JOURNAL OF BACTERIOLOGY, Aug. 1990, p. 4696–4700
0021-9193/90/084696-05$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 172, No. 8

Molecular Cloning, Nucleotide Sequence, and Expression of *shl*, a New Gene in the 2-Minute Region of the Genetic Map of *Escherichia coli*

GUY LECLERC, GILLES NOËL, AND GABRIEL R. DRAPEAU*
Department of Microbiology and Immunology, Université de Montréal, Montréal, Québec, Canada H3C 3J7

Received 19 October 1989/Accepted 24 May 1990

Cells of *Escherichia coli* that harbor *supH* (an allele of the wild-type gene *serU*) are sensitive to UV irradiation and temperature and appear to have an impaired cell division control mechanism. We found that a gene located at the 2-min region, designated *shl*, inhibited the growth of *supH*-harboring cells when carried by a high-copy-number plasmid, whereas the same plasmid had no visible effect when present in parental cells. The amino acid sequence predicted from the nucleotide sequence of the *shl* gene indicated a similarity to the GalR and LacI repressor proteins, suggesting it is a transcription regulator. The sequence between the promoter and the structural genes revealed the presence of a short open reading frame of 28 amino acid residues followed by a segment of 81 base pairs. These structural features suggest that a transcription antitermination mechanism may be involved in the regulation of expression of the *shl* gene. The possibility that *shl* is a regulator of *serU* is discussed.

Recent studies have demonstrated that the *ftsM1* division mutation of *Escherichia coli* is in *serU*, a gene which encodes the dispensable serine tRNA-2 (3, 4, 11, 20). The *supH* and *ftsM1* mutations are probably identical. Both relieve the auxotrophies associated with the presence of the *leuB6* and *ilvD145* mutations, and both confer sensitivity to high temperatures and to irradiation by UV light at the permissive temperature. A lower colony-forming ability on L agar plates without added NaCl was also reported for the cells harboring either one of these mutations (4). Because UV sensitivity and colony formation on salt-free L agar are largely suppressed by the presence of a plasmid bearing a wild-type *ftsZ* gene and also because UV sensitivity is a phenotype of the division mutant *lon* (8), it was postulated that the activity of *supH* causes a cell division defect attributed to the synthesis of a specific faulty protein.

In the course of experiments aimed at locating and identifying the locus encoding this putative protein, we observed that a new unidentified gene from the 2-min region severely inhibited the growth of *supH* cells when carried by a multicopy plasmid, whereas the same plasmid had no detectable effect in the parental cell. This observation was of interest because the 2-min region contains several genes which are responsible for septum formation and cell division (1). We now report the location, cloning, and nucleotide sequence of this new gene. It will be shown that the deduced amino acid sequence of the encoded protein has a similarity with those of the Gal and Lac repressor proteins. This gene is also preceded by a leader region which appears to code for a short peptide of 28 amino acid residues.

Several plasmids obtained by the mini-Mu in vivo cloning procedure as described previously (14) and which harbored various segments of wild-type chromosomal DNA from the *leu* region were isolated after selection for prototrophy in a *leu* mutant. None of these plasmids seemed to have any growth inhibitory effect in a wild-type host, but when transferred into GD40 (*supH*) cells, most prevented colony formation on the minimal medium. This growth inhibition was also observed on L agar plates but not on the Casamino Acids medium. However, even if the colony-forming abilities appeared normal on the latter medium, microscopic examination of the cells revealed extensive filamentation, indicating that the inhibition of colony formation on the Casamino Acids medium was only less severe. DNA from one of these plasmids was digested with *Hin*dIII, and the fragments were isolated from an agarose gel and inserted into *Hin*dIII-restricted pBR322. Following propagation in AB1157, plasmid DNA was isolated and used for the transformation of AB1157 and GD40. One plasmid, pGL22, which carried a DNA fragment of 1.8 kilobases (kb), gave no transformant with GD40 on the minimal medium or on L agar plates, whereas the transformation efficiency appeared to be normal with AB1157, indicating that the inhibitory activity was due to a specific gene carried by this plasmid. (The acronym *shl* for suppressor H-linked phenotype is proposed for this gene, and it will be referred to as such hereafter in this report.) The 1.8-kb *Hin*dIII fragment had been located in the region between *ilvH* and *ftsI* (*pbpB*) (13) (see also Fig. 3). This was also confirmed by the nucleotide sequence data reported below.

The DNA sequence of the *shl* gene was carried out by the dideoxy-chain terminating method (16) with the sequencing system obtained from United States Biochemicals. By using the deletion method of Henikoff (7), the two strands of the 1.8-kb *Hin*dIII DNA fragment, inserted in M13mp19, were sequenced. This is shown in Fig. 1 together with an amino acid translation. The 5′-terminal region of the sequence was found to contain the terminal end of the *ilvH* gene, the sequence of which has been reported (18). Two differences were noted between the published sequence and that presented in Fig. 1. First, at positions 7 and 8 the sequence GA had been inverted, and second, an A was missing at position 89. As a result, the terminating codon would not be expected to be TAA at position 95 but rather the TGA codon located eight base pairs downstream.

Two open reading frames (ORF) were found. One has 28 codons and starts at position 117, and the other comprises

* Corresponding author.

```
                                                                              -35
AAGCTTGATGCATTTTTAGCATCGATTCGCGATGTGGCGAAAATTGTGGAGGTTGCTCGCTCTGGTGTGGT

                 -10        SD             MetArgAsnLeuGlnProAsnMetSer
CGGACTTTCGCGCGGCGATAAAATAATGCGTTGAGAATGATCTCAATGCGCAATTTACAGCCCAACATGTCA
                                           117

ArgTrpAlaPhePheAlaLysSerValGlyThrTrpAsnLysSerSerCysArgSer***
CGTTGGGCTTTTTTGCGAAATCAGTGGGAACCTGGAATAAAAGCAGTTGCCGCAGTTAATTTTCTGCGC
                                                           2Ø3
                                                                   SD
TTAGATGTTAATGAATTTAACCCATACCAGTACAATGGCTATGGTTTTTACATTTTACGCAAGGGGCAATT

MetLysLeuAspGluIleAlaArgLeuAlaGlyValSerArgThrThrAlaSerTyrValIleAsnGlyLys
GTGAAACTGGATGAAATCGCTCGGCTGGCGGGAGTGTCGCGGACCACTGCAAGCTATGTTATTAACGGCAAA
285
AlaLysGlnTyrArgValSerAspLysThrValGluLysValMetAlaValValArgGluHisAsnTyrHis
GCGAAGCAATACCGTGTGAGCGACAAAACCGTTGAAAAAGTCATGGCTGTGGTGCGTGAGCACAATTACCAC
                          385
ProAsnAlaValAlaAlaGlyLeuArgAlaGlyArgThrArgSerIleGlyLeuValIleProAspLeuGlu
CCGAACGCCGTGGCAGCTGGGCTTCGTGCTGGACGCACACGTTCTATTGGTCTTGTGATCCCCGATCTGGAG
                                                     485
AsnThrSerTyrThrArgIleAlaAsnTyrLeuGluArgGlnAlaArgGlnArgGlyTyrGlnLeuLeuIle
AACACCAGCTATACCCGCATCGCTAACTATCTTGAACGCCAGGCGCGGCAACGGGGTTATCAACTGCTGATT

AlaCysSerGluAspGlnProAspAsnGluMetArgCysIleGluHisLeuLeuGlnArgGlnValAspAla
GCCTGCTCAGAAGATCAGCCAGACAACGAAATGCGGTGCATTGAGCACCTTTTACAGCGTCAGGTTGATGCC
         585
IleIleValSerThrSerLeuProProGluHisProPheTyrGlnArgTrpAlaAsnAspProPheProIle
ATTATTGTTTCGACGTCGTTGCCTCCTGAGCATCCTTTTTATCAACGCTGGGCTAACGACCCGTTCCCGATT
                                     685
ValAlaLeuAspArgAlaLeuAspArgGluHisPheThrSerValValGlyAlaAspGlnAspAspAlaGlu
GTCGCGCTGGACCGCGCCCTCGATCGTGAACACTTCACCAGCGTGGTTGGTGCCGATCAGGATGATGCCGAA
                                                                  785
MetLeuAlaGluGluLeuArgLysPheProAlaGluThrValLeuTyrLeuGlyAlaLeuProGluLeuSer
ATGCTGGCGGAAGAGTTACGTAAGTTTCCCGCCGAGACGGTGCTTTATCTTGGTGCGCTACCGGAGCTTTCT

ValSerPheLeuArgGluGlnGlyPheArgThrAlaTrpLysAspAspProArgGluValHisPheLeuTyr
GTCAGCTTCCTGCGTGAACAAGGTTTCCGTACTGCCTGGAAAGATGATCCGCGCGAAGTGCATTTCCTGTAT
                     885
AlaAsnSerTyrGluArgGluAlaAlaAlaGlnLeuPheGluLysTrpLeuGluThrHisProMetProGln
GCCAACAGCTATGAGCGGGAGGCGGCTGCCCAGTTATTCGAAAAATGGCTGGAAACGCATCCGATGCCGCAG
                                                   985
AlaLeuPheThrThrSerPheAlaLeuLeuGlnGlyValMetAspValThrLeuArgArgAspGlyLysLeu
GCGCTGTTCACAACGTCGTTTGCGTTGTTGCAAGGAGTGATGGATGTCACGCTGCGTCGCGACGGCAAACTG

ProSerAspLeuAlaIleAlaThrPheGlyAspAsnGluLeuLeuAspPheLeuGlnCysProValLeuAla
CCTTCTGACCTGGCAATTGCCACCTTTGGCGATAACGAACTGCTCGACTTCTTACAGTGTCCGGTGCTGGCA
    1Ø85
ValAlaGlnArgHisArgAspValAlaGluArgValLeuGluIleValLeuAlaSerLeuAspGluProArg
GTGGCTCAACGTCACCGCGATGTCGCAGAGCGTGTGCTGGAGATTGTCCTGGCAAGCCTGGACGAACCGCGT
                                  1185
LysProLysProGlyLeuThrArgIleLysArgAsnLeuTyrArgArgGlyValLeuSerArgSer***
AAGCCAAAACCTGGTTTAACGCGCATTAAACGTAATCTCTATCGCCGCGGCGTGCTCAGCCGTAGCTAAGCC

GCGAACAAAAATACGCGCCAGGTGAATTTCCCTCTGGCGCGTAGAGTACGGGACTGGACATCAATATGCTTA
AAGTAAATAAGACTATTCCTGACTATTATTGATAAATGCTTTTAAACCCGCCCGTTAATTAACTCACCAGCT
GAAATTCACAATAATTAAGTGATATCGACAGCGCGTTTTTGCATTATTTTGTTACATGCGGCGATGAATTGC
CGATTTAACAAACACTTTTCTTTGCTTTTGCGCAAACCCGCTGGCATCAAGCGCCACACAGACGTAACAAGG
ACTGTTAACCGGGGAAGATATGTCCTAAAATGCCGCTCGCGTCGCAAACTGACACTTTATATTTGCTGTGGA
AAATAGTGAGTCATTTTAAAACGGTGATGACGATGAGGGATTTTTTCTTACAGCTATTCATAACGTTAATTT
GCTTCGCACGTTGGACGTAAAATAAACAACGCTGATATTAGCCGTAAACATCGGGTTTTTTACCTCGGTATG
CCTTGTGACTGGCTTGACAAGCTT
```

FIG. 1. Nucleotide sequence of the 1.8-kb *Hin*dIII fragment and the predicted amino acid sequences of the leader peptide and the Shl protein. The leader polypeptide is encoded by nucleotides 117 through 201, and the major ORF of the *shl* structural gene is shown from nucleotides 285 to 1287. The potential ribosome-binding sites (SD) for the leader and *shl* gene are underlined. The −10 and −35 regions are underlined and labeled. Dyad symmetries are indicated by horizontal arrows. The sites of insertion of the S end and the adjacent *lacZ* region of the mini-Mu replicons pGL15 and pGL19 are indicated by vertical arrows (see Fig. 3).

334 codons. The latter is postulated to use GUG as the initiation codon at position 285. A promoter sequence with high similarity to the consensus *E. coli* promoter sequence (6), notably at the −10 region, was observed. It would appear to overlap with the coding region of *ilvH*. In fact, only 11 base pairs would separate the predicted TGA stop codon of *ilvH* and the initiation codon ATG of the first ORF. A prospective Shine-Dalgarno sequence, GAGA, is located 9 base pairs upstream of this initiation codon. A recognizable ribosome-binding site, AAGG, precedes the major ORF, but there does not appear to be a readily identifiable promoter sequence. Another relevant structure of the DNA sequence is the presence of three potential stem loop structures. One is located immediately downstream of the assumed −10 region. It encompasses the *ilvH* stop codon, the ribosome-binding site, and the ATG codon of the first ORF. A second perfect hairpin structure of 8 base pairs followed by a run of six uridine residues in the transcript is observed some 10

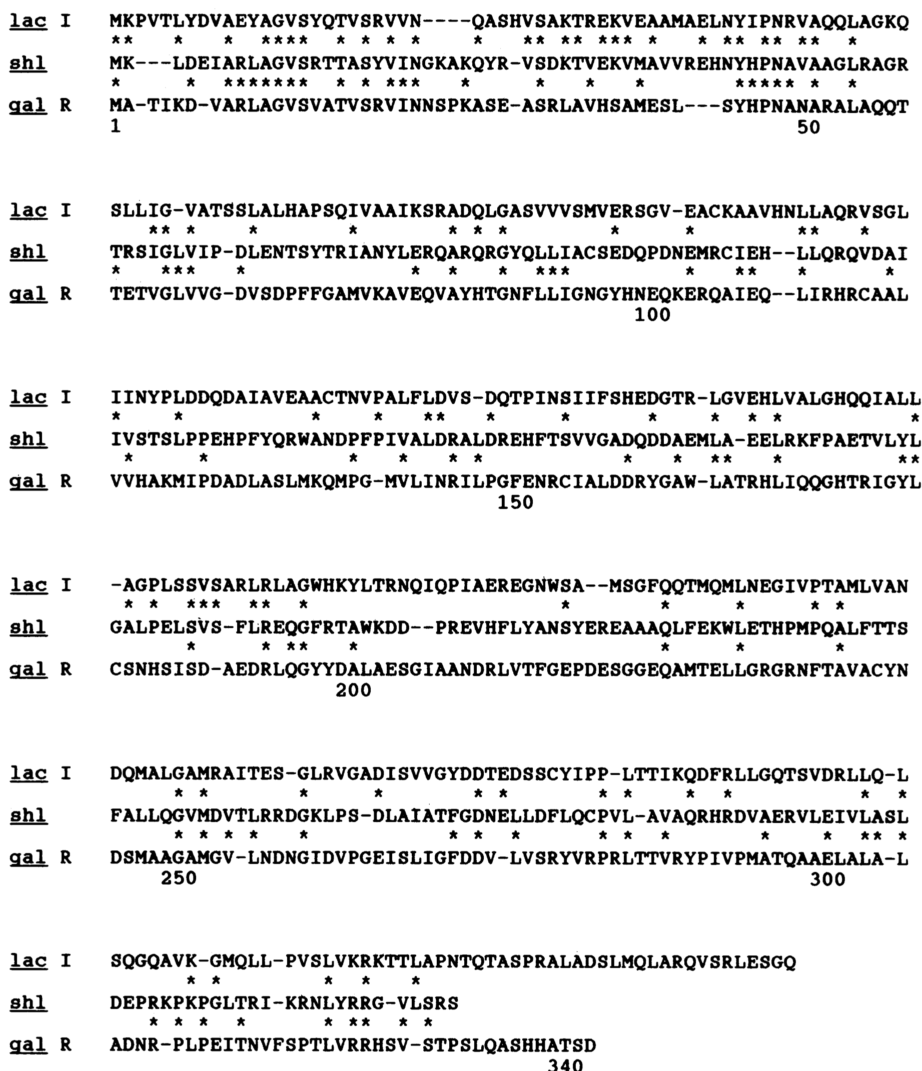

FIG. 2. Comparison of the amino acid sequences of the LacI, Shl, and GalR repressor proteins. Identical amino acids are indicated with an asterix. Gaps have been inserted in the sequences to maximize sequence identities. Numbers refer to the amino acid residues of the GalR repressor.

base pairs downstream of the ATG codon, a structure that has the properties of a *rho*-independent terminator (15). Similar terminators within the ORF of the leader regions of the *pyrBI* and the *bgl* operons of *E. coli* and of the amidase gene of *Pseudomonas aeruginosa* have been reported and shown to play a critical role in the expression of their respective operons (2, 10, 12, 17). A third hairpin structure, which is also followed by an uridine-rich sequence, is detected between the two ORFs.

The *shl* major coding region would encode a protein of 334 amino acid residues. A similarity search with GenBank (Release 55) and National Biomedical Research Foundation (Release 18) sequence libraries revealed considerable sequence similarity between the deduced amino acid sequence of Shl and those of the Lac and Gal repressors (21). The amino acid identities are highlighted in Fig. 2. It can be seen that the similarity extends over most of the amino acid sequence of the protein, but it is particularly strong in the amino-terminal region, a region considered to contain the DNA-binding site (sequence 1 to 56) and commonly referred to as the headpiece. The degree of sequence similarity for this region is very high, being 45 and 52% for the Gal and the Lac repressor, respectively. Furthermore, the first 22 amino acids of the protein could form a helix-turn-helix DNA-binding motif.

To determine whether the short ORF preceding the *shl* gene is transcribed and translated in vivo, a plasmid carrying the postulated coding region of the leader polypeptide was fused in frame to the *lacZ* structural gene by using the mini-Mu replicon pEG109 (5). This is represented in Fig. 3. One of the plasmids, pGL15, had the fusion site within codon 22 of the leader polypeptide, while in pGL19, the fusion was at amino acid residue 210 of the major ORF (see Fig. 1). Since plasmid pGL15 conferred a lactose-positive phenotype to the harboring cells on X-gal plates, the putative leader region of *shl* is likely translated into a peptide. However, the β-galactosidase activity expressed by the fusion plasmid pGL15 was very low, considering that it is a high-copy-number plasmid (Table 1). The possibility that the presence of other genes which are located upstream of the fusion site had an influence on the expression of the leader sequence was examined. Plasmid DNA from pGL15 was

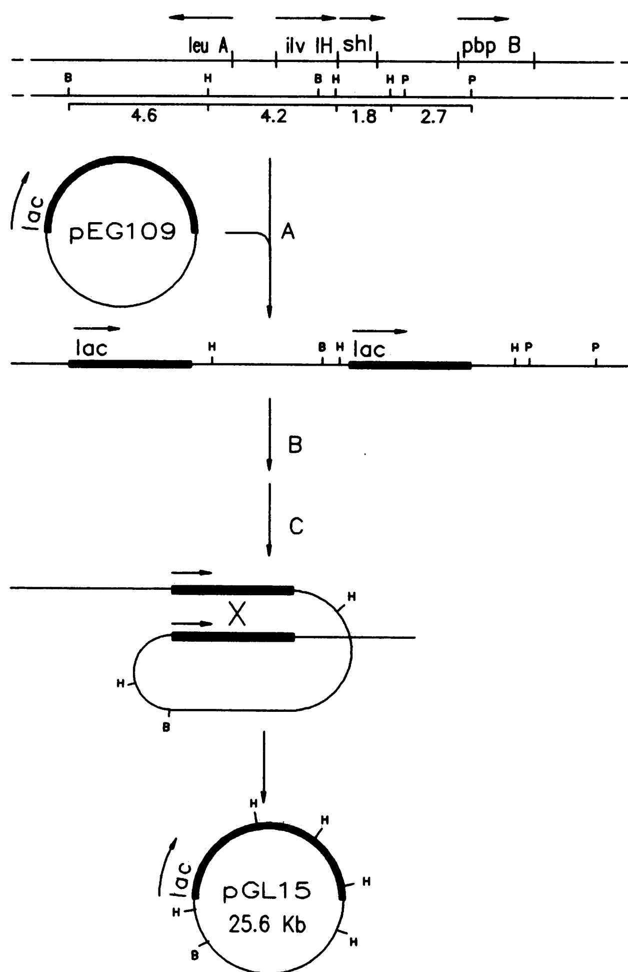

FIG. 3. Construction of the gene fusions. The mini-Mu replicon pEG109 (thick bar) transposes many times in each cell, and occasionally a fragment of host DNA bearing the *leu-pbpB* region becomes flanked by two copies of mini-Mu in the same orientation (A). Packaging (B) then proceeds by a head-full mechanism, and after infection of a recipient cell (C), recombination occurs between mini-Mu homologous sequences to form plasmid pGL15. The orientation of the *lac* region is indicated with an arrow. Distances between restriction sites are reported in kilobase pairs. The arrows indicate the direction of transcription. Restriction enzyme abbreviations: E, *Eco*RI; H, *Hin*dIII; B, *Bam*HI; P, *Pst*I.

cleaved with *Hin*dIII, and a 7.4-kb fragment which contained the leader-*lacZ* fusion and only a short segment (about 100 nucleotides) from the *ilvH* gene was inserted into the *Hin*dIII site of pGL25, a plasmid with the F replication origin derived from pDF41 (9). As shown in Table 1, this single-copy plasmid was found to convey about the same level of β-galactosidase activities as the multicopy plasmid pGL15, suggesting that the presence of one or more of the genes or DNA sequences located upstream of the fusion site negatively affected the expression of the leader-*lacZ* fusion in pGL15. Alternatively, the reduced β-galactosidase activity of the leader-*lacZ* fusion in pGL15 could be due to the presence of the *ilvIH* operon, which would antagonize by oncoming transcription and translation the initiation of transcription of the *shl* leader, which is transcribed in the same direction. Unlike the situation for the leader sequence, the activities of the *shl-lacZ* fusion in pGL19 were considerably

TABLE 1. β-Galactosidase activities of leader- and *shl-lacZ* gene fusions carried by different plasmids

Plasmids	Replication origin and characteristics	Activity (units)
pGL15	p15A; mini-Mu replicon; leader-*lacZ* fusion	27
pGL19	p15A; mini-Mu replicon; *shl-lacZ* fusion	810
pGL25	F; leader-*lacZ* fusion	38
pGL27	F; *shl-lacZ* fusion	40

higher, suggesting that the presence of the upstream DNA segment did not have the same interfering effect with the expression of the *shl* gene. Also shown in Table 1 are the β-galactosidase activities of the *shl-lacZ* gene fusion carried by the mini-F plasmid pGL27. The activities were 20-fold lower than for pGL19, a result expected on the basis of the lower number of gene copies per cell.

The homology that the Shl protein shares with the Gal and Lac repressor proteins strongly suggests that it is a transcriptional regulator. When it is carried by a high-copy-number plasmid, *shl* has a growth inhibitory effect only in *supH* mutant cells. The poor viability of these cells could be explained by an effect of *supH* promoting the expression of *shl*. This appears improbable because we have found that the presence of the *shl* gene on the pKK223-3 expression vector did not inhibit growth of wild-type cells even under fully induced conditions (results not shown). An alternative possibility would be that the Shl protein regulates the synthesis or the activity of *serU* (*supH*). In the region very near the *shl* gene is the *serR* gene, mutations of which affect the levels of seryl tRNA synthetase (19). It is plausible that an excess of such a regulator would be more dramatic in a *supH* mutant than in wild type, since one substrate of the synthetase, serine tRNA-2, is altered in the mutant. In fact, evidence that an increase in the *supH* activity is growth inhibitory was suggested from the observation that *supH* could be cloned only on a pSC101-derived plasmid (11). Therefore, it is possible that *shl* is in fact *serR*, a regulator of *serU* expression.

We are indebted to Normand Brisson and Daniel Gautheret for their help with the analysis of the nucleotide and amino acid sequences and Gilles Leclerc for helpful suggestions during these studies. We thank M. J. Casadaban for the mini-Mu replicon pEG109 and Marie Lesage for expert typing of the manuscript.

This work was supported by the Medical Research Council of Canada.

LITERATURE CITED

1. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, Edition 7. Microbiol. Rev. **47**:180–230.
2. **Drew, R., and N. Lowe.** 1989. Positive control of *Pseudomonas aeruginosa* amidase synthesis is mediated by a transcription anti-termination mechanism. J. Gen. Microbiol. **135**:817–823.
3. **Eggertsson, G.** 1968. Suppressors causing temperature sensitivity of growth in *Escherichia coli*. Genetics **60**:269–280.
4. **Eggertsson, G., and E. A. Adelberg.** 1965. Map positions and specificities of suppressor mutations in *Escherichia coli* K-12. Genetics **52**:319–340.
5. **Groisman, E. A., B. A. Castilho, and M. J. Casadaban.** 1984. *In vivo* DNA cloning and adjacent gene fusing with a mini-Mu-*lac* bacteriophage containing a plasmid replicon. Proc. Natl. Acad. Sci. USA **81**:1480–1483.
6. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.

7. **Henikoff, S.** 1987. Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol. **155**:156–165.
8. **Howard-Flanders, P., E. Simpson, and K. Theriot.** 1964. A locus that controls filament formation and sensitivity to irradiation in *Escherichia coli* K-12. Genetics **49**:237–246.
9. **Kahn, M., R. Kolter, C. Thomas, D. Figurski, R. Meyer, E. Remant, and D. R. Helinski.** 1979. Plasmid cloning vehicles derived from plasmids ColE1, F, R6K, and RK2. Methods Enzymol. **68**:268–280.
10. **Kenneth, L. R., F. E. Powell, and C. L. Turnbough, Jr.** 1985. Role of translation and attenuation in the control of *pyrBI* operon expression in *Escherichia coli* K-12. J. Bacteriol. **163**: 991–999.
11. **Leclerc, G., C. Sirard, and G. R. Drapeau.** 1989. The *Escherichia coli* cell division mutation *ftsM1* is in *serU*. J. Bacteriol. **171**:2090–2095.
12. **Mahadevan, S., and A. Wright.** 1987. A bacterial gene involved in transcription antiterminator: regulation at a rho-independent terminator in the *bgl* operon of *E. coli*. Cell **50**:485–494.
13. **Nakamura, M., I. N. Maruyama, M. Soma, J. Kato, H. Suzuki, and Y. Hirota.** 1983. On the process of cellular division in *Escherichia coli*: nucleotide sequence of the gene for penicillin-binding protein 3. Mol. Gen. Genet. **191**:1–9.
14. **Noël, G., and G. R. Drapeau.** 1986. Identification of new cell division genes in *Escherichia coli* by using extragenic suppressors. J. Bacteriol. **165**:399–404.
15. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
16. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
17. **Schnetz, K., and B. Rak.** 1988. Regulation of the *bgl* operon of *Escherichia coli* by transcriptional antitermination. EMBO J. **7**:3271–3277.
18. **Squires, C. H., M. DeFelice, J. Devereux, and J. M. Calvo.** 1983. Molecular structure of *ilvIH* and its evolutionary relationship to *ilvG* in *Escherichia coli* K12. Nucleic Acids Res. **11**:5299–5313.
19. **Theall, G., K. B. Low, and D. Söll.** 1979. Regulation of the biosynthesis of aminoacyl-tRNA synthetases and of tRNA in *Escherichia coli*. IV. Mutants with increased levels of leucyl- or seryl-tRNA synthetase. Mol. Gen. Genet. **169**:205–211.
20. **Thorbjarnardóttir, S., H. Uemura, T. Dingermann, T. Rafnar, S. Thorsteinsdóttir, D. Söll, and G. Eggertsson.** 1985. *Escherichia coli supH* suppressor: temperature-sensitive missence suppression caused by an anticodon change in $tRNA_2^{Ser}$. J. Bacteriol. **161**:207–211.
21. **Wilcken-Bergmann, B. V., and B. Müller-Hill.** 1982. Sequence of galR gene indicates a common evolutionary origin of *lac* and *gal* repressor in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **79**:2427–2431.