JOURNAL OF BACTERIOLOGY, Dec. 2001, p. 6943–6946 Vol. 183, No. 23
0021-9193/01/$04.00+0 DOI: 10.1128/JB.183.23.6943–6946.2001

NOTES

IS*186* Insertion at a Hot Spot in the *lon* Promoter as a Basis for Lon Protease Deficiency of *Escherichia coli* B: Identification of a Consensus Target Sequence for IS*186* Transposition

L. SAISREE,[1] MANJULA REDDY,[1] AND J. GOWRISHANKAR[1,2]*

Centre for Cellular and Molecular Biology, Hyderabad 500007,[1] and Centre for DNA Fingerprinting and Diagnostics, Hyderabad 500076,[2] India

Received 2 July 2001/Accepted 5 September 2001

The radiation sensitivity of *Escherichia coli* B was first described more than 50 years ago, and the genetic locus responsible for the trait was subsequently identified as *lon* (encoding Lon protease). We now show that both *E. coli* B and the first reported *E. coli* K-12 *lon* mutant, AB1899, carry IS*186* insertions in opposite orientations at a single site in the *lon* promoter region and that this site represents a natural hot spot for transposition of the insertion sequence (IS) element. Our analysis of deposited sequence data for a number of other IS*186* insertion sites permitted the deductions that (i) the consensus target site sequence for IS*186* transposition is 5′-$(G)_{\geq 4}(N)_{3-6}(C)_{\geq 4}$-3′, (ii) the associated host sequence duplication varies within the range of 6 to 12 bp and encompasses the $N_{(3-6)}$ sequence, and (iii) in a majority of instances, at least one end of the duplication is at the G-N (or N-C) junction. IS*186*-related sequences were absent in closely related bacterium *Salmonella enterica* serovar Typhimurium, indicating that this IS element is a recent acquisition in the evolutionary history of *E. coli*.

The heritability of radiation sensitivity and radiation resistance traits in bacteria was first studied by Evelyn Witkin more than 50 years ago (26, 27), at a time when the genetic basis of inheritance in these organisms was still a matter of debate (14, 26, 27). Her choice of *Escherichia coli* B as the strain for these studies was historically significant, not least because it is the most radiation sensitive of all natural isolates of *E. coli* (14). In 1964, Howard-Flanders et al. (12) reported isolation of the first *E. coli* K-12 *lon* mutant, AB1899, and stated that "as regards sensitivity to radiation, . . . K-12 *lon*$^-$ resemble strains B. . . " That *E. coli* B (as well as its radiation-resistant suppressor derivative *E. coli* B/r [26, 27]) is also a natural *lon* mutant was shown subsequently by Donch and Greenberg (9). Deficiency of the Lon protease (encoded by *lon*) in the B strains is one reason that they have been developed as *E. coli* hosts of choice for overproduction of recombinant heterologous proteins (2, 24, 28).

We now report that the *lon* mutations in both *E. coli* B and AB1899 (*E. coli* K-12) were caused by IS*186* insertions in opposite orientations at a single site within the spacer region of the *lon* promoter and that this site is a natural hot spot for IS*186* transposition. An unusual feature of this insertion sequence (IS) element is that, unlike most other transposons (7, 25), it is associated with a variable length of duplication of the flanking host sequence at the sites of its insertion (5, 23).

* Corresponding author. Mailing address: Centre for DNA Fingerprinting and Diagnostics, ECIL Rd., Nacharam, Hyderabad 500076, India. Phone: 91-40-7155609. Fax: 91-40-7155610. E-mail: shankar@www.cdfd.org.in.

Independent *lon*::IS*186* insertions at a single site in four strains. We showed earlier (21) that K-12 strain GJ1823 had suffered a spontaneous and unselected IS*186* insertion at the same site in the spacer region of the *lon* promoter as that identified by Ignatov and Chistyakova (13) on a plasmid derivative (pBLI) with the cloned *lon* gene. The genotype description for strain AB1899 at the *E. coli* Genetic Stock Center website (http://cgsc.biology.yale.edu) indicates, as a personal communication from D. A. Vlazny and C. W. Hill, that the *lon-1* mutation in the strain is caused by IS*186* insertion although its position is not specified.

These facts provided us the rationale to compare, at the molecular level, the features of the *lon* mutation in strains AB1899, B/r, and GJ1823 with that published earlier for the plasmid pBLI by Ignatov and Chistyakova (13). We initially performed PCRs using pairwise combinations of the following three primers, designated A, B, and C, respectively: 5′-TGAC CAAGCAGTATCAGG-3′, 5′-AAGATCGTTTACACCCGG CT-3′, and 5′-GAGAAGCTGATTATATCGT-3′. Primers A and B were designed to amplify in PCR the region from −392 to +277 of *lon* (relative to the start site of *lon* transcription, taken as +1). Primer C is complementary to a site within the 1.34-kb IS*186* element, situated 0.6 kb away from one of its ends. With primer pair A and B, the PCR product obtained with genomic DNA of wild-type *E. coli* strain MG1655 as the template was 1.3 kb smaller than that obtained with genomic DNA from each of the *lon* mutants AB1899, GJ1823, and B/r (Fig. 1). With the DNA templates from the three *lon* mutant strains, primer pair A and C on the one hand and pair B and C on the other behaved mutually exclusively, in that a PCR amplification product was obtained only with the latter for

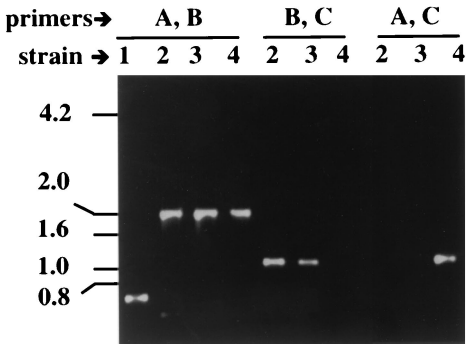

FIG. 1. PCR analysis of *lon* locus in various *E. coli* strains. Primers A and B (corresponding to two sites in *lon*) and primer C (corresponding to an internal site in IS*186*) were used for PCR in the pairwise combinations indicated with genomic DNA templates of the following strains: 1, MG1655 (K-12 wild type); 2, GJ1823 (K-12 *lon*); 3, AB1899 (K-12 *lon*); 4, B/r (*lon*). Methods for PCR and electrophoresis on 0.9% agarose gel were as described previously (22). At the left are shown the positions of DNA markers (sizes in kilobases). Estimated sizes of the observed PCR products were as follows: MG1655 (primers A and B), 0.7 kb; all *lon* mutants (primers A and B), 2.0 kb; AB1899 and GJ1823 (primers B and C), 1.0 kb; *E. coli* B/r (primers A and C), 1.1 kb.

strains AB1899 and GJ1823 and only with the former for strain B/r (Fig. 1). From these findings and from the calculated sizes of the PCR products, one could conclude that all three strains have suffered IS*186* insertion mutations in the vicinity of the *lon* promoter (data not shown) and that the orientation of IS*186* in strains AB1899 and GJ1823 was opposite to that in *E. coli* B/r.

DNA sequence analysis of the PCR products obtained above allowed the precise determination of both the site of *lon*::IS*186* insertion and the extent of the flanking target sequence duplication for strains AB1899, GJ1823, and B/r. These data are

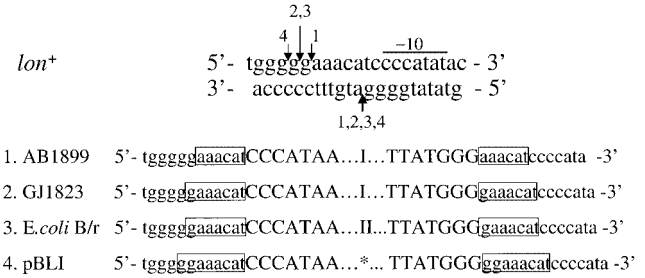

FIG. 2. Molecular identities of various *lon*:: IS*186* insertions. Shown on top is the (double-stranded) sequence of the relevant region of the wild-type *lon*$^+$ locus (4), and the −10 motif in its promoter region is overlined. Beneath are given (lines 1 through 4) the *lon* locus sequences determined in this study (using an automated DNA sequencer and associated protocols) for K-12 strains AB1899 and GJ1823 and for *E. coli* B/r, as well as that reported earlier (13) for plasmid pBLI. In each case, the IS*186* sequence (uppercase) is denoted in an abbreviated form by a few nucleotides from the inverted repeat at either end and by the orientation of IS insertion (marked, respectively, as I or II depending on whether the *Pst*I site in IS*186* is to the left or to the right of the *Bam*HI site within the element); the orientation in pBLI (*) was not reported (13). The IS*186* sequence itself is represented as having a 25-bp pair of inverted repeats as its ends (15). The flanking host sequence duplications are boxed. Arrows, postulated pairs of staggered endonucleolytic cleavage sites (1 through 4, as above) in the target to generate the four different IS*186* insertions.

TABLE 1. Target site sequences for IS*186* insertion

Accession no.[a]	Target site sequence (5′–3′)[b]
AL512784	AGGGGGCG ↓ GGGGCAGCACCC ↓ CCCTCC
AL392173	TGTGGGG ↓ TTGGTGCCCC ↓ CGCA
AL049646	GGCG ↓ GGGGCTG ↓ CCCCCACC
AL080249	AGGTG ↓ GGGGGCAGCCCC ↓ CGTCC
AC007999	CGGGG ↓ GGCTGA ↓ CCCCCT
AL390241	GGTGGGGGG ↓ CAGCCCCC ↓ GCCCGG
AL355498	GGCG ↓ GGGAACATCCC ↓ ACCCC
AL445189	GGGGGAG ↓ GGGGTATAA ↓ CCCCACCCC
AC006487	CG ↓ GGGGGCGG ↓ CCCCCG
AC004448	GGCG ↓ GGGGCTGA ↓ CCCCCACC
Z97985	CCGGGCG ↓ GGGGGCTGA ↓ CCCCCCCCACC
AC006137	CGCG ↓ GGGCTCCG ↓ CCCCCGCC
AF211971	TGGGGGG ↓ GTCAGCC ↓ CCCCGCC
AC074331	TGGGG ↓ GGCAGC ↓ CCCCGCC
Z78058	GGGCGGGG ↓ GTAGTGAC ↓ CCCGCC
AE000112[c]	AGTGGCGGGG ↓ ATCACT ↓ CCCCGCCGTT
AE000163[c]	TTGTGCGGGG ↓ AGTAAT ↓ CCCCGCATCC
AE000327[c]	AAAGCCGGGG ↓ ATAATT ↓ CCCCGGTTTT

[a] Accession numbers are taken from entries in the GenBank sequence database.

[b] Vertical arrows demarcate the host sequence that underwent duplication following IS*186* transposition. The runs of Gs and Cs which are postulated to define the consensus for IS*186* insertion (see text) are underlined. It has been assumed that the IS*186* element carries a pair of 25-bp inverted repeats at its ends (15).

[c] Entry pertaining to the native IS*186* insertions in the wild-type *E. coli* K-12 chromosome.

compared in Fig. 2 with the findings for the *lon*::IS*186* mutation on plasmid pBLI reported earlier (13). The results indicate that, although all four insertions occurred at a single site in the spacer region of the *lon* promoter (given that the sequences reading from the left to the start of IS*186* in all of them are identical), each insertion is molecularly distinct from the other three in terms of IS*186* orientation or length of target sequence duplication or both. The inferred pair of sites of staggered endonucleolytic cleavage of the *lon* target sequence during the process of IS*186* transposition in each of the four instances are depicted in Fig. 2.

Copy number of IS*186* in different strains. Whole-genome sequence analysis of wild-type *E. coli* K-12 strain MG1655 has shown that there are three copies of IS*186* in its chromosome (4) (Table 1), and a previous Southern blot hybridization study came to the same conclusion (15). The latter study also reported that *E. coli* B has the same IS*186* copy number as K-12. Our present results (that *E. coli* B is a *lon*::IS*186* mutant) therefore raised the possibility that IS*186* transposition into *lon* may have occurred by a cut-and-paste mechanism (7) from another chromosomal site in the B strain, followed by the rejoining (that is, recircularization) of the two ends of the broken donor backbone at the latter site. However, such rejoining is considered to be very rare in bacteria (7).

To address this issue further, we examined the chromosomal IS*186* copy number in MG1655, K-12 *lon*::IS*186* mutants AB1899 and GJ1823, and *E. coli* B derivatives B/r and BL21 (24) by Southern blot hybridization. A radiolabeled *Bam*HI-*Pst*I fragment of IS*186* was used as probe for hybridization with genomic DNA preparations of the various strains after digestion with enzyme *Eco*RV, *Bgl*I, or *Pvu*II, none of which cuts within the IS*186* element. The results, shown in Fig. 3, clearly demonstrate that MG1655 has three copies of IS*186* and that

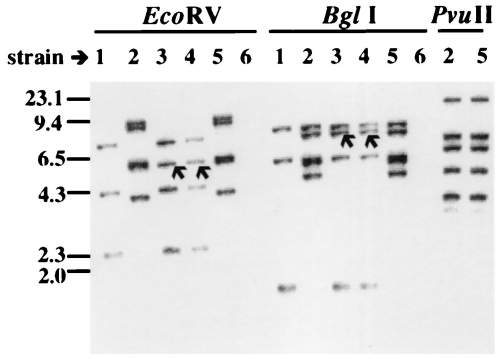

FIG. 3. Southern blot hybridization of genomic-DNA preparations from various strains using the IS*186* probe. Experimental methods were as described previously (22). The DNA preparations were digested with *Eco*RV, *Bgl*I, or *Pvu*II as indicated and subjected to Southern blot hybridization with a ^{32}P-labeled *Pst*I-*Bam*HI internal fragment of IS*186* (obtained from plasmid pHYD138 [21]), followed by autoradiography. Strains employed: 1, MG1655 (K-12 wild type); 2, B/r; 3, AB1899 (K-12 *lon*); 4, GJ1823 (K-12 *lon*); 5, BL21 (*E. coli* B); 6, *S. enterica* LT2. At the left are shown the positions of migration of DNA markers (sizes in kilobases). Arrows, *lon*-specific bands in the lanes corresponding to AB1899 and GJ1823; their sizes (5.6 [*Eco*RV] and 8.4 kb [*Bgl*I]) are consistent with those expected from IS*186* insertion in the cognate *lon*-bearing restriction fragments of wild-type K-12 (4, 20).

K-12 *lon*::IS*186* derivatives AB1899 and GJ1823 each have an additional copy which correlates with the insertion in *lon*. The two *E. coli* B derivatives each have five copies of IS*186*, the most plausible explanation for which is that they represent insertions in the loci homologous to the three of wild-type K-12, with additional copies in *lon* and near *hokX* (16). In another study, Birkenbihl and Vielmetter (3) also concluded that an IS*186* insertion into a new locus in a K-12 strain represented an additional (fourth) copy of the IS element.

In the Southern blot hybridization experiment (Fig. 3), we also included bacterial genomic DNA from strain LT2 of closely related organism *Salmonella enterica* serovar Typhimurium and observed no signal for the IS*186* probe. Our finding is consistent with the fact that IS*186* sequences have not so far been detected in the unfinished genome sequence of *S. enterica* (http://www.tigr.org). IS*186* has also not been detected in the genome of *E. coli* O157:H57 (10, 17) or in any of the other sequenced bacterial genomes, indicating that this IS element, much like others such as IS*2*, IS*4*, and IS*30* (8), may be a recent acquisition in the evolutionary history of *E. coli*. The findings of Pedersen and Gerdes (16) suggest that IS*186* is present in *E. coli* C and also in not less than 44 of the 72 natural *E. coli* isolates of the ECOR collection.

Consensus sequence for IS*186* transposition. Different IS elements are known to vary in the extent to which they exhibit target site selectivity for transposition (7, 25). A consensus sequence for IS*186* transposition has not previously been determined, although it is known that the insertions occur in G-C rich target sequences (5, 15, 23, 25). Because many eukaryotic sequences cloned in *E. coli* hosts are G-C rich, several instances of IS*186* insertions in them are identified in the GenBank database. We analyzed all the deposited entries to deduce that sequence 5′-$(G)_{\geq 4}(N)_{3-6}(C)_{\geq 4}$-3′ is the consensus target recognition site for IS*186* transposition (Table 1). The *lon* hot spot fits the consensus, as do all but one of the entries listed in Table 1. Our analysis also indicates that, even as its length varies between 6 and 12 bp, the associated host sequence duplication always encompasses the $(N)_{3-6}$ sequence of the target recognition site. Furthermore, in more than half of the instances examined at least one end of the duplication (that is, at least one site of endonucleolytic cleavage during transposition) is at the G-N (or N-C) junction (Table 1).

These findings may offer approaches to study the mechanism of IS*186* transposition, of which at present little is known. Our analysis suggests that, in IS*186* transposition, there is an initial specificity in target sequence recognition by transposase but that the sites where subsequent endonucleolytic cleavages occur are less constrained; such a phenomenon invites comparisons with both the mechanism of type I restriction endonuclease action (18) and that of rearrangements of the immunoglobulin genes mediated by the RAG1 and RAG2 proteins, where a preference for G-C-rich sequences at the target sites (11) and the generation of duplications of variable length (1) have also been observed.

It has been suggested that IS*186* belongs to the IS*4* family of insertion elements (5, 15, 19, 25), but the family as currently defined includes members that generate variable (IS*4* and IS*186*) as well as fixed lengths of target site duplications. The length of the duplications associated with transposition of IS*50* (Tn*5*), which is another member of the IS*4* family, is usually 9 bp but rarely could also be 10 bp (6). Finally, IS*1* is also known to generate host sequence duplications of various lengths (25) but is not a member of the IS*4* family.

We thank Mary Berlyn for bacterial strains and acknowledge the assistance of Mehar Sultana and N. Nagesh with oligonucleotide synthesis and DNA sequencing, respectively. We also thank one of the anonymous reviewers for suggesting the analogy between IS*186* transposition on the one hand and restriction endonuclease action and antibody gene rearrangement on the other.

J. G. is an honorary faculty member of the Jawaharlal Nehru Centre for Advanced Scientific Research.

REFERENCES

1. **Agrawal, A., Q. M. Eastman, and D. G. Schatz.** 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature **394:**744–751.
2. **Bhandari, P., and J. Gowrishankar.** 1997. An *Escherichia coli* host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer. J. Bacteriol. **179:**4403–4406.
3. **Birkenbihl, R. P., and W. Vielmetter.** 1991. Completion of the IS map in *E. coli*: IS*186* positions on the *E. coli* K12 chromosome. Mol. Gen. Genet. **226:**318–320.
4. **Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1462.
5. **Chong, P., I. Hui, T. Loo, and S. Gillam.** 1985. Structural analysis of a new GC-specific insertion element IS*186*. FEBS Lett. **192:**47–52.
6. **Chu, C. C., and A. J. Clark.** 1989. A 10- rather than 9-bp duplication associated with insertion of Tn*5* in *Escherichia coli* K-12. Plasmid **22:**260–264.
7. **Craig, N. L.** 1996. Transposition, p. 2339–2362. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.
8. **Deonier, R. C.** 1996. Native insertion sequence elements: locations, distributions, and sequence relationships, p. 2000–2011. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.

9. **Donch, J., and J. Greenberg.** 1968. Ultraviolet sensitivity gene of *Escherichia coli* B. J. Bacteriol. **95:**1555–1559.
10. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C.-G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8**(Suppl.)**:**47–52.
11. **Hiom, K., M. Melek, and M. Gellert.** 1998. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell **94:**463–470.
12. **Howard-Flanders, P., E. Simson, and L. Theriot.** 1964. A locus that controls filament formation and sensitivity to radiation in Escherichia coli K-12. Genetics **49:**237–246.
13. **Ignatov, K. B., and L. G. Chistyakova.** 1997. Integration of the insertion element IS186 into the −10 region of the heat shock promoter of the *Escherichia coli lon* gene. Bioorg. Khim. **23:**18–20. (In Russian.)
14. **Kelner, A., W. D. Bellamy, G. E. Stapleton, and M. R. Zelle.** 1955. Symposium on radiation effects on cells and bacteria. Bacteriol. Rev. **19:**22–44.
15. **Kothary, R. K., D. Jones, and E. P. M. Candido.** 1985. IS*186*: an *Escherichia coli* insertion element isolated from a cDNA library. J. Bacteriol. **164:**957–959.
16. **Pedersen, K., and K. Gerdes.** 1999. Multiple *hok* genes on the chromosome of *Escherichia coli*. Mol. Microbiol. **32:**1090–1102.
17. **Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.
18. **Redaschi, N., and T. A. Bickle.** 1996. DNA restriction and modification systems, p. 773–781. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. American Society for Microbiology, Washington, D.C.
19. **Rezsöhazy, R., B. Hallet, J. Delcour, and J. Mahillon.** 1993. The IS*4* family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. **9:**1283–1295.
20. **Rudd, K. E.** 1998. Linkage map of *Escherichia coli* K-12, edition 10: the physical map. Microbiol. Mol. Biol. Rev. **62:**985–1019.
21. **SaiSree, L., M. Reddy, and J. Gowrishankar.** 2000. *lon* incompatibility associated with mutations causing SOS induction: null *uvrD* alleles induce an SOS response in *Escherichia coli*. J. Bacteriol. **182**:3151–3157.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Sengstag, C., S. Iida, R. Hiestand-Nauer, and W. Arber.** 1986. Terminal inverted repeats of prokaryotic transposable element IS*186* which can generate duplications of variable length at an identical target sequence. Gene **49:**153–156.
24. **Studier, F. W., and B. A. Moffatt.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189:**113–130.
25. **Syvanen, M.** 1988. Bacterial insertion sequences, p. 331–356. *In* R. Kucherlapati and G. R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.
26. **Witkin, E. M.** 1946. Inherited differences in sensitivity to radiation in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **32:**59–68.
27. **Witkin, E. M.** 1947. Genetics of resistance to radiation in *Escherichia coli*. Genetics **32:**221–248.
28. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA **80:**1194–1198.