JOURNAL OF BACTERIOLOGY, Apr. 1989, p. 1974–1981
0021-9193/89/041974-08$02.00/0
Copyright © 1989, American Society for Microbiology
Vol. 171, No. 4

Nucleotide Sequence of the McrB Region of *Escherichia coli* K-12 and Evidence for Two Independent Translational Initiation Sites at the *mcrB* Locus

TROY K. ROSS,† ERIC C. ACHBERGER, AND H. DOUGLAS BRAYMER*
Department of Microbiology, Louisiana State University, Baton Rouge, Louisiana 70803

Received 15 September 1988/Accepted 6 January 1989

The McrB restriction system of *Escherichia coli* K-12 is responsible for the biological inactivation of foreign DNA that contains 5-methylcytosine residues (E. A. Raleigh and G. Wilson, Proc. Natl. Acad. Sci. USA 83:9070–9074, 1986). Within the McrB region of the chromosome is the *mcrB* gene, which encodes a protein of 51 kilodaltons (kDa) (T. K. Ross, E. C. Achberger, and H. D. Braymer, Gene 61:277–289, 1987), and the *mcrC* gene, the product of which is 39 kDa (T. K. Ross, E. C. Achberger, and H. D. Braymer, Mol. Gen. Genet., in press). The nucleotide sequence of a 2,695-base-pair segment encompassing the McrB region was determined. The deduced amino acid sequence was used to identify two open reading frames specifying peptides of 455 and 348 amino acids, corresponding to the products of the *mcrB* and *mcrC* genes, respectively. A single-nucleotide overlap was found to exist between the termination codon of the *mcrB* gene and the proposed initiation codon of the *mcrC* gene. The presence of an additional peptide of 33 kDa in strains containing various recombinant plasmids with portions of the McrB region has been reported by Ross et al. (Gene 61:277–289, 1987). The analysis of frameshift and deletion mutants of one such hybrid plasmid, pRAB-13, provided evidence for a second translational initiation site within the McrB open reading frame. The proposed start codon for translation of the 33-kDa peptide lies 481 nucleotides downstream from the initiation codon for the 51-kDa *mcrB* gene product. The 33-kDa peptide may play a regulatory role in the McrB restriction of DNA containing 5-methylcytosine.

The phenomenon of host-controlled restriction-modification enables the bacterial cell to identify and inactivate foreign DNA molecules that enter the organism. DNA is susceptible to restriction by the endonucleolytic activity of the host unless the DNA is protected by modifications at certain nucleotide bases within particular recognition sequences. The most common modifications of the DNA are due to the methylation of specific nucleotides (i.e., the N-6 position of adenine residues, the 5 position of cytosine residues, or the N-4 position of cytosine residues) at the sequences recognized by site-specific methylases (1, 5, 6, 49).

Although most restriction-modification systems use methylation as the means of protection, it is now apparent that methylation of bases at some DNA sequences confers sensitivity to restriction (25). One such system, encoded by *mrr* (for methylated adenine recognition and restriction), targets DNA containing N-6-methylated adenine residues within certain sequences (17). In addition, there have been several reports of restriction systems which are specific for DNA containing 5-methylcytosine residues (4, 27, 31). Two systems specific for 5-methylcytosine DNA, designated McrA and McrB, were shown to function in many common laboratory host strains of *Escherichia coli* (30, 31). Mrr and Mcr methyl-specific restriction systems can significantly hinder DNA-cloning procedures since the DNA of many organisms, both procaryotic and eucaryotic, contains methylated adenine or cytosine residues (11, 30).

A consensus recognition sequence has not been deduced for the McrA restriction system; however, the common recognition sequence for the McrB system is GmC (31). The McrB restriction system will recognize this target sequence in 5-hydroxymethylcytosine-containing DNA from T-even bacteriophage that is not glucosylated (31, 32) as well as in DNA methylated at the 5 position of cytosine or the N-4 position of cytosine (e.g., M · *Pvu*II-methylated DNA). The McrB system was initially mapped at 99 min on the conventional *E. coli* map (31), and a subsequent report precisely located the DNA encoding the system adjacent to the *hsdS* gene of the type I restriction system, *Eco*K (34). The McrB function is provided by the products of at least two genes, *mcrB* (33) and *mcrC* (T. K. Ross, E. C. Achberger, and H. D. Braymer, Mol. Gen. Genet., in press). Transcription of the *mcrB* gene initiates 710 base pairs beyond the end of the *Eco*K *hsdS* gene. Analysis of protein synthesis directed by plasmids containing DNA from this region showed that the product of *mcrB* is a protein of 51 kilodaltons (kDa). There is a requirement for a 39-kDa protein, encoded by the *mcrC* gene, for McrB-directed restriction (Ross et al., in press). The *mcrC* gene was located adjacent to the distal end of the *mcrB* gene. In addition to these proteins, another peptide of 33 kDa was shown to be encoded by the McrB region of the chromosome.

This report presents the DNA sequence of the McrB region of the *E. coli* K-12 chromosome and describes the organization of its constituent *mcrB* and *mcrC* genes. We address the location of the 33-kDa-peptide-coding region and suggest a possible role for this peptide.

MATERIALS AND METHODS

Bacteria, phages, and plasmids. The bacterial strains, phage vectors, and plasmid vectors used in this study are described in Table 1. Recombinant plasmids are all deriva-

* Corresponding author.
† Present address: Department of Biochemistry, University of Wisconsin, Madison, WI 53706.

TABLE 1. Bacterial strains, phages, and plasmid vectors

Strain, phage, or plasmid	Relevant features	Source or reference(s)
E. coli K-12		
K802[a]	*hsdR2 mcrB1 mcrA*	Wood (47); Raleigh and Wilson (31)
χ2813[a]	*recA56* derivative of K802	Constructed by Roy Curtiss III
JM107	*hsdR17 mcrA*	Yanisch-Perron et al. (48); Raleigh and Wilson (31)
NM522	Δ(*lac-pro*) F′ *lacZ* ΔM15 $lacI^q$ *hsd*Δ5	Gough and Murray (15)
JM83	*ara* Δ(*lac-proAB*) *rpsL* (*strA*) (ϕ80 *lacZ*ΔM15)[b]	Vieira and Messing (45)
CSR603[c]	*recA uvrA6 phr-1*	Sancar et al. (36)
Phages		
M13mp18	Vector for sequencing	Yanisch-Perron et al. (48)
M13mp19	Vector for sequencing	Yanisch-Perron et al. (48)
Plasmids		
pACYC184	Cm^r Tc^r	Chang and Cohen (7)
pUC8	Ap^r	Vieira and Messing (45)

[a] Provided by New England BioLabs.
[b] Phage ϕ80 *lacZ*ΔM15 lysogen.
[c] Provided by B. Bachmann.

tives of vector pUC8 and are shown in Fig. 1. The methods used to construct pRAB-13H, pRAB-13M1, pRAB-13M2, pRAB-17, pRAB-18, and pRAB-12 are described below. The construction of plasmids pRAB-13 and pRAB-14d has already been described (33; Ross et al., in press).

Media. Strains were routinely subcultured on LB medium (26) or LB with agar added at 1.5% (L agar). Screening of the

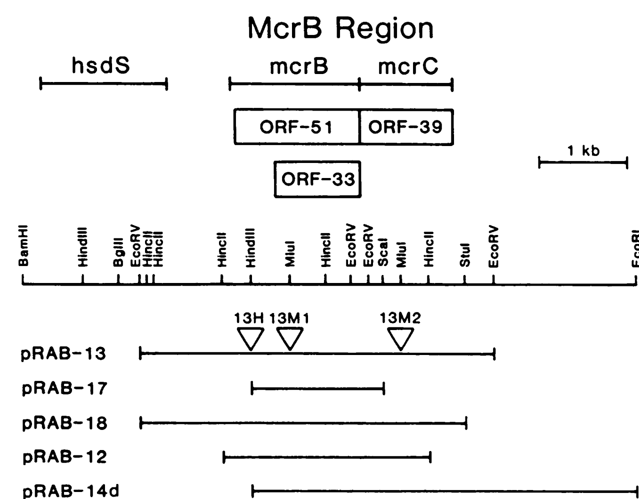

FIG. 1. A restriction map of the area from the *E. coli* K-12 chromosome showing the location of the *hsdS* gene (2, 35) and the adjacent McrB region. Within the McrB region are marked the deduced locations of the *mcrB* and *mcrC* loci and ORFs for the 51-, 39-, and 33-kDa peptides. The segments of DNA used to construct the recombinant plasmids were derived as described in the present study or as described elsewhere (33, 34). The locations of the frameshift mutations created in plasmid pRAB-13 (▽) are shown.

lactose phenotype in *E. coli* strains containing pUC8-derived plasmids was done on MacConkey agar (Difco Laboratories, Detroit, Mich.). Antibiotics were added to growth media at the following concentrations: ampicillin, 50 μg/ml; kanamycin, 50 μg/ml; chloramphenicol, 50 μg/ml (chloramphenicol was added to cultures at 200 μg/ml for plasmid amplification). L-agar plates and water top agar (0.65% agar) were used for plating M13 phage.

Chemicals and enzymes. Restriction enzymes were from Bethesda Research Laboratories, Inc., Gaithersburg, Md., or from New England BioLabs, Inc., Beverly, Mass. Prestained protein M_r markers and *E. coli* DNA polymerase I large fragment were from Bethesda Research. M · *Alu*I methylase was from New England BioLabs. Labeled methionine was from Dupont, NEN Research Products, Boston, Mass. Ethidium bromide, bromphenol blue, lysozyme, Triton X-100, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal), methyl-β-D-thiogalactoside, and antibiotics were from Sigma Chemical Co., St. Louis, Mo. Technical-grade CsCl was from KBI, Division of Cabot Corp., Reading, Pa. Agarose, acrylamide, glycine, ammonium persulfate, bisacrylamide, urea, and TEMED (*N*,*N*,*N*′,*N*′-tetramethylethylenediamine) were from Bio-Rad Laboratories, Richmond, Calif. Dideoxynucleoside triphosphates were from P-L Biochemicals, Inc., Milwaukee, Wis. Deoxynucleoside triphosphates were from Boehringer Mannheim Biochemicals, Indianapolis, Ind. All other chemicals were reagent grade and commercially available.

General molecular cloning techniques. Plasmid DNA was prepared from cells by a Triton X-100 lysis procedure, followed by ethidium bromide-CsCl equilibrium density gradient centrifugation (8, 16, 43). The method used for rapid, small-scale isolation of plasmid DNA was a modification of the boiling method of Holmes and Quigley (18). The procedures for restriction and ligation of DNAs and transformation of bacterial strains have been described elsewhere (33). For preparative digestions, 20 μg of plasmid DNA was digested with the appropriate restriction enzyme(s) and was electrophoresed on an agarose gel. The desired fragment was isolated by electroelution into a trough (24) and was further purified by the use of an ELUTIP-d column (Schleicher & Schuell, Inc., Keene, N.H.). Purification of DNA fragments from acrylamide gels has been described previously (34).

Construction of frameshift mutants of pRAB-13. Frameshift mutations in plasmid pRAB-13 were created by linearizing the plasmid molecule by partial digestion with either *Hin*dIII or *Mlu*I and repairing the 3′ recessed termini of each end, using the 5′-to-3′ polymerase activity of the large fragment of DNA polymerase I. Plasmid pRAB-13 DNA (10 μg) was partially digested with the appropriate restriction endonuclease, and the linear form of the plasmid was purified after agarose gel electrophoresis. The ends of the plasmid were made flush by using DNA polymerase I large fragment and each of the four deoxynucleotides or, for *Mlu*I-digested DNA, dGTP and dCTP (24). The plasmid DNA was circularized by using T4 DNA ligase and was subsequently used to transform strain χ2813 to Ap^r. The original pRAB-13 plasmid contains two sites for both *Hin*dIII and *Mlu*I. Since the recombinant plasmids with filled-in *Hin*dIII or *Mlu*I restriction sites have lost one of two sites for the respective enzyme, digestion with that enzyme yielded a linear form of the plasmid. Double digestions with a restriction enzyme having a unique recognition site in pRAB-13 in combination with either *Hin*dIII or *Mlu*I were performed to identify which of the *Hin*dIII or *Mlu*I restriction sites were destroyed in the mutants.

McrB restriction assay. To test *E. coli* strains for their ability to restrict 5-methylcytosine-containing DNA, the efficiency of transformation for methylated or unmethylated pACYC184 plasmid DNA was measured. Plasmid pACYC184 was methylated by using M · *Alu*I DNA methylase under reaction conditions specified by the manufacturer. Procedures for the preparation of competent cells (34) and for transformations (43) have been described previously. Portions of each transformation mixture were plated on L agar with either chloramphenicol and ampicillin or, in the case of the *trans* complementation assay, chloramphenicol, ampicillin, and kanamycin and were incubated overnight at 37°C. The data were normalized to total transformants per 0.1 μg of pACYC184 DNA, and the ratio of transformants obtained with unmethylated pACYC184 DNA relative to the number of transformants obtained with M · *Alu*I-methylated pACYC184 DNA was calculated for each strain.

Maxicell analysis of plasmid-encoded proteins. Strain CSR603 was transformed with plasmid pUC8 or recombinant derivatives of pUC8 (designated pRAB), and the plasmid-encoded proteins produced in the transformants were radiolabeled with L-[^{35}S]methionine as already described (38). The maxicells were harvested by centrifugation at 4°C and were suspended in 0.2 ml of sample buffer (36). A portion of each maxicell preparation was placed in a boiling water bath for 5 min. Samples were electrophoresed on sodium dodecyl sulfate-12% polyacrylamide gels (22), and the labeled proteins were detected by autoradiography. Prestained protein molecular weight markers were used as standards. The relative amounts of proteins were quantified from the autoradiogram by using a video densitometer (Bio-Rad). The results were normalized to the number of methionines in each peptide. When comparisons of the levels of peptide synthesis between strains were made, the results were reported relative to the production of the constitutively expressed β-lactamase gene present in the vector.

DNA sequence determination. The preparation of single-stranded DNA templates from single plaques of recombinant M13 phage lysates has been described elsewhere (15). Hybrid phages were identified as white plaques by plating dilutions of each lysate on strain NM522 in the presence of 30 μl of X-Gal (20 mg/ml in *N*,*N*-dimethylformamide) and 20 μl of the inducer methyl-β-D-thiogalactoside (24 mg/ml). DNA fragments subcloned in either M13mp18 or M13mp19 were sequenced by the chain termination method, using dideoxynucleoside triphosphates (39). The sequencing reactions were routinely primed with the universal pentadecamer primer from New England BioLabs. For larger DNA fragments which did not contain convenient restriction sites, it was necessary to use custom-synthesized DNA primers and the Sequenase DNA sequencing kit from U.S. Biochemical Corp., Cleveland, Ohio. The sequencing reactions were fractionated on 0.4-mm-thick 6% denaturing polyacrylamide gels or buffer gradient gels (3). The nucleotide sequences of both strands of the DNA were determined to generate the data presented below. DNA templates with significant sequence overlap were used to insure the fidelity of the sequence.

RESULTS

Mutagenesis of the McrB region for *E. coli* K-12. The construction of pRAB-13 by the insertion of a 3.6-kilobase *Eco*RV DNA fragment from the McrB region into vector pUC8 was reported previously (33) (Fig. 1). McrB activity was determined by the decrease in transformation efficiency observed when strain χ2813(pRAB-13) was transformed with M · *Alu*I-methylated DNA compared with unmethylated DNA. A 29-fold decrease in efficiency was measured in this strain when the transforming DNA was M · *Alu*I methylated. In contrast, the negative control, χ2813(pUC8), which has an McrB$^-$ phenotype, showed a value of 0.9 (Table 2). Strain JM107, which has the McrB$^+$ phenotype encoded on the chromosome, exhibited a 7.2-fold drop in transformation efficiency with M · *Alu*I-methylated DNA (Table 2).

TABLE 2. Restriction of M · *Alu*I-methylated pACYC184 DNA

Host strain (plasmid)[a]	No. of transformants obtained with DNA[b]		Unmethylated/methylated[c]
	Unmethylated	Methylated	
χ2813(pRAB-13)	2.2×10^4	7.6×10^2	29
χ2813(pRAB-13H)	2.1×10^4	2.5×10^4	0.8
χ2813(pRAB-13M1)	5.1×10^4	6.6×10^4	0.8
χ2813(pRAB-13M2)	7.3×10^3	5.6×10^3	1.3
χ2813(pRAB-17)	7.5×10^4	7.2×10^4	1.0
χ2813(pRAB-12)	6.4×10^4	5.6×10^4	1.1
χ2813(pRAB-14d)	5.3×10^4	4.7×10^4	1.1
χ2813[pRAB-13M1, pRAB-13M2(Kn)]	1.4×10^4	2.8×10^3	5.0
χ2813[pRAB-12(Kn), pRAB-14d]	2.3×10^4	1.4×10^4	1.6
χ2813(pUC8)[d]	2.4×10^3	2.6×10^3	0.9
JM107[e]	2.6×10^4	3.6×10^3	7.2

[a] Strains and plasmid vectors are described in Table 1, and recombinant plasmids are shown in Fig. 1.

[b] These data have been normalized to reflect the total number of transformants per 0.1 μg of pACYC184 DNA.

[c] The ratio of transformants obtained with unmethylated pACYC184 DNA relative to transformants obtained with methylated pACYC184 DNA.

[d] The strain represents an McrB$^-$ negative control.

[e] This strain has the functional McrB regional encoded on the chromosome (30).

The proteins produced by plasmid pRAB-13 in maxicells were detected by autoradiography. In addition to the two forms of the β-lactamase protein detected in the pUC8 negative control lane (Fig. 2, lane A), pRAB-13 produced three peptides. The 51-kDa product of the *mcrB* gene and the 39-kDa product of the *mcrC* gene, in addition to a third peptide with a molecular mass of approximately 33 kDa, were detected in the lane representing pRAB-13 (Fig. 2, lane B). The relative amounts of the three peptides determined from autoradiograms were 3:1:3 for the 51-, 39-, and 33-kDa peptides, respectively. It was reported previously that the *mcrC* gene product, together with the *mcrB* gene product, is required for McrB-directed restriction of 5-methylcytosine DNA and that the coding region for McrC lies near the end of the McrB-coding region (Ross et al., in press). The location of the coding region for the 33-kDa peptide, however, could not be deduced in previous reports (33, 34). Three derivatives of the prototype plasmid, pRAB-13, each containing unique frameshift mutations, were used to precisely orient the coding regions for each of the three proteins described above. One such plasmid, pRAB-13H, was constructed by inserting four nucleotides at the *Hin*dIII site within pRAB-13, thereby causing a +1 shift in the translational reading frame at that site (Fig. 1). This mutant plasmid did not impart the McrB phenotype upon strain χ2813 (Table 2). When this mutant was analyzed in maxicells, it was evident that the McrB$^-$ phenotype resulted from the lack of the 51-kDa *mcrB* gene product (Fig. 2, lane C). Plasmid pRAB-13H did produce the 39-kDa *mcrC* gene product and the 33-kDa peptide.

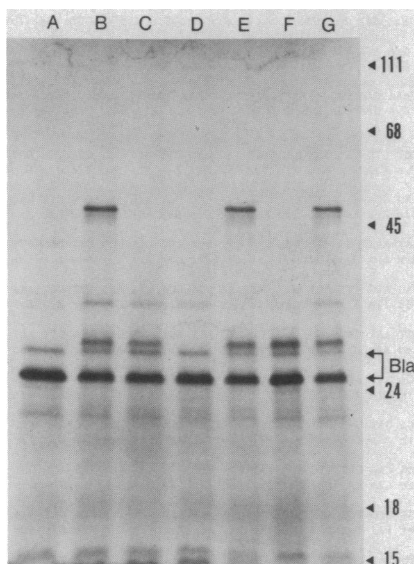

FIG. 2. Autoradiograph of sodium dodecyl sulfate-12% polyacrylamide gel depicting plasmid-produced proteins, which were labeled with L-[^{35}S]methionine in a maxicell system. Plasmids: lane A, pUC8; lane B, pRAB-13; lane C, pRAB-13H; lane D, pRAB-13M1; lane E, pRAB-13M2; lane F, pRAB-17; lane G, pRAB-18. The M_r values (in thousands) corresponding to the following protein markers, from top to bottom, are shown: phosphorylase B, bovine serum albumin, ovalbumin, α-chymotrypsinogen, β-lactoglobulin, and lysozyme. The unprocessed and processed forms of the β-lactamase (Bla) protein are marked.

The other two independent frameshift mutants were constructed by inserting four nucleotides at each of the two *Mlu*I sites within pRAB-13. The mutant plasmid with a +1 shift in the reading frame at the left-hand *Mlu*I site of pRAB-13 was designated pRAB-13M1 (Fig. 1). The plasmid isolated with a +1 shift in the reading frame at the right-hand *Mlu*I site was designated pRAB-13M2 (Fig. 1). Neither strain χ2813 (pRAB-13M1) nor strain χ2813(pRAB-13M2) exhibited the McrB$^+$ phenotype (Table 2). The maxicell analysis of proteins encoded by these mutant plasmids revealed that CSR603(pRAB-13M1) produced neither the 51-kDa McrB protein nor the 33-kDa peptide but produced the 39-kDa McrC protein (Fig. 2, lane D). Since a frameshift mutation at a single locus (left-hand *Mlu*I site) eliminated both the 51- and 33-kDa peptides, we infer that the coding regions for these proteins overlap downstream of the *Hin*dIII site (Fig. 1). While the 39-kDa protein can be translated independently of the 33- or 51-kDa peptides, the levels of synthesis were decreased. On the basis of densitometric measurements from autoradiograms, the amount of the 39-kDa protein produced in the pRAB-13M1-containing strain relative to the production of the labeled β-lactamase products was one-half that observed for strains in which the 51- and 33-kDa peptides were also produced (e.g., pRAB-13-containing strains). The frameshift at the right-hand *Mlu*I site in pRAB-13M2 permitted production of both the 51-kDa McrB protein and the 33-kDa peptide but not of the 39-kDa McrC protein (Fig. 2, lane E).

Further subcloning of the McrB region. Two additional subclones were used to define the limits of the protein-coding regions within the McrB region. Plasmid pRAB-17 was constructed by ligating the purified 1.4-kilobase *Hin*dIII-*Sca*I restriction fragment from pRAB-13 with *Hin*dIII- and *Hin*cII-digested pUC8 DNA. The isolated plasmid subclone, pRAB-17, was not able to impart the *mcrB*$^+$ phenotype to strain χ2813 (Table 2). The maxicell analysis of this plasmid showed that it produced only the 33-kDa peptide (Fig. 2, lane F). Therefore, this protein can be translated independently of the 51- and 39-kDa proteins.

Another subclone, pRAB-18, was constructed by deleting all insert DNA downstream from the *Stu*I site in pRAB-13 (Fig. 1). Plasmid pRAB-18 is missing approximately 0.3 kilobases of DNA from the right-hand end of the segment cloned in pRAB-13 (Fig. 1). Since the intact 39-kDa protein was produced in CSR603(pRAB-18), *mcrC* does not extend beyond the *Stu*I site (Fig. 1). Although pRAB-18 encodes the same proteins as does the prototype plasmid, pRAB-13 (Fig. 2, lane G), poor growth of χ2813 containing pRAB-18 did not permit the accurate determination of McrB activity.

Complementation of the McrB$^+$ phenotype by using two hybrid plasmids. The McrB$^+$ phenotype was imparted to χ2813 by transforming this strain with two frameshift mutant derivatives of plasmid pRAB-13. Plasmid pRAB-13M1, which produces the 39-kDa McrC protein, and plasmid pRAB-13M2, which produces the 51-kDa McrB protein and the 33-kDa peptide, were established in strain χ2813. This was accomplished by first inserting a kanamycin resistance cassette (GenBlock; Pharmacia, Inc., Piscataway, N.J.) into the *Sca*I site within the β-lactamase-coding region of plasmid pRAB-13M2. This new derivative of pRAB-13M2 was, therefore, Ap^s and Kn^r and was designated pRAB-13M2 (Kn). Plasmids pRAB-13M1 and pRAB-13M2(Kn) were then used to transform χ2813 to Ap^r and Kn^r. The transformed strain, χ2813[pRAB-13M1, pRAB-13M2(Kn)], was found to impart the McrB$^+$ phenotype, although neither plasmid species alone imparted McrB activity in this strain (Table 2).

Additional *trans* complementation data revealed a possible role for the 33-kDa peptide in McrB restriction. Hybrid plasmid pRAB-12, consisting of a 2.3-kilobase partially digested *Hin*cII restriction fragment subclone in pUC8 (Fig. 1), produces the 51-kDa *mcrB* gene product and the 33-kDa peptide but not the 39-kDa *mcrC* gene product (Ross et al., in press). Another plasmid, pRAB-14d, produces the 33-kDa peptide and the 39-kDa McrC protein but not the 51-kDa McrB protein (33). Neither strain χ2813(pRAB-12) nor χ2813(pRAB-14d) possessed McrB activity (Table 2). By using the same strategy as above, plasmid pRAB-12 was made Ap^s and Kn^r and was transformed in strain χ2813 along with pRAB-14d. Although pRAB-12 produces the 51-kDa McrB protein and the 33-kDa peptide and pRAB-14d produces the 39-kDa McrC protein and the 33-kDa peptide, strain χ2813[pRAB-12(Kn), pRAB-14d] exhibited very low levels of McrB restriction activity (Table 2). Agarose gel electrophoresis of plasmid DNA isolated from this strain demonstrated that pRAB-12(Kn) and pRAB-14d were present in equal quantities (data not shown).

Since each set of experiments using two plasmids in *trans* employed plasmids with sequence homology, it was necessary to rule out interplasmid recombination as the source of the results. Restriction enzyme digestions of the plasmids isolated from the tested strains revealed no discernible changes. In addition, when isolated plasmid mixtures [i.e., pRAB-13M1 plus pRAB-13M2(Kn) or pRAB-12(Kn) plus pRAB-14d] were used to transform χ2813 to either Ap^r or Kn^r, strains with a single plasmid species were McrB$^-$. This indicated that no recombination had occurred to generate a single plasmid with wild-type copies of both genes.

The nucleotide sequence of the McrB region of *E. coli* K-12. The segment of DNA represented in Fig. 3 consists of 2,695

```
GTTAACCACCGGGAGCCTTCCACTGACTCAATAGAAACTTTCCCCCTCAGTAAATATTTACCAGTCTGAT       70
                   "-35"                   "-10"      +1
TTTGCAGTAAAAATCTATTGTTTCAGTACGTTGCGAAAGCGATAATAGAGGCTTAGCAATGAGGAAGGCA      140

TATCTT ATG GAA TCT ATT CAA CCC TGG ATT GAA AAA TTT ATT AAG CAA GCA CAG CAA CAA   200
       Met Glu Ser Ile Gln Pro Trp Ile Glu Lys Phe Ile Lys Gln Ala Gln Gln Gln

CGT TCG CAA TCC ACT AAA GAT TAT CCA ACG TCT TAC CGT AAC CTG CGA GTA AAA TTG AGT   260
Arg Ser Gln Ser Thr Lys Asp Tyr Pro Thr Ser Tyr Arg Asn Leu Arg Val Lys Leu Ser

TTC GGT TAT GGT AAT TTT ACG TCT ATT CCC TGG TTT GCA TTT CTT GGA GAA GGT CAG GAA   320
Phe Gly Tyr Gly Asn Phe Thr Ser Ile Pro Trp Phe Ala Phe Leu Gly Glu Gly Gln Glu

GCT TCT AAC GGT ATA TAT CCC GTT ATT CTC TAT TAT AAA GAT TTT GAT GAG TTG GTT TTG   380
Ala Ser Asn Gly Ile Tyr Pro Val Ile Leu Tyr Tyr Lys Asp Phe Asp Glu Leu Val Leu

GCT TAT GGT ATA AGC GAC ACG AAT GAA CCA CAT GCC CAA TGG CAG TTC TCT TCA GAC ATA   440
Ala Tyr Gly Ile Ser Asp Thr Asn Glu Pro His Ala Gln Trp Gln Phe Ser Ser Asp Ile

CCT AAA ACA ATC GCA GAG TAT TTT CAG GCA ACT TCG GGT GTA TAT CCT AAA AAA TAC GGA   500
Pro Lys Thr Ile Ala Glu Tyr Phe Gln Ala Thr Ser Gly Val Tyr Pro Lys Lys Tyr Gly

CAG TCC TAT TAC GCC TGT TCC CAA AAA GTC TCA CAG GGT ATT GAT TAC ACC CGA TTT GCC   560
Gln Ser Tyr Tyr Ala Cys Ser Gln Lys Val Ser Gln Gly Ile Asp Tyr Thr Arg Phe Ala

TCT ATG CTG GAC AAC ATA ATC AAC GAC TAT AAA TTA ATA TTT AAT TCT GGC AAG AGT GTT   620
Ser Met Leu Asp Asn Ile Ile Asn Asp Tyr Lys Leu Ile Phe Asn Ser Gly Lys Ser Val

ATT CCA CCT[ATG]TCA AAA ACT GAA TCA TAC TGT CTG GAA GAT GCG TTA AAT GAT TTG TTT   680
Ile Pro Pro[Met]Ser Lys Thr Glu Ser Tyr Cys Leu Glu Asp Ala Leu Asn Asp Leu Phe

ATC CCT GAA ACC ACA ATA GAG AGC GAT ACT CAA ACG ATT AAC CAT CAA AAA AAA TAT TAT   740
Ile Pro Glu Thr Thr Ile Glu Ser Asp Thr Gln Thr Ile Asn His Gln Lys Lys Tyr Tyr

CCT CCA GGG GCC GCC CGG CGT TGG AAA AAC CTT TGT TGC ACG CGT CTG GCT TAC TTG CTG   800
Pro Pro Gly Ala Ala Arg Arg Trp Lys Asn Leu Cys Cys Thr Arg Leu Ala Tyr Leu Leu

ACA GGA GAA AAG GCT CGA ACG AAA TAT GGT TCA GTT CAT CAA TCT TAT AGC TAT GAG GAT   860
Thr Gly Glu Lys Ala Arg Thr Lys Tyr Gly Ser Val His Gln Ser Tyr Ser Tyr Glu Asp

TTT ATA CAG GGT ATC GTC CGA ATG GCG TCG GCT TCC GTA AAG ACG CAT ATT TAC AAT TTT   920
Phe Ile Gln Gly Ile Val Arg Met Ala Ser Ala Ser Val Lys Thr His Ile Tyr Asn Phe

TGT CAG CAA GCT AAA GAG CAG CCA GAG AAA AAG TAT ATT TTT ATT ATA GAT GAA ATC AAT   980
Cys Gln Gln Ala Lys Glu Gln Pro Glu Lys Lys Tyr Ile Phe Ile Ile Asp Glu Ile Asn

CGT GCC AAT CTC AGT AAA GTA TTT GGC GAA GTG ATG ATG TTA ATG GAA CAT GAT AAA CGA  1040
Arg Ala Asn Leu Ser Lys Val Phe Gly Glu Val Met Met Leu Met Glu His Asp Lys Arg

GGT GAA AAC TGG TCT GTT CCC CTA ACC TAC TCC GAA AAC GAT GAA GAA CGA TTC TAT GTC  1100
Gly Glu Asn Trp Ser Val Pro Leu Thr Tyr Ser Glu Asn Asp Glu Glu Arg Phe Tyr Val

CCG GAG AAT GTT TAT ATC ATC GGT TTA ATG AAT ACT GCC GAT CGC TCT CTG GCC GTT GTT  1160
Pro Glu Asn Val Tyr Ile Ile Gly Leu Met Asn Thr Ala Asp Arg Ser Leu Ala Val Val

GAC TAT GCC CTA CGC AGA CGA TTT TCT TTC ATA GAT ATT GAG CCA GGT TTT GAT ACA CCA  1220
Asp Tyr Ala Leu Arg Arg Arg Phe Ser Phe Ile Asp Ile Glu Pro Gly Phe Asp Thr Pro

CAG TTC CGG AAT TTT TTA CTG AAT AAA AAA GCA GAA CCT TCA TTT GTT GAG TCT TTA TGC  1280
Gln Phe Arg Asn Phe Leu Leu Asn Lys Lys Ala Glu Pro Ser Phe Val Glu Ser Leu Cys

CAA AAA ATG AAC GAG TTG AAC CAG GAA ATC AGC AAA GAG GCC ACT ATC CTT GGG AAA GGA  1340
Gln Lys Met Asn Glu Leu Asn Gln Glu Ile Ser Lys Glu Ala Thr Ile Leu Gly Lys Gly

TTC CGC ATT GGG CAT AGT TAC TTC TGC TGT GGG TTG GAA GAT GGC ACC TCT CCG GAT ACG  1400
Phe Arg Ile Gly His Ser Tyr Phe Cys Cys Gly Leu Glu Asp Gly Thr Ser Pro Asp Thr

CAA TGG CTT AAT GAA ATT GTG ATG ACG GAT ATC GCC CCT TTA CTC GAA GAA TAT TTC TTT  1460
Gln Trp Leu Asn Glu Ile Val Met Thr Asp Ile Ala Pro Leu Leu Glu Glu Tyr Phe Phe

GAT GAC CCC TAT AAA CAA CAG AAA TGG ACC AAC AAA TTA TTA GGG GAC TCA TAGTG GAA C  1520
Asp Asp Pro Tyr Lys Gln Gln Lys Trp Thr Asn Lys Leu Leu Gly Asp Ser End
                                                                    Met Glu G

AG CCC GTG ATA CCT GTC CGT AAT ATC TAT TAC ATG CTT ACC TAT GCA TGG GGT TAT TTA C 1580
ln Pro Val Ile Pro Val Arg Asn Ile Tyr Tyr Met Leu Thr Tyr Ala Trp Gly Tyr Leu G

AG GAA ATT AAG CAG GCA AAC CTT GAA GCC ATA CCC GGT AAC AAT CTT CTT GAT ATC CTG G 1640
ln Glu Ile Lys Gln Ala Asn Leu Glu Ala Ile Pro Gly Asn Asn Leu Leu Asp Ile Leu G

GG TAT GTA TTA AAT AAA GGG GTT TTA CAG CTT TCA CGC CGA GGG CTT GAG CTT GAT TAC A 1700
ly Tyr Val Leu Asn Lys Gly Val Leu Gln Leu Ser Arg Arg Gly Leu Glu Leu Asp Tyr A

AT CCT AAC ACC GAG ATC ATT CCT GGC ATC AAA GGG CGA ATA GAG TTT GCT AAA ACA ATA C 1760
sn Pro Asn Thr Glu Ile Ile Pro Gly Ile Lys Gly Arg Ile Glu Phe Ala Lys Thr Ile A

GC GGC TTC CAT CTT AAT CAT GGG AAA ACC GTC AGT ACT TTT GAT ATG CTT AAT GAA GAC A 1820
rg Gly Phe His Leu Asn His Gly Lys Thr Val Ser Thr Phe Asp Met Leu Asn Glu Asp T

CG CTG GCT AAC CGA ATT ATA AAA AGC ACA TTA GCC ATA TTA ATT AAG CAT GAA AAG TTA A 1880
hr Leu Ala Asn Arg Ile Ile Lys Ser Thr Leu Ala Ile Leu Ile Lys His Glu Lys Leu A

AT TCA ACT ATC AGA GAT GAA GCT CGT TCA CTT TAT AGA AAA TTA CCG GGC ATT AGC ACT C 1940
sn Ser Thr Ile Arg Asp Glu Ala Arg Ser Leu Tyr Arg Lys Leu Pro Gly Ile Ser Thr L

TT CAT TTA ACT CCG CAG CAT TTC AGC TAT CTG AAT GGC GGA AAA AAT ACG CGT TAT TAT A 2000
eu His Leu Thr Pro Gln His Phe Ser Tyr Leu Asn Gly Gly Lys Asn Thr Arg Tyr Tyr L

AA TTC GTT ATC AGT GTC TGC AAA TTC ATC GTC AAT AAT TCT ATT CCA GGT CAA AAC AAA G 2060
ys Phe Val Ile Ser Val Cys Lys Phe Ile Val Asn Asn Ser Ile Pro Gly Gln Asn Lys G

GA CAC TAC CGT TTC TAT GAT TTT GAA AGA AAC GAA AAA GAG ATG TCA TTA CTT TAT CAA A 2120
ly His Tyr Arg Phe Tyr Asp Phe Glu Arg Asn Glu Lys Glu Met Ser Leu Leu Tyr Gln L

AG TTT CTT TAT GAA TTT TGC CGT CGT GAA TTA ACG TCT GCA AAC ACA ACC CGC TCT TAT T 2180
ys Phe Leu Tyr Glu Phe Cys Arg Arg Glu Leu Thr Ser Ala Asn Thr Thr Arg Ser Tyr L

TA AAA TGG GAT GCA TCG AGT ATA TCG GAT CAG TCA CTT AAT TTG TTA CCT CGA ATG GAA A 2240
eu Lys Trp Asp Ala Ser Ser Ile Ser Asp Gln Ser Leu Asn Leu Leu Pro Arg Met Glu T

CT GAC ATC ACC ATT CGC TCA TCA GAA AAA ATA CTT ATC GTT GAC GCC AAA TAC TAT AAG A 2300
hr Asp Ile Thr Ile Arg Ser Ser Glu Lys Ile Leu Ile Val Asp Ala Lys Tyr Tyr Lys S

GC ATT TTT TCA CGA CGA ATG GGA ACA GAA AAA TTT CAT TCG CAA AAT CTT TAT CAA CTG A 2360
er Ile Phe Ser Arg Arg Met Gly Thr Glu Lys Phe His Ser Gln Asn Leu Tyr Gln Leu M

TG AAT TAC TTA TGG TCG TTA AAG CCT GAA AAT GGC GAA AAC ATA GGG GGG TTA TTA ATA T 2420
et Asn Tyr Leu Trp Ser Leu Lys Pro Glu Asn Gly Glu Asn Ile Gly Gly Leu Leu Ile T

AT CCC CAC GTA GAT ACC GCA GTG AAA CAT CGT TAT AAA ATT AAT GGC TTC GAT ATT GGC T 2480
yr Pro His Val Asp Thr Ala Val Lys His Arg Tyr Lys Ile Asn Gly Phe Asp Ile Gly L

TG TGT ACC GTC AAT TTA GGT CAG GAA TGG CCG TGT ATA CAT CAA GAA TTA CTC GAC ATT T 2540
eu Cys Thr Val Asn Leu Gly Gln Glu Trp Pro Cys Ile His Gln Glu Leu Leu Asp Ile P

TC GAT GAA TAT CTC AAATAAGCGGTAAGTTTTATAGAAAATACGCTCCCGACACCACATCGGGAGCGGCTCATCA 2615
he Asp Glu Tyr Leu LysEnd

TCAGGAGGCGATCACCTCCACCAGCACATCCGTTGCCCAACTCTCTTCGCTGTTGGCGACATTCCCTTCCTTGCAGGCCT 2695
```

FIG. 3. Nucleotide sequence for DNA between the *Hinc*II site, located just downstream of the *hsdS* gene, and the *Stu*I site (Fig. 1) is presented. The −35 and −10 regions of the *mcrB* promoter and the transcriptional start site (at position 125) for *mcrB* are indicated. Other features shown are the 1-nucleotide overlap (at position 1514) between the termination codon of McrB and the initiation codon for McrC and the ATG initiation codon (boxed) that marks the location of a second translational start site within the McrB ORF.

nucleotides and encompasses a region of the *E. coli* K-12 chromosome required for McrB restriction of 5-methylcytosine-containingDNA. The first 324 bases of this sequence are identical to the sequence reported by Gough and Murray (15) for a short region beyond the end of the *hsdS* gene. The −35 and −10 regions of the previously identified *mcrB* gene promoter, along with the *mcrB* transcriptional start (33), are labeled (Fig. 3). The presumptive initiation codon for the McrB translation is shown at position 147. Just upstream from this methionine is a purine-rich region that could serve as a ribosome-binding site (44). By using this methionine as the initiation codon, a peptide of 455 amino acids is predicted. This corresponds to a protein with a molecular weight of 52,705, which is in good agreement with the estimate of 51 kDa for the size of the *mcrB* gene product (Fig. 2).

The putative start codon for McrC translation is a GTG triplet at position 1514. There is a purine-rich region 7 bases upstream of this codon that could serve as a ribosome-binding site for translation of this protein. This arrangement would provide for a 1-nucleotide overlap between the stop codon for McrB translation and the initiation codon for McrC translation. By using this GTG triplet as the McrC initiation codon, a peptide of 348 amino acids is predicted. This corresponds to a protein with a molecular weight of 41,500 and is in agreement with the 39-kDa estimate for the size of the McrC protein from maxicell analyses (Fig. 2).

Existence of overlapping open reading frames (ORFs) within the McrB region. Analyses of protein profiles produced from strains containing the various plasmid constructions were consistent with the existence of an overlapping ORF that encodes the 33-kDa peptide. Examination of the nucleotide sequence of the *mcrB* gene indicated that the ORF of the 33-kDa peptide must be in the same reading frame as that of the 51-kDa protein. The ATG codon at position 630 is the probable initiation codon for the 33-kDa peptide (Fig. 3, box). This second translational start in the McrB ORF could initiate a peptide of 294 amino acids that correlates to a peptide with a molecular weight of 34,204.

ATP-binding domains within the McrC protein. A report by Walker et al. (46) compared the amino acid compositions of a number of ATP-binding proteins with the sequence of the alpha- and beta-subunits of the *E. coli* ATP synthase protein complex. Two domains with highly conserved amino acid sequences were identified. Domain A has the consensus sequence Gly-X_4-Gly-Lys-Thr-X_6-Ile/Val, and domain B generally has the sequence Arg/Lys-X_3-Gly-X_3-Lys-hydrophobic-hydrophobic-hydrophobic, followed by an aspartic

TABLE 3. Alignment of homologous sequences in adenine nucleotide-binding proteins

Protein	First Residue	Sequence	Reference
Domain A[a]			
RecA Protein	61	I V E I Y G P E S S G K T T L T L Q V I	(37)
Bovine ATPase β	153	I G L F - G G A G V G K T V F I M E L I	(46)
E. coli ATPase α	165	E L I I - G D R G T G K T A L A I D A I	(14)
E. coli ATPase β	146	V G L F - G G A G V G K T V N M M E L I	(40)
Adenylate kinase	10	I I F V V G G P G S G K G T Q C E K I V	(29)
E. coli McrC	79	A K T I R G F H L N H G K T V S T F D M	
Domain B[b]			
Bovine ATPase β	244	Y F R D Q E G Q D V L L F I (D) N I F R F	(46)
E. coli ATPase α	268	Y F R D - R G E D A L I I Y (D) D L S K Q	(14)
E. coli ATPase β	230	K F R D - E G R D V L L F V (D) N I Y R Y	(40)
Adenylate kinase	105	F E R K - I G Q P T L L L Y V (D) A G P E	(29)
Phosphofructokinase	88	Q L K K - H G I Q G L V V I G G (D) G S Y	(21)
E. coli McrC	288	S L K P E N G E N I G G L L I Y P H V (D)	

[a] For domain A, the boxed sequences are functionally identical.

[b] For domain B, the third box contains conserved hydrophobic residues, and encircled amino acids indicate aspartic acid residues that could bind magnesium ions.

acid residue that might be involved in the binding of magnesium ions.

Since it has been reported that sequence-specific McrB restriction of 5-methylcytosine-containing DNA in vitro is dependent on added ATP (E. Sutherland and E. A. Raleigh, Abstr. Workshop on Biological DNA Methylation, p. 35, 1988), a computer-assisted search was made for the existence of ATP-binding domains within the McrB and McrC proteins (9). A sequence beginning at amino acid 79 within the McrC protein corresponds to the consensus sequence for domain A of ATP-binding proteins. The highly conserved Gly-Lys-Thr sequence is present in the McrC amino acid sequence; however, it is shifted by 1 residue (Table 3). A putative B domain is located at amino acid 288 in McrC. Again, a shift is required to locate the conserved hydrophobic beta-sheet structure followed by an aspartic acid residue (Table 3). The existence of these regions of homology may be indicative of ATP binding within the McrC protein.

DISCUSSION

The analyses of three independent +1 frameshift mutants of plasmid pRAB-13 and of two additional subclones of the prototype plasmid permitted the precise assignment of the ORFs to the three proteins produced from the McrB region (Fig. 1). None of the three frameshift mutants of pRAB-13 imparted the McrB$^+$ phenotype to strain χ2813 (Table 2) because either the 51-kDa *mcrB* gene product (Fig. 2, lanes C and D) or the 39-kDa *mcrC* gene product (Fig. 2, lane E) was missing. The gene product from pRAB-13M1, together with those from pRAB-13M2, did complement the McrB$^-$ phenotype of strain χ2813 when both plasmids were maintained in the same cell (Table 2). This confirms the idea that at least two proteins, the 51-kDA *mcrB* gene product and the 39-kDa *mcrC* gene product, are required for *E. coli* K-12 to express McrB restriction activity.

The nucleotide sequence of a 2,695-base-pair segment of DNA details the organization of the coding regions within the McrB region of *E. coli* K-12. The deduced amino acid sequence is consistent with the production of a 51-kDa protein and a 39-kDa protein. The DNA sequence revealed that the proposed initiation codon of the McrC protein overlaps the termination of the *mcrB* gene by 1 nucleotide. The overlap of the *mcrB* and *mcrC* genes (Fig. 3) is consistent with translational coupling between these loci (28, 41, 42). When translation of the upstream gene was prematurely terminated (e.g., the pRAB-13M1 plasmid), production of the downstream 39-kDa protein decreased by one-half compared with that observed with the strain carrying pRAB-13. The magnitude of this decrease is comparable to that of other systems in which both genes possess functional ribosome-binding sites (42). Interestingly, the levels of the 39-kDa protein detected in maxicells are one-third the levels detected for either the 33- or 51-kDa peptide. Thus six peptides would terminate at the end of the *mcrB* ORF for each peptide

initiated at the *mcrC*-coding region. Translational coupling has been proposed for the *hsdM* and *hsdS* genes of the *Eco*K system located adjacent to the *mcrB* gene (23).

On the basis of the analysis of point mutations, the 51- and 33-kDa peptides are produced from overlapping ORFs within the *mcrB* locus. This is consistent with a previous study that reported the production of truncated versions of both the 51- and 33-kDa peptides from a hybrid plasmid containing a deletion of DNA encoding the C-terminal end of the *mcrB* gene product (33). Since there is only one ORF in this region long enough to encode these peptides (Fig. 3), the peptides must be translated in phase. On the basis of the analysis of mutants, peptide length, and sequence homology with other translational starts, the ATG codon located at position 630 is the most likely translational start site for the 33-kDa peptide (Fig. 3).

Production of the 33-kDa peptide alone (pRAB-17) or in combination with the 39-kDa protein (pRAB-14d) did not impart a detectable McrB$^+$ phenotype (Table 2). Since the 33-kDa peptide shares the same amino acid sequence with the C-terminal end of the 51-kDa protein, it could compete with the 51-kDa protein for DNA- or protein-binding sites. This could explain the lack of McrB$^+$ phenotype in strains containing pRAB-14d (i.e., a plasmid that encodes the 33- and 51-kDa peptides) together with plasmid pRAB-12(Kn), which encodes the 33- and 51-kDa peptides. If the 33-kDa peptide was overproduced, it could sequester some factor from the 51-kDa protein, such as the 39-kDa protein. Regulation of protein function by a smaller peptide synthesized from an internal translational initiation site within the same ORF has been proposed for several systems (10, 12, 13, 19, 20). In two instances, the Tn*5* transposase and the bacteriophage fl gene *II* protein, overproduction of the internal start protein in *trans* clearly inhibited the processes mediated by the full-length protein (13, 19, 20). An alternate explanation is that pRAB14d encodes a negative regulatory element within the DNA downstream of the *Stu*I site (Fig. 1). Strain χ2813(pRAB-18) (i.e., a plasmid in which DNA downstream of the *Stu*I site has been deleted) exhibited poor growth, possibly due to low viability. No growth aberration was observed when pRAB-18 was transformed in strains with an McrB$^+$ phenotype, such as JM83 and CSR603. The presence of a regulatory element downstream of the *mcrC* gene could also influence the restriction of DNA with modified cytosine residues other than 5-methylcytosine.

Another aspect of the McrB phenotype is the restriction of 5-hydroxymethylcytosine-containing DNA from T-even phage that is not glucosylated (31). This function was formerly referred to as Rgl (restricts glucoseless phage) restriction (32). Although the 39-kDa McrC protein, together with the 51-kDa McrB protein, is necessary for sequence-specific restriction of 5-methylcytosine DNA, the *mcrC* gene product is not essential for restriction of phage T4 DNA, containing nonglucosylated hydroxymethylcytosine residues (Ross et al., in press). Interestingly, the McrC protein is predicted to possess 18 more basic amino acid residues than acidic amino acid residues. A role for this protein may be to complex with the *mcrB* gene product and alter the specificity of McrB restriction activity. In this model, the role of the 33-kDa peptide would be to moderate the levels of the 39-kDa protein such that some of the 51-kDa protein remains uncomplexed. Preliminary evidence indicates that strains containing various plasmid constructions which exhibit high levels of restriction of 5-methylcytosine DNA have reduced levels of nonglucosylated T4 phage restriction. Further experiments will be necessary to determine whether these two aspects of the McrB restriction system are mutually exclusive. We plan to test the effects of overproduction and nonproduction of the 33-kDa peptide on the two activities.

ACKNOWLEDGMENTS

We thank Bob Belas for performing the DNA Sequenase reactions and Joe Fitzgibbon for assisting with the computer searches.

This work was supported in part by the Louisiana State University Agricultural Experiment Station.

LITERATURE CITED

1. **Arber, W.** 1974. DNA modification and restriction. Prog. Nucleic Acid Res. **14**:1–37.
2. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
4. **Blumenthal, R. M., S. A. Gregory, and J. S. Cooperider.** 1985. Cloning of a restriction-modification system from *Proteus vulgaris* and its use in analyzing a methylase-sensitive phenotype in *Escherichia coli*. J. Bacteriol. **164**:501–509.
5. **Boyer, H. W.** 1971. DNA restriction and modification mechanisms in bacteria. Annu. Rev. Microbiol. **25**:153–176.
6. **Butkus, V., S. Klimasauskas, L. Petrauskiene, Z. Maneliene, A. Lebionka, and A. Janulaitis.** 1987. Interaction of *Alu*I, *Cfr*6I, and *Pvu*II restriction-modification enzymes with substrates containing either N4-methylcytosine or 5-methylcytosine. Biochim. Biophys. Acta **909**:201–207.
7. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134**: 1141–1156.
8. **Clewell, D. B.** 1972. Nature of ColE1 plasmid replication in *Escherichia coli* in the presence of chloramphenicol. J. Bacteriol. **110**:667–676.
9. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
10. **Dunn, J. J., and F. W. Studier.** 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. **166**:477–535.
11. **Ehrlich, M., and R. Y. Wang.** 1981. 5-Methylcytosine in eukaryotic DNA. Science **212**:1350–1357.
12. **Eisenberg, S., and M. Finer.** 1980. Cleavage and circularization of single-stranded DNA: a novel enzymatic activity of ϕX174 A* protein. Nucleic Acids Res. **22**:5305–5315.
13. **Fulford, W., and P. Model.** 1988. Regulation of bacteriophage fl DNA replication. I. New functions for genes *II* and *X*. J. Mol. Biol. **203**:49–62.
14. **Gay, N. J., and J. E. Walker.** 1981. The *atp* operon: nucleotide sequence of the region encoding the alpha-subunit of *Escherichia coli* ATP synthase. Nucleic Acids Res. **9**:2187–2194.
15. **Gough, J. A., and N. E. Murray.** 1983. Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. **166**:1–19.
16. **Guerry, P., D. J. LeBlanc, and S. Falkow.** 1973. General method for the isolation of plasmid deoxyribonucleic acid. J. Bacteriol. **116**:1064–1066.
17. **Heitman, J., and P. Model.** 1987. Site-specific methylases induce the SOS DNA repair response in *Escherichia coli*. J. Bacteriol. **169**:3243–3250.
18. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**: 193–197.
19. **Isberg, R. R., A. L. Laazaar, and M. Syvanen.** 1982. Regulation of Tn5 by the right-repeat proteins: control at the level of the transposition reaction. Cell **30**:883–892.
20. **Johnson, R. C., and W. S. Reznikoff.** 1984. Role of the IS*50R* proteins in the promotion and control of Tn5 transposition. J. Mol. Biol. **177**:645–661.
21. **Kolb, E., P. J. Hudson, and J. I. Harris.** 1980. Phosphofructoki-

nase: complete amino-acid sequence of the enzyme from *Bacillus stearothermophilus*. Eur. J. Biochem. **108**:587–597.

22. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
23. **Loenen, W. A. M., A. S. Daniel, H. D. Braymer, and N. E. Murray.** 1987. Organization and sequence of the *hsd* genes of *Escherichia coli* K-12. J. Mol. Biol. **198**:159–170.
24. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Marinus, M. G.** 1987. DNA methylation in *Escherichia coli*. Annu. Rev. Genet. **21**:113–131.
26. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Noyer-Weidner, M., R. Diaz, and L. Reiners.** 1986. Cytosine-specific DNA modification interferes with plasmid establishment in *Escherichia coli* K-12: involvement of *rglB*. Mol. Gen. Genet. **205**:469–475.
28. **Oppenheim, D. S., and C. Yanofsky.** 1980. Translational coupling during expression of the tryptophan operon of *E. coli*. Genetics **95**:785–795.
29. **Pai, E. G., W. Sachsenheimer, R. H. Schirmer, and G. E. Schultz.** 1977. Substrate positions and induced-fit in the crystalline adenylate kinase. J. Mol. Biol. **114**:37–45.
30. **Raleigh, E. A., N. E. Murray, H. Revel, R. M. Blumenthal, D. Westaway, A. D. Reith, P. W. J. Rigby, J. Elhai, and D. Hanahan.** 1988. McrA and McrB restriction phenotypes of some *E. coli* strains and implication of gene cloning. Nucleic Acids Res. **16**:1563–1575.
31. **Raleigh, E. A., and G. Wilson.** 1986. *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. USA **83**:9070–9074.
32. **Revel, H.** 1967. Restriction of non-glucosylated T-even bacteriophages: properties of permissive mutants of *Escherichia coli* B and K-12. Virology **31**:688–701.
33. **Ross, T. K., E. C. Achberger, and H. D. Braymer.** 1987. Characterization of the *Escherichia coli* modified cytosine restriction (*mcrB*) gene. Gene **61**:277–289.
34. **Ross, T. K., and H. D. Braymer.** 1987. Localization of a genetic region involved in McrB restriction by *Escherichia coli* K-12. J. Bacteriol. **169**:1757–1759.
35. **Sain, B., and N. E. Murray.** 1980. The *hsd* (host specificity) genes of *E. coli* K-12. Mol. Gen. Genet. **180**:35–46.
36. **Sancar, A., A. M. Hack, and W. D. Rupp.** 1979. Simple method for identification of plasmid-coded proteins. J. Bacteriol. **137**: 692–693.
37. **Sancar, A., C. Stachelek, W. Konigsberg, and W. D. Rupp.** 1980. Sequences of the *recA* gene and protein. Proc. Natl. Acad. Sci. USA **77**:2611–2615.
38. **Sancar, A., R. P. Wharton, S. Seltzer, B. M. Kacinski, N. D. Clark, and W. D. Rupp.** 1981. Identification of the *uvrA* gene product. J. Mol. Biol. **148**:45–62.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
40. **Saraste, M., A. Eberle, N. J. Gay, M. J. Runswick, and J.E. Walker.** 1981. The *atp* operon: nucleotide sequence of the genes for the gamma, beta, and epsilon subunits of *Escherichia coli* ATP synthase. Nucleic Acids Res. **9**:5287–5296.
41. **Schoner, B. E., R. M. Belagaje, and R. G. Schoner.** 1986. Translation of a synthetic two-cistron mRNA in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **83**:8506–8510.
42. **Schumperli, D., K. McKenney, D. A. Sobieski, and M. Rosenberg.** 1982. Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. Cell **30**:865–871.
43. **Selker, E., K. Brown, and C. Yanofsky.** 1977. Mitomycin C-induced expression of *trpA* of *Salmonella typhimurium* in the plasmid ColE1. J. Bacteriol. **129**:388–394.
44. **Shine, J., and L. Dalgarno.** 1974. The 3′ terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
45. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
46. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**: 945–951.
47. **Wood, W. B.** 1966. Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affected the restriction and modification of DNA. J. Mol. Biol. **16**:118–133.
48. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
49. **Yuan, R.** 1981. Structure and mechanism of multifunctional restriction endonucleases. Annu. Rev. Biochem. **50**:285–315.