Activation of a Cryptic Gene by Excision of a DNA Fragment

LAURA L. PARKER, PAUL W. BETTS, AND BARRY G. HALL*

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

Received 9 March 1987/Accepted 25 September 1987

The cryptic *bgl* operon in *Escherichia coli* K-12 strain 1011A contains a 1.4-kilobase-pair fragment of foreign DNA within the *bglF* structural gene. The active allele found in its descendant strain, MK1, required the precise excision of that insertion for its activation. Molecular and genetic approaches have shown that strain 1011A possessed an active (*bglR*⁺) rather than a silent wild-type (*bglR*⁰) allele of the regulatory region and that this change was caused by a point mutation. Our model for the retention of cryptic genes (B. G. Hall, S. Yokoyama, and D. H. Calhoun, *Mol. Biol. Evol.* 1:109-124, 1983) suggested that the insertion might have been selected to silence a disadvantageous *bglR*⁺ allele. We examined the genealogy of strain MK1 and found that the insertion of foreign DNA was not selected for that reason, since it preceded the change to *bglR*⁺. This means that the change to *bglR*⁺ was also not selected, since the presence of the insertion would not allow expression of the operon. We have calculated the probability of isolating a *bglR*⁺ mutation by chance alone as less than 10⁻⁸. We suggest that mutation rates estimated under the usual conditions of exponential growth may be irrelevant to the frequencies of these events under natural conditions.

Genes for utilization of the β -glucoside sugars arbutin and salicin are normally cryptic in wild-type *Escherichia coli* strains (15). Mutations at the *bgl* locus increase the rate of expression of these genes to a level suitable for growth on arbutin and salicin (16) and thus confer a Bgl⁺ phenotype on the cell. It has been shown that insertion of IS1 or IS5 into a 47-base-pair region upstream from the start of transcription activates the *bgl* operon (17). In addition, ethyl methanesulfonate induced point mutations in the cyclic-AMP receptor protein binding site, and particular *gyrA* and *gyrB* mutations are capable of activating the operon (17).

We are interested in the molecular evolution of cryptic genes and in the variety of mechanisms by which different cryptic operons can be activated (6, 11). A previously described mutant of the *bgl* operon in strain MK1 (10) was isolated as a spontaneous mutant arising as a papilla on the surface of a colony of the parental strain, 1011A, growing on a salicin tetrazolium indicator plate. In the course of these studies, we examined the activated *bgl* operon in strain MK1. In this paper we show that, instead of integration of an insertion element into the *bglR* region, the activation of the *bgl* operon in strain MK1 involved excision of a DNA fragment from within a structural gene of the *bgl* operon.

MATERIALS AND METHODS

***E. coli* K-12 strains.** The strains used in this study are listed in Table 1. Plasmid pAR6 (16) contains the wild-type *bgl* allele inserted into the *EcoRI* site of pBR322. Plasmid pUF652 (this study) contains the active *bgl* allele inserted into the *EcoRI* site of pHSG415 (7).

Genetic nomenclature for the *bgl* operon. Until recently, *bglC* was used to designate the gene specifying the β -glucoside transport system, and *bglS* was used to designate the gene specifying a positive regulatory locus (15). Two recent papers (13, 18) reversed those definitions. To avoid further confusion, Barbara J. Bachmann, curator of the *E. coli* Genetic Stock Center, after consultation with the *E. coli* Genetic Stock Center Advisory Committee, has renamed these loci (B. Bachmann, personal communication). In this

paper, we use the new designations as they will appear in the next edition of the linkage map of *E. coli*. *bglF* now designates the β -glucoside-specific enzyme II of the phosphotransferase system, i.e., the transport protein. *bglG* now designates the gene encoding the positive regulator of expression of the *bgl* operon.

Culture media and conditions. Minimal media were prepared from a phosphate-buffered mineral salts solution (5) plus a sugar as a carbon source. Plates contained 1.5% agar as a solidifying agent. Glucose was used at 2 g/liter, and salicin was used at 1.1 g/liter. Required amino acids were used at 100 mg/liter, except for proline, which was used at 20 mg/liter.

Molecular techniques. Methods for the preparation of genomic DNA, isolation and manipulation of plasmids, isolation and labeling of probe DNA, and DNA-DNA hybridizations were as previously described (5). *bgl*-specific probes were made from fragments of plasmid pAR6 (17) as specified below.

Genetic methods. Bacterial matings were performed by mixing donor and recipient cells in a 1:1 ratio and incubating them in L broth for 2 h at 37°C. Exconjugants were selected for *his*⁺ on minimal glucose medium supplemented with arginine and methionine. *bglR*⁺ recombinants were selected on minimal salicin medium containing arginine and methionine.

Mutation rates. Mutation rates were measured by growing 100 1-ml cultures from inocula of 200 to 400 cells in minimal medium containing a concentration of glucose that limited the final density to about 2×10^8 cells per culture. Each culture was plated onto salicin minimal medium. Mutation rates were calculated by the method of Luria and Delbrück (12) with a computer program generously provided by J. Cairns.

RESULTS

Identification of an insertion into the *bgl* operon. As part of another study, the structure of the active *bgl* operon in strain MK1 was compared with that of the cryptic operon in its parent. Genomic DNA was prepared from the Bgl⁺ mutant strain MK1 and from its parent, strain 1011A. The DNAs were digested with restriction endonucleases, and the frag-

* Corresponding author.

TABLE 1. *E. coli* K-12 strains

Strain	Genotype
1011A.....	HfrC <i>spc</i> Δ <i>lacZ</i> W4680 Δ <i>egbA</i> 1011 <i>bglR</i> ⁺ , insertion:: <i>bglF</i>
AB3303	F ⁻ <i>thi</i> <i>hisG</i> <i>argE</i> <i>lacY</i> <i>galK</i> <i>xyl</i> <i>mtl</i> <i>rpsL</i> <i>tsx</i> <i>supE</i>
CSH62T	HfrH <i>thi</i> Δ <i>bgl</i> - <i>pho</i> <i>tna</i> ::Tn10
CSH62TG	HfrH <i>thi</i> Δ <i>bgl</i> - <i>pho</i> <i>tna</i> ::Tn10 <i>gyrA</i>
HfrC	HfrC <i>metB</i> <i>relA</i> <i>spoT</i> λ ⁺
K-12	Reference wild type, λ ⁺
MK1	<i>bglF</i> ⁺ mutant of 1011A
W1895	Stable clone from HfrC
χ 235.....	λ ⁻ mutant of W1895
χ 340.....	<i>proB</i> mutant of χ 235
χ 342.....	<i>proC</i> mutant of χ 235

ments were separated by agarose gel electrophoresis and transferred to a nylon membrane. The transferred fragments were hybridized to ³⁵S-labeled *bgl* operon DNA fragments obtained from plasmid pAR6 (17). It was expected that any insertion of DNA would be found in the regulatory region of strain MK1, as this is the mechanism of spontaneous activation that has been previously demonstrated (17). In contrast to this expectation, when *EcoRI* digests were probed with the 3.5-kilobase (kb) *HindIII* fragment from pAR6, the fragment from strain MK1 was larger than that from the parent, strain 1011A. Further experiments indicated that a *HpaI*-*PvuII* fragment is 1.4 kb larger in strain 1011A than in strain MK1 (Fig. 1A). A *HindIII*-*HpaI* digestion (Fig. 1B, lanes 1 and 2) shows two fragments of 2.0 and 2.6 kb in strain 1011A, which correspond to a single band of 3.2 kb in strain MK1. Since the insert contains no *HpaI* sites (Fig. 1A), this

FIG. 1. The *bgl* operon of strain 1011A contains a 1.4-kb insertion. (A) Lanes 1 and 2: 1011A and MK1, respectively, were digested simultaneously with *HpaI* and *PvuII* and probed with the 1.6-kb *HpaI*-*PvuII* fragment from pAR6. Arrows indicate sizes of bands in kilobases. (B) Lanes 1 and 2: 1011A and MK1, respectively, were digested simultaneously with *HindIII* and *HpaI*. Both were probed with the 12-kb *EcoRI* fragment from pAR6.

FIG. 2. Map of *bgl* region of *E. coli* K-12 showing insertion into *bglF*. Abbreviations: E, *EcoRI*; H, *HindIII*; Hp, *HpaI*; Pv, *PvuII*. The solid black region is external to the operon (13, 17, 18).

indicates that the insertion contains a *HindIII* site. Thus the parent, strain 1011A, possesses an insertion into that part of the *bgl* operon corresponding to *bglF*, the gene specifying the phosphoenolpyruvate-dependent transport protein of the *bgl* system (13). We have localized the insertion to the 334-base-pair *HinfI*-*DdeI* fragment extending from base pair 2637 to 2971 of the *bgl* operon (18; data not shown). The *bgl* operon in strain MK1 became active upon excision of the 1.4-kb insert. Figure 2 shows a map of the *bgl* region indicating the position of the 1.4-kb insertion.

Strain 1011A carries a *bglR*⁺ point mutation. Excision of the insertion from within *bglF* cannot, by itself, account for the *Bgl*⁺ phenotype of strain MK1, because the wild-type operon cannot be expressed. Strain MK1 must carry either a mutation in *bglR* or a mutation in one of the gyrase genes, *gyrA* or *gyrB*, which lie outside the *bgl* region. To distinguish between these possibilities, the *bgl*⁺ operon from strain MK1 was cloned as a 12.0-kb *EcoRI* fragment inserted into the *EcoRI* site of vector pHSG415 (7). The resulting plasmid was designated pUF652. Isogenic strains CSH62T and CSH62TG (*gyrA*) were transformed with pUF652. If a mutation in either the *gyrA* or the *gyrB* gene of strain MK1 was the activating factor, then it would be expected that the cloned *bgl*⁺ allele from MK1 would be inactive in a wild-type background. The allele remained active regardless of genetic background, indicating that the mutation responsible for the *Bgl*⁺ phenotype of strain MK1 was within the *bgl* operon.

Although the *EcoRI* fragments from the active allele (pUF652) and the cryptic allele (pAR6) appeared to be the same size, it was possible that the allele in MK1 had been activated by insertion of a small mobile DNA element into the *bglR* region. We compared the MK1 allele in plasmid pUF652 with the wild-type allele in plasmid pAR6. No differences could be detected when both plasmids were digested with three restriction enzymes with 4-base-pair recognition sites (*HinfI*, *MboI*, and *HaeIII*) and probed with the 12.0-kb *EcoRI* fragment from plasmid pAR6 (data not shown), suggesting that excision of the 1.4-kb insertion was precise. The failure to detect a structural difference between the active *bgl* allele of strain MK1 and the wild-type (cryptic) *bgl* allele present in plasmid pAR6 suggests that MK1 was activated by a point mutation or some other relatively small lesion in the *bglR* region.

Because activation of the *bgl* operon required excision of the 1.4-kb insertion from within the *bglF* gene, it seemed likely that the *bglR*⁺ mutation was already present in parent

strain 1011A. To test this hypothesis, we mated strain 1011A with the F^- strain AB3033 and selected $his^+ strA$ exconjugants. *bglR* is separated by 2.3 kb from the insertion into *bglF*, so that if strain 1011A carries *bglR*⁺, about 1% of the exconjugants are expected to carry *bglR*⁺ and *bglF*⁺. Among 2,500 exconjugants, 30 carried *bglR*⁺ and *bglF*⁺, indicating that the *bglR*⁺ mutation was present in 1011A.

Nature of the insertion into *bglF*. To the limits of restriction enzyme analysis, the excision event appears to be precise. This suggests that the 1.4-kb fragment of foreign DNA is a mobile DNA element such as an insertion sequence. However, the insertion contains a single *HindIII* site, a single *DdeI* site, and two *HinI* sites. This is inconsistent with IS1, which contains a *PstI* site and no *HindIII* site (14); IS2, which contains a *HpaI* site and five *NciI* sites (3); IS3, which contains two *HindIII* sites and a *PstI* site (19); IS4, which contains a *PstI* site and no *HindIII* sites (8); IS5, which contains an *EcoRI* site and no *HindIII* site (1); IS30, which contains an *NciI* site and no *DdeI* site (2); and IS186, which contains a *PstI* site and no *HindIII* site (9). It is unlikely that any mutant insertion sequences would be different from the wild type at more than one restriction site. We conclude that this insertion is either an insertion sequence which has not been previously described or a fragment of unknown origin generated by nonhomologous recombination.

History of the insertion and the *bglR*⁺ mutation. The model proposed by Hall et al. (6) for retention of cryptic genes in microbial populations suggests that active alleles of cryptic genes are disadvantageous in commonly encountered environmental conditions. It therefore seemed possible that the insertion was selected when an active *bgl* allele became disadvantageous. To test this hypothesis, we attempted to identify an ancestor of strain MK1 which carries *bglR*⁺ and lacks the insertion. Such a strain would be phenotypically *Bgl*⁺.

Figure 3 shows the genealogy of strain MK1. Strain χ 235 was not available to us. None of the available ancestral strains was phenotypically *Bgl*⁺. To determine whether the *Bgl*⁺ phenotypes could be due to the presence of the insertion, DNA was isolated from the available ancestral strains and from strain MK1. The foreign DNA insertion was absent in the original K-12 strain, in W1895, and in strain MK1, but it was present in strains χ 340, χ 342, and DB313 (Fig. 4). The insertion must therefore have been present in strain χ 235 and probably arose during UV curing of bacteriophage lambda. The foreign DNA insertion was also found in all the strains derived from strain χ 235 except for strain MK1 (data not shown).

Strain 1011A can mutate spontaneously to *Bgl*⁺ (10). Since strain W1895 lacks the insertion and is phenotypically *Bgl*⁺, this suggests that a *bglR*⁺ mutation must have arisen at some step between W1895 and strain 1011A in the lineage.

To determine whether a *bglR*⁺ mutation was present in strain χ 235, strain AB3303 was mated with strains χ 340 and χ 342. Among 14,000 $his^+ strA$ exconjugants from each mating, no *bglR*⁺ *bglF*⁺ recombinants were recovered, indicating that both strains carry the silent wild-type allele *bglR*⁰.

Frequency of the excision event. To estimate the excision rate, we performed a Luria-Delbrück fluctuation test (12), as described in Materials and Methods. Cultures of strain 1011A were plated onto salicin minimal medium at a mean density of 1.7×10^8 cells per culture. After 4 days of incubation, no colonies were present on any of the plates. The mutation rate estimated from the Poisson distribution is therefore less than 4.1×10^{-11} per cell division, and we can

have 97% confidence that the excision rate is less than 2×10^{-10} per cell division. Excision is clearly a very rare event.

Rate of mutation from *bglR*⁰ to *bglR*⁺. We estimated the rate of mutation from *bglR*⁰ to *bglR*⁺ for strain HfrC from a fluctuation test in which 100 independent cultures containing 3×10^8 cells were plated onto salicin minimal medium. The mean number of colonies per plate was 42.8. The distribution of colonies per plate gave a good fit to expectation when an average of 6 mutations occurred during the growth of the culture (mutation rate during growth = 2×10^{-8} per cell division) and 18 mutations occurred after the cells were plated (mutation rate on plates = 6×10^{-8}). No distinction was made between those mutants arising as the result of insertions into *bglR* and those arising from point mutations.

DISCUSSION

We have shown that activation of the *bgl* genes of strain MK1 required the precise excision of a 1.4-kb fragment of DNA found in *bglF*. The insertion into *bglF* appears to be quite stable, since it has persisted through this lineage and over a period of about 20 years, and the frequency of its excision is quite rare. It is therefore somewhat surprising that it was possible to isolate a *Bgl*⁺ mutant from strain 1011A.

The presence of the insertion cannot be explained in terms of a selective advantage as envisioned by the cryptic-gene model, since the *bgl* operon had not been decrytified by a *bglR*⁺ mutation at the time the insertion occurred.

It is likely that the insertion occurred when W1895 was irradiated with UV to cure the resident lambda prophage. Although we have no way to estimate the frequency of such an insertion event, it seems likely that the insertion is a rare event. It is difficult to imagine that this insertion into a silent operon was advantageous, and it therefore seems likely that it was recovered due to a low-probability chance event.

Strain DB313 (χ 342) was the starting strain for a series of studies in this laboratory on the *ebg* operon (4); we can therefore be certain that there was no selection for functions related to utilization of β -glucoside sugars between strains DB313 and strain 1011A of the lineage. Nevertheless, an unselected *bglR*⁺ mutation was isolated in the course of seven steps, none of which involved mutagenesis. The spontaneous frequency of the *bglR*⁺ mutation is 2×10^{-8} to 6×10^{-8} . Point mutations in *bglR* have been reported as occurring at a frequency of 17 of 1,000 spontaneous *bglR*⁺ mutations (18), suggesting that the overall frequency of spontaneous *bglR*⁺ mutations is on the order of 10^{-9} . It is difficult to believe that there was even an indirect selection for the *bglR*⁺ mutation, since the presence of the insertion into *bglF* still prevented the operon from functioning. We cannot formally rule out the possibility that the 1.4-kb insertion encodes some useful function and that there may have been selection for the *bglR*⁺ mutation that activated the operon. The fact that for a high level of expression the *bgl* operon requires induction by β -glucosides (15) makes this possibility appear remote. The probability of recovering a random unselected point mutation in *bglR* between strains DB313 and 1011A is therefore on the order of 10^{-8} .

The history of strain MK1 is a series of highly improbable events. The recovery of the insertion was unlikely, but it was obtained following treatment with UV, which may have increased the probability of the insertion event. The excision event was also unlikely, since the event occurs at a frequency of less than 10^{-10} . On the other hand, the excision

FIG. 3. Genealogy of strain 1011A.

event was selected on a plate that could plausibly have had 10^{11} cells on it.

The event that is most difficult to attribute to chance is the mutation from *bglR*⁰ to *bglR*⁺. That mutation arose during one of seven steps separating strain χ342 from strain 1011A (Fig. 3). The probability of isolating the *bglR*⁺ mutant by chance alone is $7 \times 6 \times 10^{-8}$, or less than one in a million.

When the probability of a hypothesis being true is less than 10^{-6} , the hypothesis is generally rejected and an alternative explanation is sought. The calculations are based on the assumption that the frequencies of events measured in healthy, growing cells in culture are relevant to cells in old colonies on plates. This may well not be the case. Between strain DS4680 and strain 1011 in the lineage (Fig. 3), there were three rounds of selection for lactose utilization. Each round involved picking papillae from several-week-old colonies on lactose tetrazolium plates. The cells in those colonies were subjected both to nutrient depletion and to

accumulation of waste products that would not occur in exponentially growing cultures in liquid medium. Both the mutation rates and the fitnesses associated with those mutations may be sufficiently different to account for the appearance of these mutations. Indeed, a similar situation was observed in the *ebg* system, where simultaneous double mutations in *ebgR* and *ebgA* appeared in mutants isolated as papillae at frequencies several orders of magnitude higher than could be explained on the basis of chance independent events (4).

These results must lead us to at least seriously question the relevance of mutation rates calculated under the optimal conditions found in exponentially growing cultures to mutation rates in natural environments. Exponential growth in liquid culture under optimal conditions of temperature, aeration, and nutrition is a very unnatural state for *E. coli* cells. Growth as an old colony under suboptimal conditions resembles natural conditions much more than do standard

FIG. 4. The *bgl* operon contains an insertion after strain W1895. DNA from *E. coli* K-12, W1895, χ 340, χ 342, DB313, and MK1 (lanes 1 to 6, respectively) was digested simultaneously with *Hpa*I and *Pvu*II and probed with the 1.6-kb *Hpa*I-*Pvu*II fragment from plasmid pAR6. The two smaller bands in the digest of MK1 DNA are the result of occasional star activity by either *Hpa*I or *Pvu*II.

growth conditions. An understanding of evolutionary rates and processes requires realistic estimates of mutation rates. Perhaps it is time for us to seriously study mutation rates under more natural conditions.

ACKNOWLEDGMENTS

We are grateful to J. Gallant and B. Bachmann for information on the history of these strains and to B. Bachmann for providing several of the strains.

This work was supported by Public Health Service grant GM 37110 from the National Institutes of Health.

LITERATURE CITED

1. Bukhari, A. I., J. A. Shapiro, and S. L. Adhya (ed.). 1977. DNA insertion elements, plasmids, and episomes. Cold Spring Har-

- bor Laboratory, Cold Spring Harbor, N.Y.
2. Dalrymple, B., P. Caspers, and W. Arber. 1984. Nucleotide sequence of the prokaryotic mobile genetic element IS30. *EMBO J.* 3:2145-2149.
3. Ghosal, D., H. Sommer, and H. Saedler. 1979. Nucleotide sequence of the transposable DNA element IS2. *Nucleic Acids Res.* 6:1111-1121.
4. Hall, B. G. 1982. Evolution on a petri dish, p. 85-150. In M. K. Hect, B. Wallace, and G. T. Prance (ed.), *Evolutionary biology*, vol. 15. Plenum Publishing Corp., New York.
5. Hall, B. G., and P. W. Betts. 1986. Cryptic genes for cellobiose utilization in natural isolates of *Escherichia coli*. *Genetics* 115: 431-439.
6. Hall, B. G., S. Yokoyama, and D. H. Calhoun. 1983. Role of cryptic genes in microbial evolution. *Mol. Biol. Evol.* 1: 109-124.
7. 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.
8. Klaer, R., S. Kuhn, E. Tillmann, H. J. Fritz, and P. Starlinger. 1981. The sequence of IS4. *Mol. Gen. Genet.* 181:169-175.
9. Kothary, R. K., D. Jones, and E. P. M. Candido. 1985. IS186: an *Escherichia coli* insertion element isolated from a cDNA library. *J. Bacteriol.* 164:957-959.
10. Krickler, M., and B. G. Hall. 1984. Directed evolution of cellobiose utilization in *E. coli* K-12. *Mol. Biol. Evol.* 1: 171-182.
11. Li, W.-H. 1984. Retention of cryptic genes in microbial populations. *Mol. Biol. Evol.* 1:212-218.
12. Luria, S. E., and M. Delbrück. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491-511.
13. 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.
14. Ohtsubo, H., and E. Ohtsubo. 1978. Nucleotide sequence of an insertion element IS1. *Proc. Natl. Acad. Sci. USA* 25:615-619.
15. Prasad, I., and S. Schaeffer. 1974. Regulation of the β -glucoside system in *Escherichia coli* K-12. *J. Bacteriol.* 120:638-650.
16. Reynolds, A. E., J. Felton, and A. Wright. 1981. Insertion of DNA activates the cryptic *bgl* operon of *E. coli* K-12. *Nature (London)* 293:625-629.
17. Reynolds, A. E., S. Mahadevan, S. LeGrice, and A. Wright. 1986. Enhancement of bacterial gene expression by insertion elements or mutations in a CAP-cAMP binding site. *J. Mol. Biol.* 191:85-95.
18. Schnetz, K., C. Toloczky, 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.
19. Sommer, H., J. Cullum, and H. Saedler. 1979. Integration of IS3 into IS2 generates a short sequence duplication. *Mol. Gen. Genet.* 177:85-89.