

Nucleotide Sequence and Functional Map of pE194, a Plasmid That Specifies Inducible Resistance to Macrolide, Lincosamide, and Streptogramin Type B Antibiotics

SUEHARU HORINOUCHI† AND BERNARD WEISBLUM*

Pharmacology Department, University of Wisconsin Medical School, Madison, Wisconsin 53706

Received 16 July 1981/Accepted 13 January 1982

pE194 is a small plasmid (isolated originally in *Staphylococcus aureus*) which confers erythromycin-inducible resistance to macrolide, lincosamide, and streptogramin type B (MLS) antibiotics. The nucleotide sequence of pE194 contains 3,728 base pairs (bp), corresponding to a molecular mass of 2.4 million daltons. By means of site-specific cleavage with restriction endonucleases and cloning resultant fragments, determinants of the two major biological functions of pE194, i.e., inducible MLS resistance and replication, could be localized and assigned to specific sequences in the plasmid. Restriction endonuclease *TaqI* cut pE194 at three sites. *TaqI* fragment A (1,443 bp) contained the determinant for inducible MLS resistance, whereas *TaqI* fragment B (1,354 bp) contained a determinant necessary for plasmid replication. Regulatory mutations resulting in constitutive expression of MLS resistance mapped in *TaqI* fragment A, whereas a mutation associated with elevated plasmid copy number was mapped in *TaqI* fragment B. Also mapping in *TaqI* fragment B was a plasmid replication determinant comprising two sets of inverted complementary repeat sequences, one of which spanned 124 bp and was adjacent to a second smaller set which was rich in guanine and cytosine residues. pE194 contained six open reading frames which were theoretically capable of coding for proteins with maximum molecular masses as follows (in daltons): A, 48,300; B, 29,200; C, 14,000; D, 13,900; E, 12,600; and F, 2,700. Insertion of plasmid pBR322 into the single *PstI* site located in frame A of pE194 resulted in a composite plasmid which could replicate in both *Bacillus subtilis* and *Escherichia coli*, suggesting that an intact polypeptide A is dispensable for both replication of pE194 and for MLS resistance. Frame B specified inducible MLS resistance, whereas frame F specified the putative peptide associated with the proposed B determinant translational attenuator. The extent to which frames C, D, and E, all contained in *TaqI* fragment B, were translated into polypeptide products is not known; however, a base change in frame E was found in a comparison between the high-copy-number mutant, *cop-6*, and the wild-type strains.

Plasmid pE194 from *Staphylococcus aureus* reported by Iordanescu et al. (13, 14) specifies erythromycin-induced resistance to macrolide, lincosamide, and streptogramin type B (MLS) antibiotics mediated by a specific N⁶-dimethylation of adenine in 23S rRNA, as a consequence of which MLS antibiotics bind to the ribosome with reduced affinity (16, 17). The molecular basis of this resistance phenotype and the regulation of its expression have been the main research theme of our laboratory (30). pE194 is of special interest because one of its subfragments obtained by digestion with *TaqI* restriction endonuclease contains the smallest DNA sequence (1,443 base pairs) capable of specifying

inducible MLS resistance (10, 11), and thus serves as a useful test system for studies at the DNA sequence level. In addition, we have shown how mutants of pE194 with altered control of either MLS resistance or replication, respectively, can be isolated preferentially by the use of MLS antibiotics as selecting agents (31, 32). We shall show below that pE194 replication function can also be mapped both functionally and by DNA sequencing methods.

Ehrlich (4) first showed the feasibility of introducing small *S. aureus* plasmids into *Bacillus subtilis* by transformation. Plasmid pE194 was introduced by Gryczan and Dubnau (5) into *B. subtilis*, making it possible to choose from a large number of mutant bacterial strains in studies of the function of this plasmid. When *B. subtilis* cells carrying wild-type pE194 are grown

† Present address: Department of Agricultural Chemistry, The University of Tokyo, Bunkyo-ku 113, Japan.

in medium containing a noninducing MLS antibiotic, e.g., tylosin or clindamycin, high-copy-number mutants of pE194 are obtained (31). One of the high-copy-number mutants, *cop-6*, was introduced into a minicell-producing strain of *B. subtilis* by Shivakumar et al. (24), who obtained detailed information on pE194-encoded peptides. Some previous reports have emphasized only individual aspects of pE194 structure and function (10, 11, 31). We therefore present here the DNA sequence of this plasmid in its entirety, including data pertinent to the mapping of biological functions and determinants of their expression or regulation.

MATERIALS AND METHODS

Bacterial strains and plasmids. Bacterial strains and plasmids used are listed in Table 1.

Plasmid DNA preparation. *Escherichia coli* and *B. subtilis* plasmids were prepared by the lysozyme-EDTA method as described by Horinouchi et al. (8, 9); plasmids from *S. aureus* were prepared similarly by use of lysostaphin instead of lysozyme as the lytic enzyme. Covalently closed circular DNA was obtained from the lysates by two cycles of ultracentrifugation using the CsCl-ethidium bromide method. The nucleotide sequence of pE194 was determined using DNA from the *cop-6* mutant.

Isolation of mutants. Mutants of pE194 which express MLS resistance constitutively were obtained by plating (inducible) *S. aureus* RN2442 on solid medium containing 10 µg of either clindamycin or tylosin per ml, as described previously (32). Copy-number mutants of pE194 were obtained by selection in the same way except that *B. subtilis* BD170 (pE194) was used as described (31). Plasmid DNA samples from constitutive mutants were analyzed by digestion with *Hinf*I, followed by dephosphorylation, end labeling, and secondary cleavage with *Mbo*I.

Enzymes. The restriction endonucleases as well as T4 polynucleotide kinase and T4 DNA ligase were purchased from New England Biolabs except for *Cla*I, which was purchased from Boehringer-Mannheim Corp. Experimental conditions used were as recommended by the vendor.

DNA sequence studies. The nucleotide sequence of pE194 was determined using the end-label method of Maxam and Gilbert (18) followed by computer analysis of the sequence data using the programs devised by Korn et al. (15). Strand separation was performed by the method of Szalay et al. (27).

RESULTS

A physical map of pE194 showing critical restriction endonuclease cleavage sites and five open reading frames deduced from computer analysis of the sequence are shown in Fig. 1. Pending identification of the polypeptides encoded by pE194 with their respective reading frames, we adopted the convention of referring to open reading frames alphabetically in order of decreasing size and assigning a polypeptide molecular weight as if the entire open reading frame were used. We used the term "open reading frame" to designate a DNA sequence which lacks stop codons when translated in a particular phase. Schematic representations of recombinants that were constructed to elucidate biological functions of pE194 are shown in Fig. 2.

A more detailed physical map of pE194, including additional restriction sites and the extent of sequencing from each of the labeled ends prepared, is shown in Fig. 3. All restriction sites shown were identified in overlapping sequence determinations; about 70% of the pE194 sequence was determined in both strands. The complete nucleotide sequence, including potential polypeptides encoded by open reading frames as well as inverted complementary repeat sequences, are shown in Fig. 4.

Inducible MLS resistance associated with *Taq*I fragment A. Analysis of the pE194 sequence showed that reading frame B is contained wholly within *Taq*I fragment A. Plasmid pC194 was used as a cloning vector to study expression of this fragment because it specifies inducible chloramphenicol resistance as a selectable marker and contains two *Cla*I sites into which *Taq*I ends can be readily ligated, since they share the common four-base sequence, TCGA, and are cleaved at the same position in this sequence. The pC194 DNA preparation was partially digested with *Cla*I to yield a significant portion of unit linear molecules. The complete pC194 DNA sequence, including details pertinent to the map location of the two *Cla*I sites, is presented in the accompanying paper (12). The pE194 DNA preparation was digested to completion using *Taq*I, followed by additional digestion with *Bst*NI and *Hga*I to cut *Taq*I fragments

TABLE 1. Bacterial strains and plasmids

Strain	Plasmid	Source	Reference	Relevant characteristics
<i>S. aureus</i> RN2442	pE194	R. P. Novick	13, 14	Inducible MLS resistance
<i>B. subtilis</i> BD430	pE194	D. Dubnau	5	Inducible MLS resistance
<i>B. subtilis</i> BD476	pE194 <i>cop-6</i>	D. Dubnau	31	High copy number
<i>B. subtilis</i> 168	pC194	D. Dean	4, 13, 14	Chloramphenicol resistance
<i>B. subtilis</i> MI112	pTL11-2	T. Tanaka	28, 29	Trimethoprim resistance, <i>leu</i>
<i>B. subtilis</i> MI112	None known	T. Tanaka	28, 29	<i>hsdR hsdM leuA8 thr-5 arg-15 recE4</i>
<i>E. coli</i> RR1	pBR322	J. Davies	1	Tetracycline and ampicillin resistances

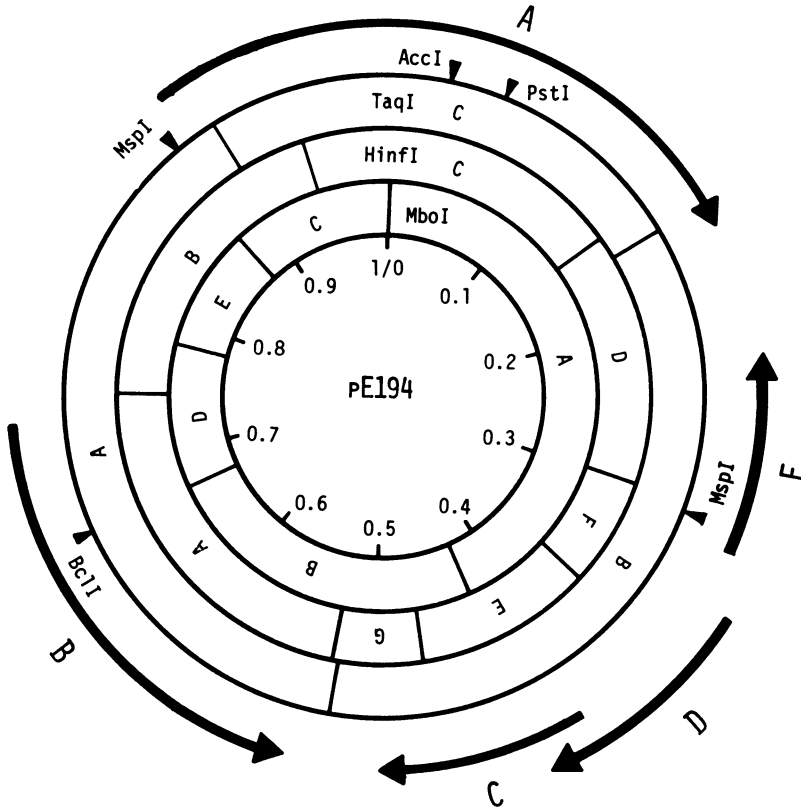


FIG. 1. Physical map of pE194 showing five open reading frames and restriction sites for *MboI*, *HinfI*, *TaqI*, as well as sites for the enzymes *BclI*, *MspI*, *AccI*, and *PstI*. Arrows indicate open reading frames and their respective 5' to 3' orientations. The open reading frames are labeled alphabetically in order of decreasing size. See text for details.

B and C, respectively, and thus the extent to which these fragments might participate in the subsequent ligation step was minimized. The two preparations were mixed in a 1:3 ratio, respectively, followed by ligation and transformation of a competent *B. subtilis* cell preparation by the nutritional step-down procedure of Dubnau and Davidoff-Abelson (3), induction for MLS resistance, as described below, and selection of transformants on enriched medium containing chloramphenicol (5 $\mu\text{g/ml}$) plus erythromycin (10 $\mu\text{g/ml}$). Transformant clones obtained in this way were tested by the cleared lysate method for the presence of a single plasmid of appropriate molecular weight, followed by digestion with restriction endonucleases and analytical polyacrylamide gel electrophoresis. The structure of pHW1 was checked by digestion with *TaqI* followed by polyacrylamide gel electrophoresis, which demonstrated the presence of the expected five pC194 *TaqI* fragments plus an additional fragment with mobility indistinguishable from that of pE194 *TaqI* fragment A.

Examination of the resistance phenotype of the transformant carrying the pE194 *TaqI* fragment A using antibiotic disks revealed that resistance was inducible; *TaqI* fragment A therefore contains both the MLS resistance determinant and information required for regulation of its expression.

Insertional inactivation of the MLS resistance determinant. A cloned trimethoprim-resistant dihydrofolate reductase determinant of *B. subtilis* was obtained by Tanaka (28, 29) by selection for resistant *B. subtilis* after ligation of digested DNA from the mutant using plasmid pTL11 as cloning vehicle and selection of trimethoprim-resistant transformants. Digestion of DNA from the recombinant plasmid pTL11-2 with *BglII* yielded a DNA fragment (1.8 megadaltons) carrying an intact trimethoprim resistance determinant capable of ligation into the single *BclI* site of pE194. pE194 digested with *BclI* and pTL11-2 digested with *BglII* were mixed in a 1:1 ratio and ligated, and the resultant reaction mixture was used to transform competent *B. subtilis* MI12

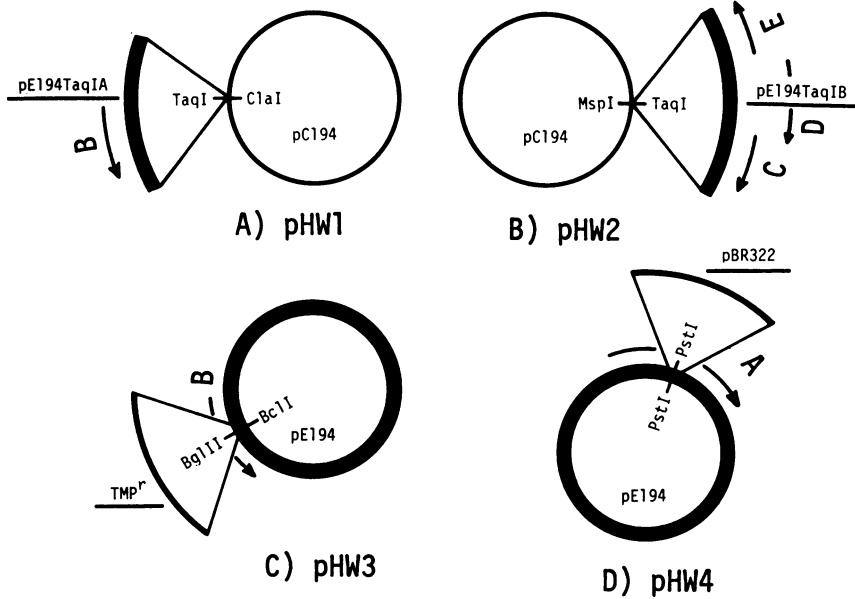


FIG. 2. Schematic diagrams summarizing plasmid constructions. These include (A) pHW1, obtained by insertion of pE194 *TaqI* fragment A into a *ClaI* site of pC194, with formation of a recombinant plasmid that confers inducible MLS plus chloramphenicol resistance; (B) pHW2, obtained by insertion of pE194 *TaqI* fragment B into the *MspI* site of pC194, formation of a plasmid in which pC194 replication function is destroyed but restored by a replication determinant present in pE194 *TaqI* fragment B; (C) pHW3, obtained by insertion of the trimethoprim (TMP)-resistant dihydrofolate reductase determinant (from *B. subtilis*) into the *BclI* site of pE194; and (D) pHW4, obtained by ligation of pE194 and pBR322 along their respective *PstI* sites which interrupts continuity of the pE194 reading A frame sequence.

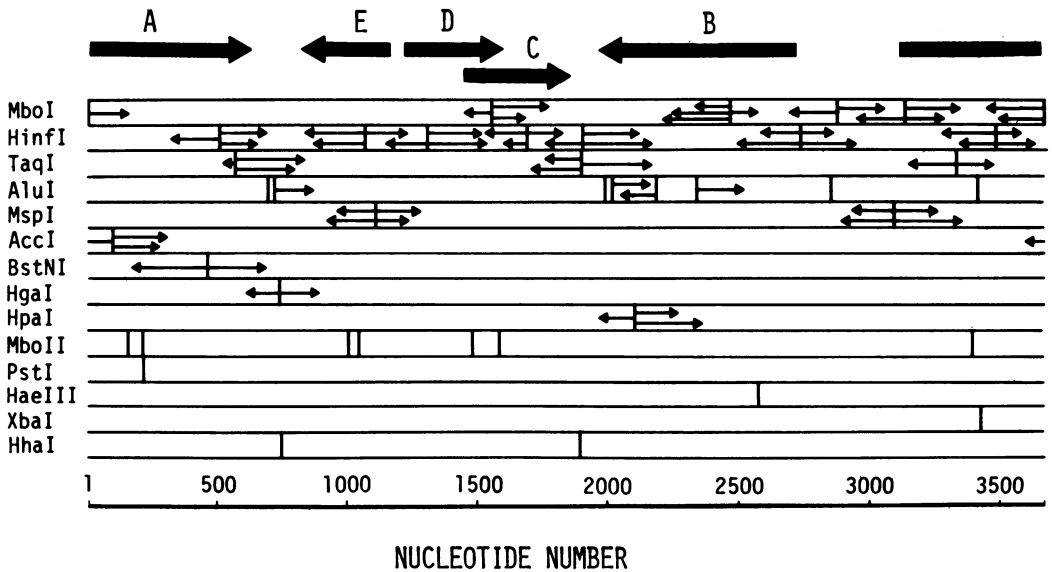


FIG. 3. Sequencing strategy for pE194. The sites used for end labeling of DNA fragments obtained by digestion with restriction endonucleases are indicated with the thinner arrows whose length and direction indicate the extent of sequence determination from these sites. The open reading frames (thicker arrows) are labeled as in Fig. 1.

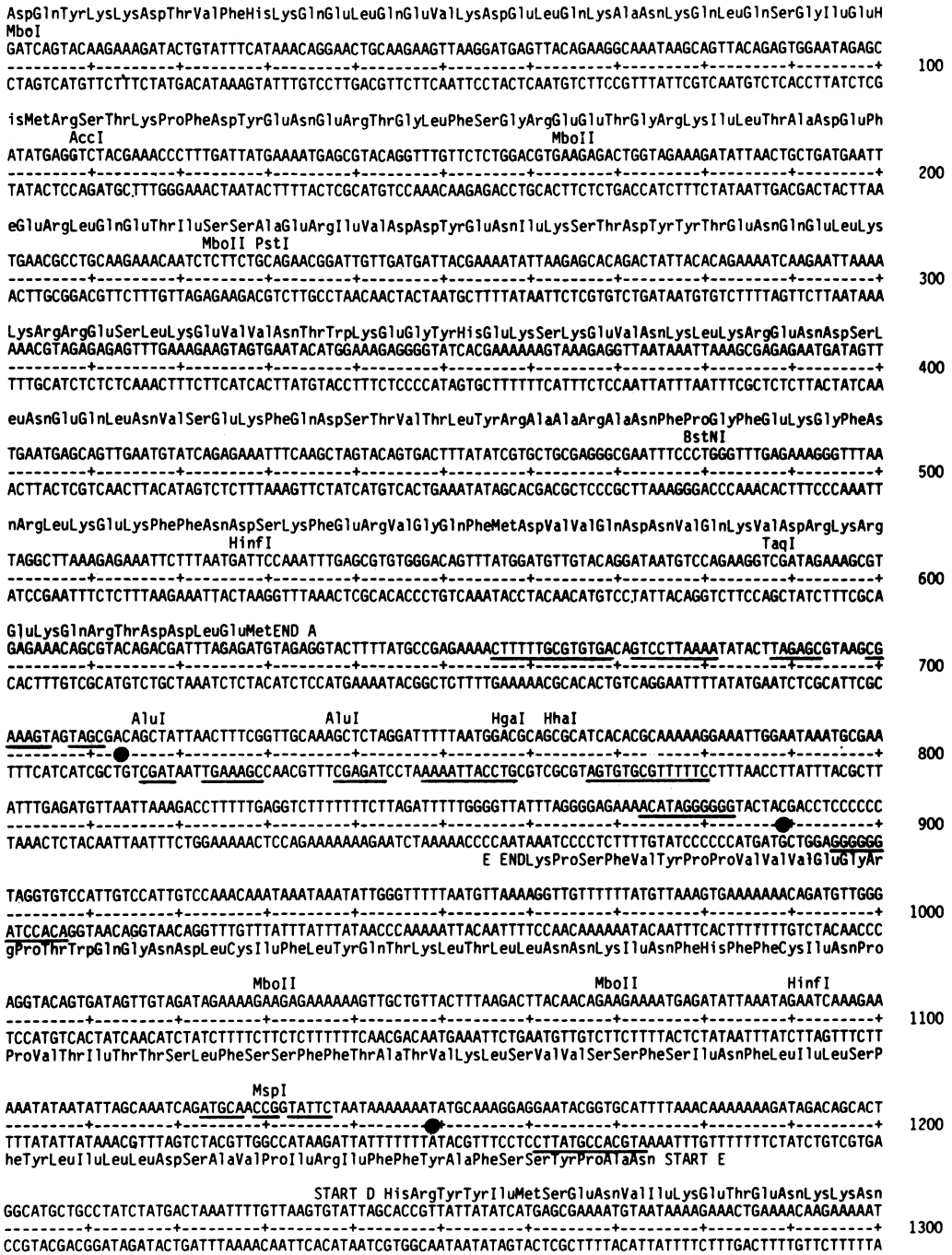


FIG. 4. The nucleotide sequence of pE194, numbered from the MboI C-A junction as reference point, is shown together with potential amino acid sequences specified by open reading frames A, B, C, D, E, and F deduced from the DNA sequence.

SerArgGlyArgAsnTrpThrPheValLeuTyrProGluSerAlaLysAlaGluTrpLeuGluTyrLeuLysGluLeuHisIluGlnPheValValSerP
HinfI
TCAAGAGGACGTAATTGGACATTTGTTTTATATCCAGAATCAGCAAAAGCCGAGTGGTTAGAGTATTTAAAAGAGTTACACATTTCAATTTGTAGTGTCTC
AGTTCTCCTGCATTAACCTGTAACAATAATATAGGCTTAGTCGTTTTCCGGCTCACCATCTCATAAATTTTCTCAATGTGTAAGTTAACATCACAGAG
1400

roLeuHisAspArgAspThrAspThrGluGlyArgMetLysLysGluHisTyrHisIluLeuValMetTyrGluGlyAsnLysSerTyrGluGlnIluLys
CATTACATGATAGGGATACTGATACAGAAGGTAGGATGAAAAAGAGCATTATCATATTTCTAGTGATGTATAGGGTAATAAATCTTATGAACAGATAAA
GTAATGTACTATCCCTATGACTATGCTTCCATCTACTTTTTCTCGTAATAGTATAAGATCACTACATCTCCCATTTATTAGAATACTGTCTATTT
START
C LeuThrGluGluLeuAsnAlaThrIluProGlnIluAlaGlySerValLysGlyLeuValArgTyrMetLeuHisMetAspAspProAsnLysPh
sIluIluAsnArgArgIluGluCysAspTyrSerAlaAspCysArgLysCysGluArgSerCysGluIluTyrAlaSerHisGluArgSerEND D
MboI MboI
ATAAATTAACAGAAGAATTGAATGCGACTATTCGCGAGATTGCGAGAGTGTAAAGGCTTCTGTGAGATATATGCTTCACATGGACGATCCTAATAAAT
TTATAATTGCTTCTTAACCTACGCTGATAAGGCGCTAACGCTCTCACACTTCCAGAACACTCTATATACGAAGTGTACCTGCTAGGATTATTTAA
1600

eLysTyrGlnLysGluAspMetIluValTyrGlyGlyValAspValAspGluLeuLeuLysLysThrThrThrAspArgTyrLysLeuIluLysGluMet
MboI
TAAATATCAAAAAGAAGATATGATAGTTTATGGCGGTGATAGTGTGATGAATTTAAAGAAAACAACACAGATAGATATAAATTAATTAAGAAATG
ATTTATAGTTTTCTTCTATACTATCAATACCGCCACATCTACAACACTTAATAATTTCTTTGTTGTTGCTATCTATTTAATTAATTTCTTTAC
1700

IluGluPheIluAspGluGlnGlyIluValGluPheLysSerLeuMetAspTyrAlaMetLysPheLysPheAspAspTrpPheProLeuLeuCysAspA
HinfI
ATTGAGTTTATTGATGAACAAGGAATCGTAGAATTAAGAGTTTAAATGGATTGCAATGAAGTTAAATTTGATGATTGGTCCCGCTTTTATGTGATA
TAAC TCAATAACTACTTGTTCCTTAGCATCTTAATTTCTCAAAATACCTAATACGTTACTTCAAAATTAACACTACTAACCAAGGGCGAAAATACACTAT
1800

snSerAlaTyrValIluGlnGluTyrIluLysSerAsnArgTyrLysSerAspArgEND C
ACTCGGCGTATGTTATTCAGAATATATAAATCAAAATCGGTATAAATCTGACCGATAGATTTGAAATTTAGGTGTCACAAGACACTTTTTTCGCACC
TGAGCGCATACAATAAGTCTTATATATTTAGTTTAGCCATATTTAGACTGGCTATGTAAAACCTTAATCCACAGTGTCTGTGAGAAAAAGCGTGG
1900

HhaI (TaqI) ClaIHinfI
AGCGAAAATCGTTTAAAGCCGACTGCGCAAAAGACATAATCGATTCAAAAAATAGGCACACGAAAAACAAGTTAAGGGATGCAGTTTATGCATCCCTT
TCGCTTTTGACCAAAATCGCGTACGCGTTTTCTGATTAGCTAAGTGTTTTTATCCGTGTCTTTTTGTTCAATTCCTACGTCAAATACGTAGGGAA
2000

AluI AluI
AACTACTTATTAATAATTTATAGCTATTGAAAAGAGATAAGAATGTTCAAAGCTAATATTGTTTAAATCGTCAATTCCTGCATGTTTTAAGGAATTG
TTGAATGAATAATTTATTAATATCGATAACTTTTTCTATTCTTAACAAGTTTCGATTATAACAATTTAGCAGTTAAGGACGTACAAAATTCCTTAAC
B ENDLysAsnPheLeuLysTyrSerAsnPheLeuSerLeuPheGlnGluPheSerIluAsnAsnLeuAspAspIluGlyAlaHisLysLeuSerAsnA
2100

HpaI
TTAAATGATTTTTGTAAATATTTCTGTATTCTTTGTTAAACCCATTTTCATAACGAAATAATATACTTTTGTTTATCTTTGTGTGATATTCTTGATT
AATTTAACTAAAAACATTTATAAAGAACATAAGAACAATTTGGGTAAAGTATTGCTTTTATTAATATGAAAACAAATAGAAACACACTATAAGAACATA
snPheGlnAsnLysThrPheIluLysLysTyrGluLysAsnValTrpLysMetValPheTyrAsnTyrLysGlnLysAspLysHisSerIluArgSerLys
2200

AluI
TTTTCTACTTAATCTGATAAGTGAGCTATTCACTTTAGGTTTAGGATGAAAAATTTCTCTTGAACCATACTTAATATAGAAATATCAACTCTGCCAT
AAAAAGATGAATTAGACTATTCCTACTGATAAGTGAATCCAAATCTACTTTTATAAGAGAACCCTGGTATGAATTATCTTTATAGTTGAAAGACGGTA
sLysArgSerLeuArgIluLeuSerSerAsnValLysProLysProHisPheTyrGluArgProValMetSerLeuIluSerIluAspValGluAlaMet
2300

AluI
TAAAGTAAATGCCAATGAGCGTTTTGTATTAAATAATCTTTAGCAAACCCGATTTCACGATTAATAAATCTCATTAGCTATACTACAAAAACAATT
ATTTTCATTACGGTTACTCGCAAAACATAAATATTAGAAAACTGTTGGGCATAAGGTGCTAATTTATTTAGAGTAATCGATATGATAGTTTTGTTAA
LeuLeuLeuAlaLeuSerArgLysThrAsnLeuLeuArgLysAlaPheGlyTyrGluValIluLeuTyrIluGluAsnAlaIluSerAspPheValIluL
2400

TTGCGTATTATATCCGACTTATGTTATAAGGTATATTACCATATATTTTATAGGATGGTTTTAGGAAATTTAAACTGCAATATATCTTGTAAAA
AACGCATAATAGGCATGAATACAATATCCATATAATGGTATATAAATATCCTAACCAAAATCCTTTAAATTTGACGTTATATAGGAACAATTTT
ysArgIluIluAspThrSerIluAsnTyrProIluAsnGlyTyrIluLysTyrSerGlnAsnLysProPheLysPheGlnLeuIluAspLysAsnLeuVa
2500

FIG. 4—Continued

BclI (MboI)
 CTTGGAATATTCGATCAACAAGTTATTTCTGTAGTTTGCATAATTTATGGTCTATTTCAATGGCAGTTACGAAATACACCTCTTTACTAATTC
 GAACCTTTAATAGCACTAGTTGTTCAAATAAAGACATCAAACGTTAATAATACCAGATAAAGTTACCGTCAATGCTTTAATGGAGAAATGATTAAG
 1G1nPheAsnAspHisAspValLeuLysAsnGluThrThrLysCysLeuLysHisAspIluGluIluAlaThrValPheAsnCysArgLysValLeuGlu 2600

HaeIII
 AAGGGTAAATGGCCTTTTCTGAGCCGATTTCAAAGATATTATCATGTTCAATTAATCTTATATTGTCATTATTTATCTATATTATGTTTGAAGTA
 TTCCCATTTTACCGAAAAGGACTCGGCTAAAGTTCTATAATAGTACAAGTAAATAGAAATAAAACAGTAATAAATAGATATAATACAAAACCTCAT
 LeuThrPheHisGlyLysGlySerGlyIluGluPheIluAsnAspHisGluAsnLeuArgIluAsnThrMetIluLysAspIluAsnHisLysSerThrI 2700

HinfI
 ATAAAGTTTGGACTGTGTTTATATTTTCTCGTTCATTATAACCCTCTTAATTTGGTTATATGAATTTTGGCTTATAACGATTCAATATAACCCTTA
 TATTTCAAACCTGACACAAAATATAAAAGAGCAAGTAAATTTGGGAGAAATTAACAATATACTTAAAACGAATAATTGCTAAGTAAATTTGGTGAAT
 IuPheAsnGlnSerHisLysIluAsnLysGluAsnMet START B F END 2800

(AluI)SacI
 TTTTTGTTGGTTGATAATGAACGTGCTGATTACAAAATACTAAAAATGCCATATTTTTCTCTCTTATAAAAATTAGTATAATTATAGCAGGAGCT
 AAAAAACAACCAACTATTACTTGACACGACTAATGTTTTATGATTTTTACGGGTATAAAAAAGGAGGAATTTAATCATATTAATATCGTCTCGA
 LysLysAsnProGlnTyrHisValThrSerIluValPheIluSerPheIluGlyMet START F 2900

MboI
 CTGATAAATATGAACATGATGAGTGATCGTTAAATTTATACTGCAATCGGATGCGATTATGAATAAAAGATATGAGAGATTTATCTAATTTCTTTTTTC
 GACTATTTATACTTGTACTACTCACTAGCAATTTAAATATGACGTTAGCCTACGCTAATAACTTATTTCTATACTCTCTAAATAGATTAAGAAAAAAG
 TTGTAATAAAAGAAAGTCTTAAAGGTTTTATAGTTTGGTCGTAGAGCACAGGTTAACGACTTAATACGAAGTAAATAAGTCTAGTGTGTAGACT
 AACATTTTTTTCTTCAAGAATTTCCAAAATATCAAACCAGCATCTCGTGTGCCAAATGCTGAATTAATGCTTCATTATTCCAGATCACAAATCTGA 3000

START A MetSerHisSerIluLeuArgValAlaArgValLysGlySerSerAsnThr
 MspI MboI
 TTATGAAATCTATATACGTTTATATATATTTATTATCCGGAGGTGATGATGCTCATTCAATTTGAGGGTCCGACAGTTAAAGGATCAAGTAAACA
 AATACTTTAGATATATGCAATATATATAAATAATAGGCCCTCCACATCGTACAGAGTAAAGTTAAAACCTCCCAACGGTCTCAATTTCTAGTTCATTATGT 3200

AsnGlyIluGlnArgHisAsnGlnArgGluAsnLysAsnTyrAsnAsnLysAspIluAsnHisGluGluThrTyrLysAsnTyrAspLeuIluAsnAlaG
 AACGGGATACAAGACATAATCAAAGAGAGAATAAAAACCTATAATAATAAAGACATAAATCATGAGGAAACATATAAAAATTTATGATTTGATTAAACGCAC
 TTGCCCTATGTTCTGTATTAGTTTCTCTCTATTTTGTATTTTGTATTTTGTATTTTAACTAAACTAATTTGCGTG 3300

IluAsnIluLysTyrLysAspLysIluAspGluThrIluAspGluAsnTyrSerGlyLysArgLysIluArgSerAspAlaIluArgHisValAspGlyLe
 AAAATATAAGTATAAAGATAAAAATTGATGAAACGATTGATGAGAATTTACAGGAAACGTAATAATTCGGTCAGATGCAATTCGACATGTGGACGGACT
 TTTTATATTTTCATATTTCTATTTTAACTACTTTGCTAACTACTCTTAATAAGTCCCTTTGCATTTTAAGCCAGTCTACGTTAAGCTGTACACCTGCCTGA 3400

uValThrSerAspLysAspPhePheAspAspLeuSerGlyGluGluIluGluArgPhePheLysAspSerLeuGluPheLeuGluAsnGluTyrGlyLeu
 MboI AluI XbaI
 GGTACAAGTGATAAAGATTTCTTTGATGATTTAAGCGGAGAAGAAATAGAACGATTTTAAAGATAGCTTGGAGTTCTAGAAAATGAATACGGTAAAG
 CCAATGTTCACTATTTCTAAAGAACTACTAAATTCGCCCTCTCTTTATCTTGTCTAAAAAATTTCTATCGAACCTCAAAGATCTTTACTTATGCCATTC 3500

GluAsnMetLeuTyrAlaThrValHisLeuAspGluArgValProHisMetHisPheGlyPheValProLeuThrGluAspGlyArgLeuSerAlaLysG
 HinfI
 GAAAATATGCTGTATGCGACTGTCCATCTGGATGAAAGAGTCCACATATGCACCTTTGGTTTTGTCCTTTAAACAGAGGACGGAGATTTGCTGCAAAAAG
 CTTTATACGACATACGCTGACAGGTAGACCTACTTTCTCAGGGTGTATACGTGAAACCAAAACAGGGAAATTTGCTCCTGCCCTCTAACAGACGTTTTCT 3600

IluGlnLeuGlyAsnLysLysAspPheThrGlnLeuGlnAspArgPheAsnGluTyrValAsnGluLysGlyTyrGluLeuGluArgGlyThrSerLysG
 AACAGTTAGGCAACAAGAAAGACTTTACTCAATACAAGATAGATTTAATGAGTATGTGAATGAGAAAGGTTATGAACCTGAAAGAGGCACGCTCAAGA
 TTGTCATCCGTTGTTCTTTCTGAAATGAGTTAATGTTCTATCTAAATTACTCATACACTACTCTTTTCCAATACTTGAACCTTCTCCGTCAGGTTTCT 3700

uValThrGluArgGluHisLysAlaMet
 GGTACAGAACGAGAACATAAAGCGATG
 CCAATGCTCTGCTCTGTATTTTCGCTAC 3728

FIG. 4—Continued

cells. The ligated DNA mixture was mixed with competent cells, and resistant transformants were selected on M9 medium containing trimethoprim (20 $\mu\text{g/ml}$). Trimethoprim-resistant, erythromycin-sensitive transformant clones obtained in this manner were checked for the *leu* phenotype, and several transformant clones with the desired phenotype were further tested by preparation of small-scale cleared lysates as described by Horinouchi et al. (8). The structure of the resultant plasmid pHW3 was checked by digestion with *Mbo*I restriction endonuclease whose site specificity subsumes those of *Bcl*II and *Bg*III. Analysis of the resultant digest by polyacrylamide gel electrophoresis revealed the presence of at least eight new restriction fragments with counterparts in the *Mbo*I digest of pTL11-2 in addition to the five bands expected from digestion of pE194.

Independent evidence that pHW3 actually contains DNA inserted into the *Bcl*II site was obtained by the heteroduplex mapping method (33) using pE194 and pHW3. Both plasmids were digested at their unique *Acc*I sites to form unit linear molecules which were denatured, mixed, and reannealed to form a mixture containing heteroduplexes of the two linear species. The resultant mixture was then analyzed by electron microscopy. From analysis of 37 such heteroduplex molecules, the insertion site was located at $0.65 \pm 1\%$ fractional lengths which would correspond to 2404 ± 37 residues from the *Acc*I site, or at residue 2515 ± 37 residues from the *Mbo*I C-E border chosen arbitrarily as reference origin of pE194. This number agrees well with the expectation deduced from the sequence that the insert should be located at the *Mbo*I site at residue 2539.

Shivakumar et al. (25) reported that induction of MLS resistance was insensitive to rifampin, from which they inferred that induction did not affect transcription directly. Determination of the nucleotide sequence of pE194 *Taq*I fragment A and the peptides encoded by it enabled us (10) to propose an explicit model for the regulation of the MLS resistance phenotype. Gryczan et al. (6) have proposed a similar model.

According to this model, pE194 *Taq*I fragment A contains two open reading frames, F and B, respectively, potentially encoding (i) a 19-amino acid peptide, and (ii) a protein containing 244 amino acids (29,000 daltons) identified as the presumptive adenine methylase which mediates the MLS resistance phenotype. A set of inverted complementary repeat sequences present at the 5' end of frame B, between residues 2711 and 2850, interposed between the putative transcription start site and the 29,000-dalton protein structural gene, has been associated with the control of expression of the 244-amino-acid

polypeptide and constitutes the most noteworthy structural feature of this region. The first set of inverted complementary repeat sequences partially overlaps frame F, the structural gene for the putative "control peptide." In the translation attenuation model, the erythromycin-induced ribosome stall in the course of synthesis of the F frame product leads to disruption of the associated conformation of this control region (10, 11). The model resembles, in part, regulation of gene expression by transcriptional attenuation mechanisms used in the control of amino acid biosynthesis, reviewed by Yanofsky (34), except that ribosome stall results from antibiotic action rather than from amino acid deficiency. Moreover, as a consequence of the redistribution of the inverted complementary repeat sequences, the ribosome loading site (Shine and Dalgarno [22]) for synthesis of the 29,000-dalton protein, normally sequestered, is unmasked together with a part of the structural gene sequence which encodes the first seven amino acids of the protein. Additional details of the promoter for this gene, including the -35, -10, and ribosome loading sites (19-22, 26), have been described (6, 10, 11). These sequences resemble corresponding sequences associated with *E. coli* promoters.

The amino acid sequence of the 29,000-dalton protein, encoded by nucleotides 2738 to 2007, is shown in Fig. 4. The unique *Bcl*II site in pE194 has been localized by both endonuclease cleavage studies and by direct sequence analysis within this region. The location of this *Bcl*II site at residue 2515 is situated 223 nucleotides downstream from the methionine start codon, well into the coding portion of the 29,000-dalton determinant.

Replication function associated with *Taq*I fragment B. A functional criterion for localization of replication determinants involves disruption of replication function by cleavage with restriction endonucleases and restoration by insertion of DNA fragments which replace the lost function. Chang and Cohen (2) have reported that a DNA fragment of pSC101, obtained by digestion with *Hpa*II and containing the replication sequences of this plasmid, could be ligated into the unique *Hpa*II site of pC194, resulting in a plasmid capable of replication in *E. coli* but not in *B. subtilis*. This suggested that the *Hpa*II site of pC194 is located in a region vital for replication. We therefore attempted to clone pE194 fragments into the pC194 *Msp*I site to determine whether ability to replicate in *B. subtilis* might be restored to pC194 cleaved with *Msp*I by insertion of a functionally active equivalent sequence from pE194. Our most informative experiments were performed by using a *Taq*I digest of pE194 as the source of test DNA for

insertion into the *MspI* site of pC194. After transformation, selection for chloramphenicol resistance, and screening for erythromycin-sensitive transformants, a plasmid, pHW2, was obtained from one of the transformant clones which contained pE194 *TaqI* fragment B. The structure of this plasmid was tested and verified by digestion with *HinI* restriction endonuclease, followed by analysis of the resultant digest using polyacrylamide gel electrophoresis. The analysis showed the expected pE194 *HinI* fragments E and F (wholly contained within pE194 *TaqI* fragment B) as well as the expected bridge fragments.

In the nucleotide sequence studies of pC194 DNA presented in the accompanying paper (12), the pC194 *MspI* site is localized in a 12-base-pair guanine-cytosine-rich inverted complementary repeat sequence. Adjacent to this sequence, we noted the presence of an unusually long inverted complementary repeat region spanning 121 bases. Insertion of these two sets of inverted complementary repeat sequences along with an additional 200 nucleotides and the chloramphenicol resistance determinant into pBR322 yielded a composite plasmid, pHW2, capable of autonomous replication in *B. subtilis*. We therefore searched for and found a similar inverted complementary repeat region in *TaqI* fragment B of pE194 consisting of (i) a larger set spanning 124 bases located between residues 656 and 780, and a second smaller set comprising 24 bases between residues 873 and 907 (Fig. 5). The two sets of inverted complementary repeat regions are unique in their respective plasmids. In view of the evidence implicating these two respective regions in plasmid replication, we infer that their unique structural similarity reflects, at least in part, the fact that they serve similar functions in plasmid replication.

In determining the DNA sequence between residues 929 and 1329 in pE194 wild type, we found a single base difference at residue 1013 (A in cop-6 and G in the wild type). This mutation affects the third position of a putative thr codon, ACT, in open frame E of the *cop-6* mutant DNA, and formally would have been derived from a thr codon, ACC. We note the location of this site in pE194 *TaqI* fragment B, 106 bases 3' to the B' set of sequences shown in Fig. 5. However, we have no explicit model yet to interpret this finding.

Attempt at insertional inactivation of the A determinant. A single *PstI* site is present in both pE194 and pBR322. Insertion into the *PstI* site of pBR322 inactivates the beta-lactamase determinant (1). pE194 and pBR322 plasmid DNA were digested with *PstI*, mixed in a 1:1 ratio, ligated, and incubated with competent *E. coli* cells prepared as described by Horinouchi et al.

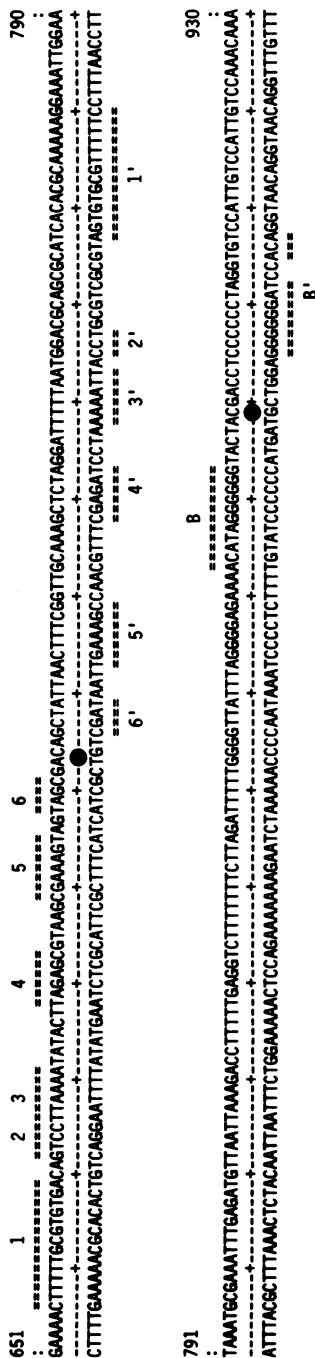


FIG. 5. Inverted complementary repeat sequences of pE194 located in *TaqI* fragment B. A 124-base region (residues 656 to 780) containing inverted complementary repeat sequences is shown together with a neighboring 34-base symmetric region (residues 873 to 907) containing a second set of inverted complementary repeat sequences. Each subsequence and its complementary counterpart are indicated by capital letters or numerals and their primed counterparts, respectively. See text for details.

(8, 9). Tetracycline-resistant transformants obtained by selection using solid enriched medium supplemented with tetracycline (10 $\mu\text{g/ml}$) were screened for ampicillin sensitivity, and cleared lysates from several such transformant clones were checked by analytical agarose gel electrophoresis. One composite plasmid DNA preparation obtained in this way was used to transform *B. subtilis* MI112. Competent cells were incubated with the composite DNA preparation for 40 min and then induced by addition of an equal volume of L-broth containing erythromycin (final concentration, 0.05 $\mu\text{g/ml}$) for an induction period of 60 min before selection by plating onto solid medium containing erythromycin (10 $\mu\text{g/ml}$). Transformants obtained after overnight incubation were checked for the presence of a plasmid that yielded the expected two fragments following digestion with *Pst*I and analytical agarose-ethidium bromide gel electrophoresis, as well as for the MLS resistance phenotype using paper disks containing erythromycin and tylosin. The structure of the composite plasmid was checked by digestion with *Pst*I, followed by agarose ethidium bromide gel electrophoresis, which yielded two fragments with mobilities of pBR322 and pE194 linear molecules. The composite plasmid formed by ligation of *Pst*I linear molecules of pBR322 and pE194 was capable of replication in both *E. coli* and *B. subtilis* after transformation and selection for tetracycline resistance in *E. coli*, or for erythromycin resistance in *B. subtilis*. These results suggest that the protein encoded by frame A is not required for either replication or expression of MLS resistance.

Examination of frame A shows the presence of the sequence GGAGG, starting at residue 3139, which could serve as the ribosome-loading site for initiation of protein synthesis, with the initiator methionine codon ATG being close by at residue 3150. Based on similarities to the frame B promoter, we would predict that the -35 and -10 sequences of frame A begin at residues 3092 and 3119, respectively. A protein encoded by frame A starting at this site and encoding the predicted 403 amino acids would have a molecular mass of 48,400 daltons. A protein band with mobility corresponding to a molecular mass of 45,000 daltons (E1 gene product) whose coding sequence contains the *Xba*I site has been reported by Shivakumar et al. (23) as one of the products synthesized by *B. subtilis* minicells carrying pE194. The E1 gene product and A frame products are most likely identical.

DISCUSSION

Our study of pE194 has concentrated on two functions of this plasmid-inducible resistance to

MLS antibiotics and plasmid replication. A second small plasmid originating in *S. aureus*, pC194 (also first described by Iordanescu and Surdeanu [13]), whose sequence is presented and analyzed in the accompanying paper (12), has served as a valuable tool in these studies because of the apparent functional and structural similarities of two sets of inverted complementary repeat sequences in pE194 and pC194, which appear to be associated with replication. The presence of three open reading frames (C, D, and E) in pE194 *Taq*I fragment B potentially encoding proteins which contain more than 100 amino acids is another noteworthy attribute of this region. This finding raises a question of the extent to which translation products of this region play a role in its function. The probability of randomly finding an open reading frame capable of encoding 100 amino acids without interruption by a single stop codon is less than 1%; however, ribosome-loading sites and sites of translation initiation do not appear to be present in association with these open frames.

When mapping the erythromycin resistance determinant of pE194, Gryczan et al. (7) noted that elimination of the pE194 *Hpa*I (*Hinc*II) site by cleavage and religation under conditions which resulted in resection of several nucleotides, or insertion of a randomly cloned *Bacillus licheniformis* DNA fragment into the *Bcl*I site, inactivated the MLS resistance determinant. The *Bcl*I site was assigned to the 29,000-dalton polypeptide promoter, a result which does not accord with sequence determination studies. Shivakumar et al. (25) also noted that RNA polymerase protected the *Bcl*II site from cleavage by this enzyme. In relation to the current complete sequence, these conclusions need reconsideration.

Data obtained from minicell experiments and from DNA sequence determination have their respective strengths and weaknesses. The DNA sequence is optimally useful in the interpretive mode and for planning critical experiments based on the location of specific restriction sites in relation to open reading frames, repeat sequences, and promoters. The existence of hypothetical proteins derived from open reading frames does not mean they necessarily exist. In particular, products encoded by the C, D, and E open reading frames on *Taq*I fragment B remain to be demonstrated. Minicells can serve as a useful test of biological function. However, the validity of data obtained from minicell preparations may be obscured by proteolysis, transcription readthrough, premature termination of translation, and formation of defective segregants during plasmid replication. Data from both systems will be useful in ascertaining additional details of pE194 function.

ACKNOWLEDGMENTS

We thank W. Reznikoff, I. Tessman, C. Murray, and J. Rabinowitz for their helpful and insightful criticism of this work. We also thank M. Fianđt and W. Szybalski for assisting us with the heteroduplex analysis, and R. Littlewood and J. Devereux for assisting us with use of the computer. L. Post, G. Davis, and M. Nomura generously provided us with much useful advice in our sequence studies. M. Jacobson, T. Baisch, J. Bacon, and P. Jaszczewski provided vital technical assistance. S. Silver helped us to improve the manuscript above and beyond his regular editorial duties.

This research was supported by The National Science Foundation (research grant PCM-719390) and by research grants from The Upjohn Co., Kalamazoo, Mich., and Eli Lilly & Co., Indianapolis, Ind.

LITERATURE CITED

- Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Cross, and S. Falkow. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2:95-113.
- Chang, S., and S. N. Cohen. 1979. High frequency transformation of *Bacillus subtilis* protoplasts by plasmid DNA. *Mol. Gen. Genet.* 168:111-115.
- Dubnau, D., and R. Davidoff-Abelson. 1971. Fate of transforming DNA following uptake by competent *Bacillus subtilis*. I. Formation and properties of the donor-recipient complex. *J. Mol. Biol.* 56:209-221.
- Ehrlich, S. D. 1977. Replication and expression of plasmids from *Staphylococcus aureus* in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 74:1680-1682.
- Gryczan, T. J., and D. Dubnau. 1978. Construction and properties of chimeric plasmids in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 75:1428-1432.
- Gryczan, T. J., G. Grandi, J. Hahn, and D. Dubnau. 1980. Conformational alteration of mRNA structure and the posttranscriptional regulation of erythromycin-induced drug resistance. *Nucleic Acids Res.* 8:6081-6096.
- Gryczan, T. J., A. G. Shivakumar, and D. Dubnau. 1980. Characterization of chimeric cloning vehicles in *Bacillus subtilis*. *J. Bacteriol.* 141:246-253.
- Horinouchi, S., T. Uozumi, and T. Beppu. 1980. Cloning of *Streptomyces* DNA into *Escherichia coli*: absence of heterospecific gene expression of *Streptomyces* genes in *E. coli*. *Agric. Biol. Chem.* 44:367-381.
- Horinouchi, S., T. Uozumi, T. Hoshino, A. Ozaki, S. Nakajima, T. Beppu, and K. Arima. 1977. Molecular cloning and in vitro transcription of *Bacillus subtilis* plasmid in *Escherichia coli*. *Mol. Gen. Genet.* 157:175-182.
- Horinouchi, S., and B. Weisblum. 1980. Posttranscriptional modification of mRNA conformation: mechanism that regulates erythromycin-induced resistance. *Proc. Natl. Acad. Sci. U.S.A.* 77:7079-7083.
- Horinouchi, S., and B. Weisblum. 1981. The control region for erythromycin resistance: free energy changes related to induction and mutation to constitutive expression. *Mol. Gen. Genet.* 182:341-348.
- Horinouchi, S., and B. Weisblum. 1982. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* 150:815-825.
- Iordanescu, S., and M. Surdeanu. 1980. New incompatibility groups for *Staphylococcus aureus* plasmids. *Plasmid* 4:256-260.
- Iordanescu, S., M. Surdeanu, P. Della Latta, and R. Novick. 1978. Incompatibility and molecular relationships between small staphylococcal plasmids carrying the same resistance marker. *Plasmid* 1:468-479.
- Korn, L. J., C. L. Queen, and M. N. Wegman. 1977. Computer analysis of nucleic acid regulatory sequences. *Proc. Natl. Acad. Sci. U.S.A.* 74:4401-4405.
- Lai, C.-J., J. E. Dahlberg, and B. Weisblum. 1973. Structure of an inducibly methylatable nucleotide sequence in 23S ribosomal ribonucleic acid from erythromycin-resistant *Staphylococcus aureus*. *Biochemistry* 12:457-463.
- Lai, C.-J., and B. Weisblum. 1971. Altered methylation of ribosomal RNA in an erythromycin-resistant strain of *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U.S.A.* 68:856-860.
- Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* 74:560-564.
- Pribnow, D. 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc. Natl. Acad. Sci. U.S.A.* 72:784-788.
- Rosenberg, M., and D. Court. 1980. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* 13:319-353.
- Schaller, H., C. Gray, and K. Herrmann. 1975. Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proc. Natl. Acad. Sci. U.S.A.* 72:737-741.
- Shine, J., and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U.S.A.* 71:1342-1346.
- Shivakumar, A. G., T. J. Gryczan, Y. I. Kozlov, and D. Dubnau. 1980. Organization of the pE194 genome. *Mol. Gen. Genet.* 179:241-252.
- Shivakumar, A. G., J. Hahn, and D. Dubnau. 1979. Studies on the synthesis of plasmid-coded proteins and their control in *Bacillus subtilis* minicells. *Plasmid* 2:279-289.
- Shivakumar, A. G., J. Hahn, G. Grandi, Y. Kozlov, and D. Dubnau. 1980. Posttranscriptional regulation of an erythromycin resistance protein specified by plasmid pE194. *Proc. Natl. Acad. Sci. U.S.A.* 77:3903-3907.
- Siebenlist, U., R. B. Simpson, and W. Gilbert. 1980. *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell* 20:269-281.
- Szalay, A. A., K. Grohman, and R. L. Sinsheimer. 1977. Separation of the complementary strands of DNA fragments on polyacrylamide gels. *Nucleic Acids Res.* 4:1569-1578.
- Tanaka, T. 1979. *recE4*-independent recombination between homologous deoxyribonucleic acid segments of *Bacillus subtilis* chromosomes. *J. Bacteriol.* 139:775-782.
- Tanaka, T., and N. Kawano. 1980. Cloning vehicles for the homologous *Bacillus subtilis* host-vector system. *Gene* 10:131-136.
- Weisblum, B. 1975. Altered methylation of ribosomal ribonucleic acid in erythromycin-resistant *Staphylococcus aureus*, p. 199-206. In D. Schlessinger (ed.), *Microbiology—1974*. American Society for Microbiology, Washington, D.C.
- Weisblum, B., M. Y. Graham, T. Gryczan, and D. Dubnau. 1979. Plasmid copy number control: isolation and characterization of high-copy-number mutants of plasmid pE194. *J. Bacteriol.* 137:635-643.
- Weisblum, B., C. Siddhikol, C.-J. Lai, and V. Demohn. 1971. Erythromycin-inducible resistance in *Staphylococcus aureus*: requirements for induction. *J. Bacteriol.* 106:835-847.
- Westmoreland, B. C., W. Szybalski, and H. Riis. 1969. Mapping of deletions and substitutions in heteroduplex DNA molecules of bacteriophage lambda by electron microscopy. *Science* 163:1343-1348.
- Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. *Nature (London)* 289:751-758.