1983; PRENTKI *et al.* 1986). Activation of silent genes by insertion sequences was first described by REYNOLDS, FELTON and WRIGHT (1981) who showed that the cryptic *bgl* operon could be activated by the insertion of IS*1* or IS*5* into a 47-bp region 78–125 bp upstream of the transcriptional start site (REYNOLDS, FELTON and WRIGHT 1981; REYNOLDS *et al.* 1986).

Our results are surprising because IS elements integrate into a region of the *cel* operon no closer than 72 bp upstream of the transcriptional start site, and as far as 180 bp upstream. Those insertion sequences do not function by creating a better promoter because the transcriptional start site is the same in each case, despite a wide range of integration sites. The basal level of transcription increases with the integration of each one of these elements, suggesting that the repressor does not bind as efficiently as it does in the wild type. Many repressors function by physically obscuring the promoter from the DNA-dependent RNA polymerase and cover a region of about 20–30 bp. It is unlikely that the primary binding site of the *cel* repressor would be found 180 bp upstream of the transcriptional start site, therefore the insertions sequences must not function by directly destroying this

site. Little is known about how a repressor protein identifies its own binding site. Sequences far upstream, while not composing the repressor site *per se*, may enable the protein to recognize a particular sequence that is farther downstream. The insertion sequences may act like enhancer elements in eukaryotes, either providing an initial binding site from which the protein later slides, or by affecting the local supercoiling of the area, creating a more easily recognizable region.

There are many similarities between the *bgl* and *cel* operons. Despite lack of sequence homology (PARKER and HALL 1990), both systems encode genes for utilization of β-glucoside sugars, and the cryptic state of both systems can be overcome by the integration of insertion sequences. IS*1* and IS*5* can activate both systems, but only the *cel* operon has been shown to be activated by IS2. Insertions in both the *bgl* and *cel* operons occur upstream of the transcription start site. In both cases, the exact mechanism of insertional activation is unknown.

Insertions sequences offer a mechanism for activating cryptic genes that is very attractive in terms of models proposed for their retention (HALL, YOKOYAMA and CALHOUN 1983; LI 1984) because IS elements are capable of excising precisely and restoring the cryptic state of the system. It has been proposed that cryptic genes are maintained in a population through repeated cycles of activation and cryptification (HALL, YOKOYAMA and CALHOUN 1983). The active allele of a particular gene may be selectively disadvantageous in one particular environment, and advantageous when the environment changes. Repeated cycles of activation and cryptification would maintain cryptic genes in a silent but potentially functional state (HALL, YOKOYAMA and CALHOUN 1983; LI 1984). An allele that is activated by an insertion sequence has a greater chance of repeated cycles of activation and cryptification because the change to the cryptic state is attainable by the precise excision of the element without affecting the properties of the structural or regulatory genes. Thus, when the silent allele is favored, an insertionally-activated allele might be silenced either by excision (restoring the wild-type sequence) or by any of numerous mutations that might disrupt the operon. When selection again favors expression, the silent wild-type operon could be activated either by an IS or by a point mutation. If, on the other hand, a second silencing mutation were present within the operon, then the probability of precise reversion of that mutation would be low and that allele might well be lost during selection for active alleles. An allele which was activated by a single base substitution has a greater chance of cryptification by a second mutation that would not be readily reversible, and thus might ultimately be lost from the population.

The *cel* operon is also activated by single base substitutions in *celD*. Amino acid substitutions of either of two lysines allow the repressor to recognize cellobiose, arbutin and to some extent salicin as inducers. Two out of three changes examined resulted in the substitution of the lysine with a glutamic acid, a change from a basic amino acid to an acidic amino acid. The other change resulted in the substitution of a lysine with a threonine, a replacement of a basic amino acid with a neutral amino acid. The amino acid substitutions appear to function by affecting the repressor-inducer interactions. Although no homology to other repressors has been found (PARKER and HALL 1990), it has been noted that *celD* contains a helix-turn-helix motif at the carboxy terminal end of the protein, a motif that is characteristic of other repressor proteins (MATTHEWS *et al.* 1982; OHLENDORF *et al.* 1982). It is thought that this motif is where the DNA binding site occurs (MATTHEWS *et al.* 1982; OHLENDORF *et al.* 1982). Mutations in *celD* map to a region upstream of the helix-turn-helix motif, possibly in the sugar-binding domain of *celD*.

It is interesting that *celD* is located in the middle of the operon and is under control of the major *cel* promoter. This implies that to maintain a silent state the *cel* operon must be transcribed at some low level. In fact, we have shown that when the wild-type operon is present on a high copy number plasmid, β-glucosidase activity can be detected. Thus, like the *sac* (PARKER and HALL 1988) and *bgl* (REYNOLDS, FELTON and WRIGHT 1981) operons, the cryptic *cel* operon is not entirely silent but allows an adequate level of transcription for repression. Likewise, in the mutant state, induction would require that only a small amount of sugar is transported. Low level transcription would permit induction.

Many mutations in *celD* would be expected to inactivate the *cel* repressor and thus confer a high level constitutive phenotype similar to that of strain MK2. Surprisingly, none of the point mutation-activated mutants were high level constitutives. In fact, none of the other mutants we have isolated have expressed the *cel* operon at a level approaching that of strain MK2. There may be selection against such high level constitutives that reduces the frequency of their recovery.

The *cel* operon is cryptic because it is in a constant state of repression. This raises the possibility that cellobiose-phosphate, arbutin-phosphate, and salicin-phosphate may not be the true inducers of the operon. In the natural environment, the *cel* operon may not be cryptic, but expressed in response to some form of the β-glucoside sugars that is unknown to us. While we cannot formally rule out this possibility, it seems

highly unlikely because most other known β-glucosides are extremely toxic and would probably be strongly selected against as inducer or substrates.

We are grateful to PAUL BETTS for his excellent technical advice. We would like to thank DANIEL HARTL for supplying plasmids pBRG36 and pLG2 which were used for construction of the IS*1* and IS*5* probes. This work was supported by U.S. Public Health Service grant GM 37110 to B.G.H.

LITERATURE CITED

AKROYD, J., B. BARTON, P. LUND, S. M. SMITH, K. SULTANA and N. SYMONDS, 1984 Mapping and properties of the *gam* and *sot* genes of phage Mu: their possible roles in recombination. Cold Spring Harbor Symp. Quant. Biol. **49:** 261–266.

BERK, A. J., and P. A. SHARP, 1977 Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digestion hybrids. Cell **12:** 721–732.

BIEL, S. W., G. ADELT and D. E. BERG, 1984 Transcriptional control of IS*1* transposition in *Escherichia coli*. J. Mol. Biol. **174:** 251–264.

DINARDO, S., K. A. VOELKEL, R. STERNGLANZ, A. E. REYNOLDS and A. WRIGHT, 1982 *Escherichia coli* DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. Cell **31:** 43–51.

GHOSAL, D., H. SOMMER and H. SAEDLER, 1979 Nucleotide sequence of the transposable DNA element IS2. Nucleic Acids Res. **6:** 1111–1121.

GROISMAN, E. A., and M. J. CASADABAN, 1986 Mini-Mu bacteriophage with plasmid replicons for in vivo cloning and *lac* gene fusing. J. Bacteriol. **168:** 357–364.

HALL, B. G., and P. W. BETTS, 1987 Cryptic genes for cellobiose utilization in natural isolates of *Escherichia coli*. Genetics **115:** 431–439.

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., and W. FAUNCE, III, 1987 Functional genes for cellobiose utilization in natural isolates of *Escherichia coli*. J. Bacteriol. **169:** 2713–2717.

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

HASHIMOTO-GOTOH, T., F. C. H. FRANKLIN, A. NORDHEI and K. N. TIMMIS, 1981 Specific-purpose plasmid cloning vectors. I. Low copy number, temperature sensitive, mobilization deficient pSC101-derived containment vectors. Gene **16:** 227–235.

JAURIN, B., and S. NORMAK, 1983 Insertion of IS2 creates a novel *ampC* promoter in *Escherichia coli*. Cell **32:** 809–816.

JORDAN, E., H. SAEDLER and P. STARLINGER, 1968 O^c and strong polar mutations in the *gal* operon are insertions. Mol. Gen. Genet. **102:** 353–363.

KLAER, R., S. KUHN, E. TILLMANN, H.-J. FRITZ and P. STARLINGER, 1981 The sequence of IS*4*. Mol. Gen. Genet. **181:** 169–175.

KLECKNER, N., D. BARKER, D. ROSS and D. BOTSTEIN, 1978 Properties of the translocatable element Tn*10* in *Escherichia coli* and bacteriophage lambda. Genetics **90:** 427–453.

KLECKNER, N., R. CHAN and B. K. TYE, 1975 Mutagenesis by insertion of a drug resistance element carrying an inverted repeat. J. Mol. Biol. **97:** 561–570.

KRICKER, M., and B. G. HALL, 1984 Directed evolution of cellobiose utilization in *E. coli* K-12. 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* K-12. Genetics **115:** 419–429.

LI, W.-H., 1984 Retention of cryptic genes in microbial populations. Mol. Ecol. Biol. **1:** 212–218.

MAHADEVAN, S., A. E. REYNOLDS and A. WRIGHT, 1987 Positive and negative regulation of the *bgl* operon of *Escherichia coli*. J. Bacteriol. **169:** 2570–2578.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MATTHEWS, B. W., D. H. OHLENDORF, W. F. ANDERSON and Y. TAKEDA, 1982 Cro repressor protein and its interaction with DNA. Cold Spring Harbor Symp. Quant. Biol. **47:** 427–433.

MILLER, J. H., 1972 *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

OHLENDORF, D. H., W. F. ANDERSON, R. G. FISHER, Y. TAKEDA and B. W. MATTHEWS, 1982 The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. Nature **298:** 718–723.

OHTSUBO, H., and E. OHTSUBO, 1978 Nucleotide sequence of an insertion element IS*1*. Proc. Natl. Acad. Sci. USA **25:** 615–619.

PARKER, L. L., and B. G. HALL, 1988 A fourth *Escherichia coli* gene system with the potential to evolve β-glucosidase utilization. Genetics **119:** 485–490.

PARKER, L. L., and B. G. HALL, 1990 Characterization and nucleotide sequence analysis of the cryptic *cel* operon of *Escherichia coli* K12. Genetics **124:** 455–471.

PRASAD, I., and S. SCHAEFLER, 1974 Regulation of the β-glucoside system in *Escherichia coli* K-12. J. Bacteriol. **120:** 638–650.

PRENTKI, P., B. TETER, M. CHANDLER and D. J. GALAS, 1986 Functional promoters created by insertion of the transposable element IS*1*. J. Mol. Biol. **191:** 383–393.

REYNOLDS, A. E., J. FELTON and A. WRIGHT, 1981 Insertion of DNA activates the cryptic *bgl* operon of *E. coli* K-12. 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.

REYNOLDS, A. E., S. MAHADEVAN, S. F. LEGRICE and A. WRIGHT, 1986 Enhancement of bacterial gene expression by insertion elements or by mutations in a CAP-cAMP binding site. J. Mol. Biol. **191:** 85–95.

ROSNER, J. L., 1972 Formation, induction, and curing of bacteriophage P1 lysogens. Virology **49:** 679–689.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCHNETZ, K., and B. RAK, 1988 Regulation of the *bgl* operon of *Escherichia coli* by transcriptional anti-termination. EMBO J. **7:** 3271–3277.

SCHNETZ, K., C. TOLOCZYKI and B. RAK, 1987 β-Glucoside (*Bgl*) operon of *Escherichia coli* K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes. J. Bacteriol. **169:** 2579–2590.

SHAPIRO, J. A., 1969 Mutations caused by the insertion of genetic material into the galactose operon of *Escherichia coli*. J. Mol. Biol. **40:** 93–105.

YANISH-PERRON, C., J. VIERA and J. MESSING, 1985 Improved M13 phage vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:** 103–119.

Communicating editor: J. R. ROTH