JOURNAL OF BACTERIOLOGY, Feb. 1991, p. 1367–1375
0021-9193/91/041367-09$02.00/0

Vol. 173, No. 4

A Family of Regulatory Genes Associated with Type II Restriction-Modification Systems

TAO TAO, JASON C. BOURNE, AND ROBERT M. BLUMENTHAL*

Department of Microbiology, Medical College of Ohio, Toledo, Ohio 43699-0008

Received 9 October 1990/Accepted 6 December 1990

Restriction-modification systems must be regulated to avoid autorestriction and death of the host cell. An open reading frame (ORF) in the *Pvu*II restriction-modification system appears to code for a regulatory protein from a previously unrecognized family. First, interruptions of this ORF result in a nonrestricting phenotype. Second, this ORF can restore restriction competence to such interrupted mutants in *trans*. Third, the predicted amino acid sequence of this ORF resembles those of known DNA-binding proteins and includes a probable helix-turn-helix motif. A survey of unattributed ORFs in 15 other type II restriction-modification systems revealed three that closely resemble the *Pvu*II ORF. All four members of this putative regulatory gene family have a common position relative to the endonuclease genes, suggesting a common regulatory mechanism.

Type II restriction-modification systems (RMS2s) produce two separate enzymes sharing the same DNA sequence specificity: a restriction endonuclease (REase) and a modification methyltransferase (MTase). Modification by the MTase protects the cell's own DNA from the REase (28, 41). Over 1,300 RMS2s have been identified, and they are produced by every known major group of prokaryotes (39). The roles of these RMS2s include defense against bacteriophage infection (26) and possibly promotion of recombination with exogenous DNA (36).

The major goal of RMS2 regulation is generally assumed to be prevention of autorestriction. That is, the REase and MTase genes must be regulated such that the cell's own DNA is fully protected before substantial REase activity appears. The need for this type of regulation is most obvious during establishment, when the genes for an RMS2 enter new host cells in which the DNA is completely unmodified. Even RMS2s that are already established may have need of this sort of regulation, to compensate for transient interruptions of protein synthesis such as occur in stationary phase, in the stringent response to amino acid limitation (11), in heat or cold shock responses (23, 31), or in response to oxidative or DNA damage (13, 46).

While the potential usefulness of regulated expression of RMS2 genes may seem obvious, the mechanisms for achieving this regulation have not been well explored. There has been one previous report of an open reading frame (ORF) that appears to influence the expression of RMS2 genes. Specifically, interruption of an ORF in the *Bam*HI RMS2 alters expression of the *Bam*HI genes (30). We have found a third ORF in the *Pvu*II RMS2 sequence with candidate expression sequences (43a), and here we present evidence that it specifies a *trans*-acting regulatory protein. This ORF, which we give the gene designation *pvuIIC* (*C* for controller; see reference 43), is required for expression of the REase gene *pvuIIR*. Furthermore, a survey of unattributed ORFs in other RMS2s reveals three, including the *Bam*HI ORF, that closely resemble *pvuIIC*.

* Corresponding author. BITNET address: BLUMENTHAL@MCOIARC.

MATERIALS AND METHODS

Strains used. The *Escherichia coli* K-12 strains used include two McrB$^{\pm}$ pairs. One pair is JM107 (50) and its mutant derivative JM107MA2 (4). The other pair is DH5α and DH5αMCR (24), both obtained from BRL/Life Technologies, Inc. The plasmids that were used include pACYC184 (12), p*Pvu*RM3.4, and p*Pvu*M1.9 (4).

Modification assays. The activity of *Pvu*II MTase was measured in two ways. Incompatibility of functional *pvuIIM* and *mcrB* genes was the basis for a transformation assay (3, 4, 37). McrB$^{\pm}$ cells were made competent by the procedure of Chung and Miller (14) and transformed in triplicate with the test plasmid at concentrations empirically shown to yield between 20 and 500 carbenicillin-resistant colonies with the control plasmid p*Pvu*M1.9. The second assay procedure involved growing λ*vir* phage on a strain carrying the test plasmid(s) and then plating triplicate dilutions of the resulting phage on strains carrying either p*Pvu*M1.9 (R^- M^+) or p*Pvu*RM3.4 (R^+ M^+). Phage stocks were grown and plated as described by Davis et al. (16).

Restriction assays. The activity of *Pvu*II REase was measured in two ways. In vivo restriction was measured by determining the efficiency of plating of unmodified λ*vir* or the efficiency of transformation of pACYC184, using procedures analogous to those described for the modification assays. In vitro restriction was measured after growing test strains overnight in Luria-Bertani medium (16) containing appropriate antibiotics. A 1.4-ml sample of the culture was pelleted, washed in 5 mM Tris-HCl (pH 8.0)–25% sucrose, and resuspended in 500 μl of LENT (0.4 mg of lysozyme per ml, 10 mM EDTA, 100 mM NaCl, 10 mM Tris-HCl [pH 7.9]). The cells were incubated for 1 h on ice and sonically disrupted, and cell debris was pelleted. Dilutions of the resulting crude extract were assayed in 20-μl reaction mixtures containing 1 μg of phage λ DNA, 10 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, 50 mM NaCl, 1 mM dithiothreitol, and 100 μg of bovine serum albumin per ml, with 1-h incubations at 37°C. The assays were resolved on 1% agarose gels and were scored for the highest dilution giving a complete digestion and for the highest dilution yielding visible amounts of all of the final digestion products even if the digest was incomplete.

Cloning and generation of mutants. The *Cla* and *Esp* mutants were generated by complete digestion of p*Pvu*

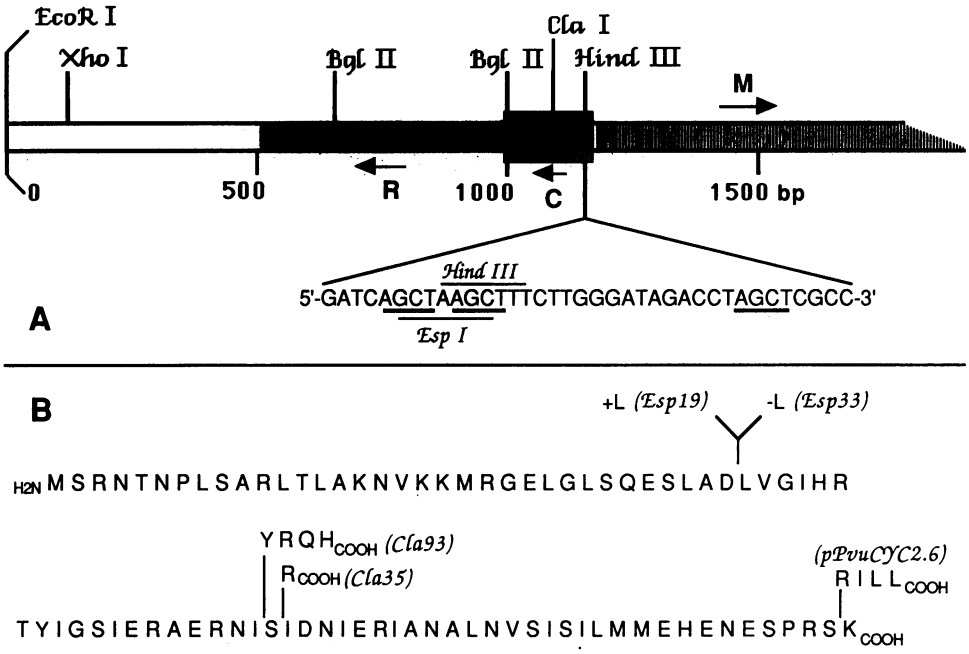

FIG. 1. (A) Genetic and restriction map of the *Pvu*II RMS2 (adapted from reference 4). The sequence around the *Hin*dIII site shows the overlapping *Esp*I site and is referred to in the Discussion. (B) Amino acid sequence of the *pvuIIC* ORF, showing the predicted results of various mutations and constructions.

RM3.4 with either *Cla*I (New England BioLabs, Inc.) or *Esp*I (Brisco, Ltd., Winthrop, Mass.) and then either filling in the 5' extensions with the large fragment of DNA polymerase I or removing the extensions with mung bean nuclease. In both cases, protocols described by Sambrook et al. (40) were followed. Following ligation and transformation, small preparations of the resulting plasmids were screened by restriction digest; candidates of the desired mutants were subjected to double-stranded dideoxy sequencing with oligonucleotide primers that had been synthesized for determining the original sequence (44). To generate a plasmid clone of *pvuIIC* that is compatible with the *Cla* and *Esp* mutant plasmids, the entire *Pvu*II RMS2 was excised from p*Pvu*RM3.4 as an *Eco*RI-*Eco*RV fragment and ligated into pACYC184 that had been digested with *Eco*RI and *Sca*I. The resulting construct was then made REase$^-$ by deleting a 0.4-kbp *Bgl*II fragment (Fig. 1A).

Sequence comparisons. FASTA analyses were carried out with a minimum match (k-tuple parameter) set to 2 (35). The RMS2 sequences scanned for potential regulatory ORFs include *Bam*HI, *Bsu*RI, *Cvi*JI, *Dde*I, *Eco*RI, *Eco*RV, *Fok*I, *Hin*fI, *Msp*I, *Pae*R7I, *Pst*I, *Pvu*II, *Rsr*I, *Sin*I, *Sma*I, and *Taq*I. The references for those sequences are in reference 48 except as noted in the text. The sequences for *Bam*HI and for *Sma*I were generously communicated prior to publication by, respectively, J. E. Brooks and J. C. Dunbar.

RESULTS

Interruption of *pvuIIC* greatly reduces expression of *pvuIIR*. The intact *Pvu*II RMS2 has been cloned as a 3.4-kb insert into pBR322, generating p*Pvu*RM3.4 (4). This entire insert has been sequenced (43a, 44); the sequence reveals an ORF between the REase and MTase genes (Fig. 1A). This *pvuIIC* ORF spans substrate sites for the restriction endonucleases *Cla*I and *Esp*I (Fig. 1A), and these sites are unique in p*Pvu*RM3.4. Neither site is within *pvuIIM* or *pvuIIR*; in the case of *Cla*I, this result has been confirmed by in vitro transcription-translation and by subcloning (4). Furthermore, analysis of the N-terminal amino acids of *Pvu*II REase protein has confirmed our assignment of the start of the *pvuIIR* ORF (2a, 43a).

Two frameshift mutations were introduced at the *Cla*I site: one by filling in *Cla*I-digested p*Pvu*RM3.4 with Klenow polymerase, which generates a new *Nru*I site as AT/CGAT becomes ATCGCGAT; and the other by digestion with mung bean nuclease, which actually removed 8 bp in the case of the mutant chosen for further study. *Esp*I generates a 3-bp overhang (GC/TNAGC), so the filled-in or nuclease-resected mutants remained in frame, adding or deleting a leucine codon. Mung bean nuclease digestion of the *Esp*I-cleaved DNA generates a new *Eco*47III site (AGCGCT). All four mutations were confirmed by restriction digestion and by DNA sequencing, and the expected changes in amino acid sequence are shown in Fig. 1B. All four mutant plasmids transform *E. coli* at the same rate as the parental plasmid, suggesting that all are as capable of establishment in new cells as the parental plasmid (not shown).

The expression of *pvuIIM* appears to be unchanged in the mutants. All four mutant plasmids are fully self-modified; that is, their *Pvu*II sites are resistant to digestion in vitro with *Pvu*II REase (not shown). All four are incompatible with the *E. coli* McrB restriction system (Table 1), which is known to digest DNA methylated by *Pvu*II MTase (3, 4, 37). Finally, bacteriophage λ, grown on *E. coli* carrying the parental or any of the four mutant plasmids, seems equally resistant to restriction by *Pvu*II R$^+$ M$^+$ strains (Table 1). If the mutations affect *pvuIIM* expression at all, they do not decrease it significantly; our data do not rule out an increase in expression.

In contrast to *pvuIIM*, the expression of *pvuIIR* is greatly reduced by all four mutations. Strains carrying any of the

TABLE 1. Effects of interrupting *pvuIIC*

Plasmid[a]	Relative methylation activity		Relative restriction activity	
	McrB restriction[b]	λ*vir* protection[c]	λ*vir* restriction[d] (PFU/ml ± SE)	pACYC restriction[e]
p*Pvu*M1.9	4.6×10^{-5} (2.2)	ND	$9.9 \pm 0.6 \times 10^{11}$ (3.4×10^{-8})	$4.2 \pm 0.4 \times 10^{4}$ (1.3×10^{-2})
p*Pvu*RM3.4	1.0×10^{-4} (1.0)	0.96 (1.00)	$3.3 \pm 1.0 \times 10^{4}$ (1.0)	$5.5 \pm 0.4 \times 10^{2}$ (1.0)
p*Pvu*RM3.4-*Cla*35	$<1.6 \times 10^{-5}$ (>6.3)	1.05 (1.09)	$1.4 \pm 0.1 \times 10^{12}$ (2.5×10^{-8})	$5.3 \pm 1.1 \times 10^{4}$ (1.1×10^{-2})
p*Pvu*RM3.4-*Cla*93	$<2.3 \times 10^{-5}$ (>4.4)	1.16 (1.21)	$1.2 \pm 0.1 \times 10^{12}$ (2.8×10^{-8})	$4.4 \pm 0.6 \times 10^{4}$ (1.3×10^{-2})
p*Pvu*RM3.4-*Esp*19	$<7.5 \times 10^{-5}$ (>1.3)	0.83 (0.86)	$1.1 \pm 0.1 \times 10^{12}$ (3.2×10^{-8})	$8.5 \pm 1.6 \times 10^{3}$ (6.5×10^{-2})
p*Pvu*RM3.4-*Esp*33	$<1.7 \times 10^{-4}$ (>5.9)	0.97 (1.01)	$1.2 \pm 0.1 \times 10^{12}$ (2.9×10^{-8})	$1.5 \pm 0.2 \times 10^{4}$ (3.6×10^{-2})

[a] Plasmid p*Pvu*RM3.4 codes for the intact *Pvu*II system; p*Pvu*M1.9 includes only *pvuIIM* intact. The *Cla* and *Esp* derivatives of p*Pvu*RM3.4 are illustrated in Fig. 1 and involve addition or deletion of several bases within the C ORF.

[b] *E. coli* DH5α and DH5αMCR were transformed with the indicated plasmid DNA. Values indicate the McrB$^+$/McrB$^-$ ratio of transformation efficiency for each plasmid. Incompatibility with McrB$^+$, due to MTase activity, results in a low ratio. Each transformation was done in triplicate on plates containing carbenicillin. Numbers in parentheses show relative McrB restriction.

[c] Stocks of λ*vir* were grown on *E. coli* JM107MA2 containing the indicated plasmid. These phage stocks were then tested for resistance to restriction by measuring the R$^+$ M$^+$/R$^-$ M$^+$ ratio of plating efficiency, using strains JM107MA2(p*Pvu*RM3.4) and JM107MA2(p*Pvu*M1.9). Each plating was done in triplicate. Numbers in parentheses show relative protection from *Pvu*II restriction. ND, Not done.

[d] An unmethylated stock of λ*vir* was plated directly on *E. coli* JM107MA2 carrying the indicated plasmid. Values are from triplicate platings. The λ genome contains 15 *Pvu*II sites. Numbers in parentheses show relative amounts of restriction.

[e] Plasmid pACYC184, which is compatible with the plasmids indicated at the left, was used to transform strain JM107MA2 carrying the indicated plasmid. Values indicate the efficiency of transformation in transformants per microgram of DNA, ± the standard error, from triplicate platings. Each transformation was plated onto medium containing both carbenicillin and chloramphenicol to select for both plasmids. pACYC184 contains two *Pvu*II sites. Numbers in parentheses show relative amounts of restriction.

mutant plasmids do not detectably restrict unmethylated bacteriophage λ or plasmid pACYC184 (Table 1), and extracts of such strains contain at least 10^4-fold less *Pvu*II REase activity than strains carrying the parental plasmid (see below).

The *pvuIIC* product acts in *trans*. The entire DNA insert from p*Pvu*RM3.4 was subcloned into pACYC184, and then a *Bgl*II fragment was deleted from within *pvuIIR*. This deletion removes 114 of the 157 codons in *pvuIIR* (Fig. 1A). Because the *pvuIIC* and *pvuIIR* ORFs overlap slightly, this deletion should also result in replacement of the carboxy-terminal lysine of *pvuIIC* with four amino acids (Fig. 1B). This construct, named p*Pvu*CYC2.6, is compatible with p*Pvu*RM3.4 and was introduced into cells containing the mutant derivatives of p*Pvu*RM3.4 to determine whether the ORF could act in *trans* (Table 2).

As expected, *pvuIIC* supplied in *trans* has no detectable effect on a subclone containing only the *pvuIIM* gene (p*Pvu*M1.9) and very little effect on the intact *Pvu*II restriction system (p*Pvu*RM3.4). In contrast, when *pvuIIC* is supplied in *trans* to any of the four *pvuIIC* mutants, the result is a roughly 10^4-fold increase in in vitro *Pvu*II REase activity and a 10^7-fold increase in restriction of λ*vir*. These results represent increases to the wild-type level of in vitro restriction activity and substantial restoration of restriction activity in vivo, indicating significant complementation despite the low copy number of p*Pvu*CYC2.6.

In theory, this result could have been due to interplasmid recombination instead of complementation, though the host strain is RecA$^-$. To address this possibility, cells from between λ*vir* plaques in the experiment described above were grown, and the plasmid pairs were isolated. Restriction digestion using *Cla*I alone or with *Xho*I, and *Esp*I alone or with *Xho*I, confirmed that the *Cla* or *Esp* mutations were still on the same plasmid as the intact *pvuIIR* gene and that the pACYC184 derivative carrying wild-type *pvuIIC* retained the *Bgl*II deletion in *pvuIIR* (not shown; see Fig. 1A).

The predicted *pvuIIC* protein resembles a DNA-binding protein. The predicted amino acid sequence of *pvuIIC* was compared with those in the NBRF-PIR data base, using the program FASTA (35), and the significant matches included some known DNA-binding proteins. One alignment stood out, having a 16-amino-acid stretch with no gaps, 11 identities and 5 conservative substitutions, and a match score nearly 10 standard deviations above the mean initial score

TABLE 2. Effect of *pvuIIC* in *trans*

E. coli JM107MA2 carrying:		Restriction activity	
Plasmid 1[a]	Plasmid 2	In vitro[b]	In vivo[c]
p*Pvu*M1.9	pACYC184	≤0/≤0	$6.3 \pm 0.9 \times 10^{12}$
	p*Pvu*CYC2.6	—/—	$1.2 \pm 0.1 \times 10^{12}$
	Ratio		5.3×10^{0}
p*Pvu*RM3.4	pACYC184	3/≥5	$5.0 \pm 0.8 \times 10^{2}$
	p*Pvu*CYC2.6	4/≥5	$2.6 \pm 0.5 \times 10^{3}$
	Ratio	−1/0	1.9×10^{1}
p*Pvu*RM3.4-*Cla*35	pACYC184	≤0/1	$5.9 \pm 0.4 \times 10^{12}$
	p*Pvu*CYC2.6	4/≥5	$5.7 \pm 0.3 \times 10^{5}$
	Ratio	≤−4/≤−4	1.0×10^{7}
p*Pvu*RM3.4-*Cla*93	pACYC184	1/2	$3.3 \pm 0.3 \times 10^{12}$
	p*Pvu*CYC2.6	≥5/≥5	$1.1 \pm 0.4 \times 10^{5}$
	Ratio	≤−4/≤−3	3.0×10^{7}
p*Pvu*RM3.4-*Esp*19	pACYC184	≤0/1	$4.1 \pm 0.6 \times 10^{12}$
	p*Pvu*CYC2.6	4/≥5	$3.4 \pm 0.2 \times 10^{5}$
	Ratio	≤−4/≤−4	1.2×10^{7}
p*Pvu*RM3.4-*Esp*33	pACYC184	≤0/1	$4.0 \pm 0.6 \times 10^{12}$
	p*Pvu*CYC2.6	4/≥5	$1.1 \pm 0.1 \times 10^{5}$
	Ratio	≤−4/≤−4	3.6×10^{7}

[a] Described in Table 1, footnote *a*. The same cells also contained either pACYC184 or its derivative containing the *pvuIIC* ORF.

[b] The dilution, as a power of 10, at which restriction activity could be detected in crude extracts of 3.6-fold-concentrated overnight cultures. The left value indicates the maximum dilution at which 1 μg of λ DNA was completely digested in 1 h at 37°C. A value of 5 thus corresponds to about 3.6×10^5 U/liter of culture. The right value indicates the maximum dilution at which all final restriction fragments were present, even if the digestion was incomplete. —, Could not be measured.

[c] Restriction activity was measured as described in Table 1, footnote *d*, by plating unmodified λ*vir* on the various strains.

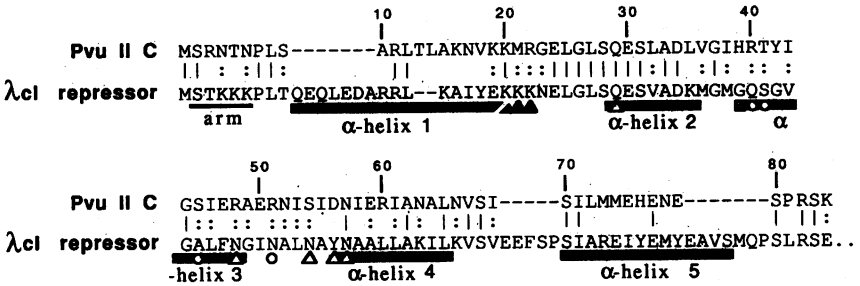

FIG. 2. Alignment of the *pvuIIC* and λ*c*I amino acid sequences. Some structural features of λ*c*I are indicated. Symbols: ○, amino acids that make base-specific hydrogen bonds to the operator; △, contacts to DNA phosphates; ▲, phosphate contacts via an ion pair; ■, an amino acid implicated in making a favorable contact to RNA polymerase.

for the entire data base. This was the match to the *c*I repressor of bacteriophage λ; the alignment is shown in Fig. 2. The match involves the DNA-binding amino portion (amino acids 1 to 92) of the 236-amino-acid λ repressor. A central element in the recognition of DNA sequences by repressor proteins is a hinged structure including two alpha helices (reviewed in references 7 and 8). All of the rules for formation of a helix-turn-helix structure are satisfied by the *pvuIIC* sequence, and its AAC score is 0.78 (see reference 7). The AAC score is a measure of homology to known helix-turn-helix motifs, and scores below 0.80 indicate strong candidates for adopting helix-turn-helix structures. The greatest similarity between the *pvuIIC* and λ*c*I sequences is from the carboxy end of helix 1 through helix 2 and into the turn. This similarity includes several amino acids thought to contact DNA phosphate groups directly or via ion pairs (27, 34; Fig. 2). In contrast, none of the amino acids thought to make base-specific hydrogen bonds in λ*c*I are identical to the aligned amino acids in the *pvuIIC* sequence. There is also limited similarity between the amino termini of the two sequences, and in λ*c*I the amino "arm" wraps around the DNA to make minor-groove contacts (33). The similarity to λ*c*I and the AAC score are consistent with the possibility that *pvuIIC* codes for an activator or repressor of transcription.

ORFs homologous to *pvuIIC* exist in other RMS2s. The occurrence of a *trans*-acting regulatory gene in the *Pvu*II RMS2 led us to search for other such genes that might have been overlooked. The sequences of 15 RMS2s were analyzed for the presence of ORFs >60 codons in length, and all such unattributed ORFs were compared with data in the NBRF-PIR data base, using the program FASTA. We found three other RMS2s with possible DNA-binding protein ORFs; *Bam*HI, *Eco*RV, and *Sma*I. The *Bam*HI ORF was identified previously by Nathan and Brooks (30). The *Sma*I ORF was identified in a published report of the *Sma*I DNA sequence as having unknown function, though a regulatory role was mentioned as a possibility (21).

We found that all four of these putative regulatory proteins have strikingly related sequences (Fig. 3). On the basis of the alignment between *pvuIIC* and λ*c*I (Fig. 2), the greatest similarity between the four RMS2 C proteins would be in the DNA-recognizing helix-turn-helix motif. A consensus sequence for that region (Fig. 3) was used to search the NBRF-PIR data base with FASTA. The two highest match scores corresponded to the immunity repressor protein from the *Bacillus* phage ϕ105 (15, 17), and a 13-kDa *Bacillus subtilis* protein believed to be a repressor for some stage in sporulation (18). These scores were, respectively, 7.7 and

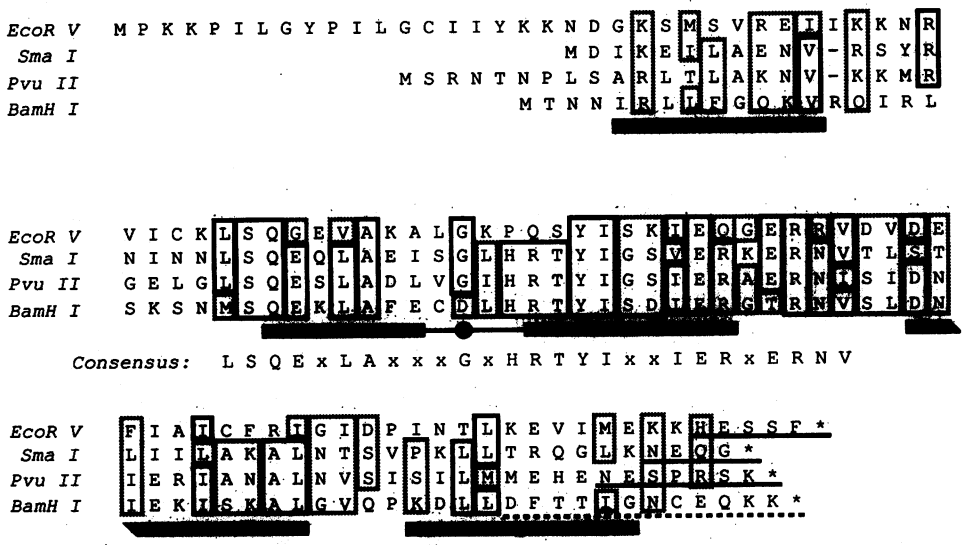

FIG. 3. Alignment of the four putative RMS2 regulatory proteins. Black boxes indicate identities in three or more of the sequences; gray boxes indicate conservative substitutions. The aligned structural features from Fig. 2 are indicated. Underlined amino acids at the carboxy termini overlap the respective REase genes.

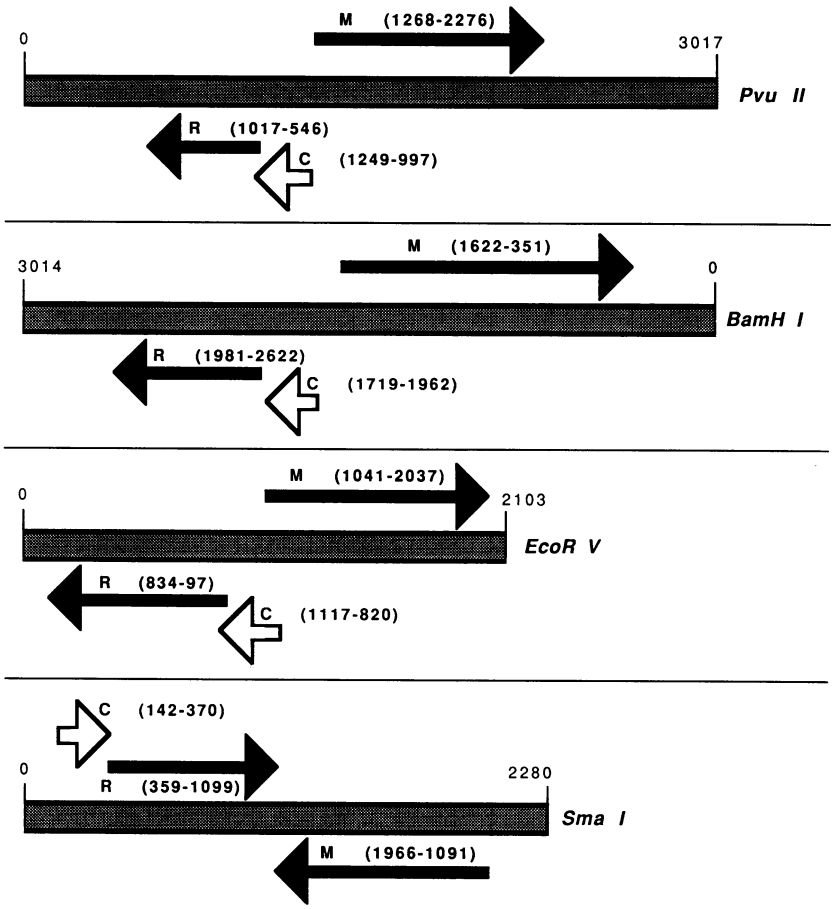

FIG. 4. Boundaries and orientations of the ORFs from four RMS2s. Numbering is from the original references.

5.5 standard deviations above the mean initial match score for the entire data base.

In addition to showing strong sequence similarity at the amino acid level, the four C ORFs have the same relative positions in their RMS2s. All four are codirectional with, and precede, the REase genes (Fig. 4). Three of the four clearly overlap the REase gene (Fig. 5); the fourth (*Bam*HI) overlaps the REase ORF, though *Bam*HI REase actually initiates further downstream (8a). No such common orientation exists relative to the MTase genes: the *Sma*I RMS2 is organized convergently, unlike the other three, and only the *ecoRVC* gene overlaps the MTase as well as the REase.

Relatedness of the four RMS2s that contain C ORFs. The strong similarity between the C proteins is in marked contrast to the MTase or REase proteins. The relatedness between two protein sequences can be assessed by the program RDF2 (35), which compares the optimal alignment between two proteins with optimal alignments between one of the proteins and random scramblings of the other. This approach minimizes the risk that two proteins will be judged as being related when they actually just have similar amino acid compositions. This analysis was applied to the REase, C, and MTase proteins for *Pvu*II, *Bam*HI, *Eco*RV, and *Sma*I. The results (Table 3) reveal three points. First, the four REase proteins are essentially unrelated to one another (though some limited similarities have been noted between the *Bam*HI and *Pvu*II REases; 43a). Second, the MTases with the exception of *Eco*RV show significant interrelatedness. The *Eco*RV MTase generates N^6-methyladenosine (5), while the other three generate N^4-methylcytosine (8a, 10, 17a). Third, even the similarity between the three N^4-methylcytosine MTases is far lower than between the C proteins.

The results from Table 3 suggest that the C genes might have spread independently of the MTase or REase genes. The codon compositions of the four RMS2s, on the other hand, reveal that all 12 REase, C, and MTase genes resemble one another more than they resemble the chromosomes of their respective host organisms. In chromosomal genes, the third position of codons has been shown to be most strongly correlated with overall genomic composition (29). For the REase, C, and MTase genes, the third-position compositions range from 22 to 32%, while the G+C contents of the chromosomal DNA in the host organisms range from 37 to 58% (32).

DISCUSSION

The *Pvu*II RMS2 appears to code for a regulatory protein. The regulation of expression of RMS2s is not well understood, though elements of the control process have been identified for a few systems. In theory, RMS2s could be regulated by passive mechanisms. For example, most RMS2 REases are homodimeric, so that subunits must accumulate before active enzyme appears, while most RMS2 MTases are active as monomers (28). This may result in a lag

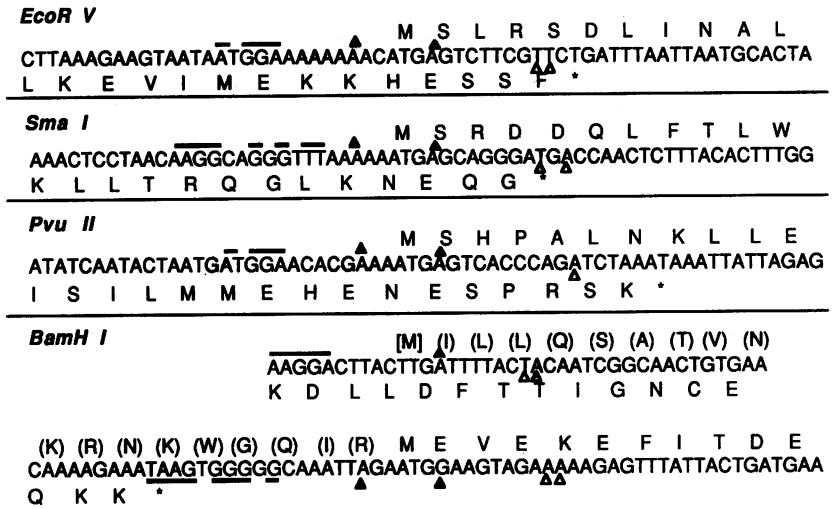

FIG. 5. Sequences of the C and REase ORF overlaps. The sense strand is shown 5′ to 3′; in each case, the C translation is underneath and the REase translation is above. Amino acids in parentheses in the *Bam*HI REase have not been found in the mature protein (8a). Features associated with *E. coli* translation initiators are indicated (20). These include homology to the 3′ end of 16S rRNA (▬), a preference for A at −3 and for a purine at +4 (▲), and a TTAA sequence at about +10 (△).

between the appearance of MTase and REase activities. An alternative passive approach is illustrated by the *Pst*I RMS2, where the stronger MTase promoter overlaps and interferes with the weaker, oppositely oriented REase promoter (45). The *Eco*RV RMS2 is proposed to have some translational regulation via a complex mRNA secondary structure (5). Only one previous report has implicated an RMS2 control mechanism involving a specific regulatory protein (30).

We report here the existence of a control factor for the *Pvu*II RMS2 that appears to be a sequence-specific DNA-binding protein (repressor, activator, or both). Confirmation awaits its overexpression, purification, and in vitro assay, but several facts support the existence and role of this protein. First, it is coded for by an ORF with features in its base sequence that correspond to those in authentic genes (43a), including potential promoter and translation initiation sequences. Second, *Pvu*II restriction is abolished by four different sequence alterations within this ORF (and outside of both the *pvuIIM* and *pvuIIR* ORFs). Third, this ORF appears to specify a *trans*-acting product. While the construct used to supply the ORF in *trans* also carries *pvuIIM*, our earlier studies showed that *pvuIIR* is poorly expressed as an independent subclone (*Hin*dIII-*Eco*RI; see Fig. 1A) when *pvuIIM* (*Hin*dIII-*Hin*dIII) is supplied in *trans* (4). Fourth, this ORF has substantial and striking sequence similarity to authentic repressor genes, including a probable helix-turn-helix motif. The most straightforward explanation of these results is that the ORF codes for a transcriptional regulatory protein. We have accordingly named this ORF *pvuIIC* (*C* for controller).

The data presented here suggest two things about the *pvuIIC* protein. First, since the product of p*Pvu*CYC2.6 is active (Table 2), the carboxy terminus might not play a major role in the mature protein (see Fig. 1B). This would be

TABLE 3. Relatedness among four RMS2s[a]

Protein	Reference sequence	Test sequence			
		*Bam*HI	*Pvu*II	*Sma*I	*Eco*RV
REase	*Bam*HI	1,074 (81.6)	51 (3.5)	35 (0.5)	28 (−0.9)
	*Pvu*II	31 (0.1)	827 (49.8)	38 (0.5)	36 (1.2)
	*Sma*I	35 (0.0)	38 (0.5)	1,265 (61.4)	58 (1.9)
	*Eco*RV	28 (−0.5)	36 (0.4)	58 (1.6)	1,255 (60.9)
C	*Bam*HI	385 (31.0)	136 (18.4)	127 (24.5)	73 (5.2)
	*Pvu*II	136 (20.7)	358 (45.7)	151 (22.0)	83 (11.3)
	*Sma*I	127 (11.0)	151 (13.7)	333 (47.8)	78 (7.1)
	*Eco*RV	73 (8.9)	83 (8.9)	78 (11.1)	500 (32.5)
MTase	*Bam*HI	2,196 (118.1)	166 (11.3)	79 (6.6)	34 (0.3)
	*Pvu*II	115 (17.2)	1,700 (128.2)	82 (7.6)	40 (0.8)
	*Sma*I	79 (3.7)	82 (4.1)	1,549 (119.9)	43 (1.8)
	*Eco*RV	34 (0.4)	40 (0.4)	43 (2.7)	1,553 (226.3)

[a] The REase, C, and MTase sequences are compared among themselves. The reference sequence is held constant, and the test sequence is randomized. Ten randomizations were done per comparison. In each case, the first number is the alignment score of the original sequences, and the value in parentheses is the number of standard deviations this alignment is above the average of randomized alignments. These values were generated by using the computer program RDF2 (35).

consistent with the greatly reduced similarity between the four C sequences in their carboxy-terminal regions (Fig. 3). Second, the null phenotype of the *Esp* mutants suggests that the altered leucine is in a critical region of the protein. From the homology to λcI (Fig. 2), this leucine should be within helix 2, the first helix of the helix-turn-helix motif. The leucine changes should not affect the formation of an alpha helix at that point but would change the spacing of conserved features.

It should be noted, however, that the effects of *pvuIIC* have been studied in a heterologous background. *E. coli* and *Proteus vulgaris* are both among the family *Enterobacteriaceae*, and it would be surprising to find major differences between these two backgrounds in the behavior of the *Pvu*II genes, but this has yet to be confirmed.

The *pvuIIC* family. Homologous RMS2 regulatory proteins are apparently produced by such genetically distant prokaryotes as *P. vulgaris*, *E. coli*, and *Serratia marcescens* (gram-negative enteric species) and *Bacillus amyloliquefaciens* (a gram-positive sporulator). Out of 16 RMS2 sequences analyzed, a quarter contained ORFs of the *pvuIIC* family, and two of these (*pvuIIC* and *bamHIC*) have been implicated as regulators (30; this report). Since this work was done, another group has informed us of an ORF of the *pvuIIC* family in association with an RMS2 that appears to generate 5-methylcytosine (*Eco*72I; 23a). Thus all three groups of MTases (5-methylcytosine, N^4-methylcytosine, and N^6-methyladenosine) have been found in association with these ORFs. With over 1,300 RMS2s now identified (42), it seems likely that many more will be found to contain genes in the same family as *pvuIIC*.

The *pvuIIC* family does not appear to be part of any previously identified family of regulatory genes. Aside from the three other C proteins, the highest FASTA match scores with *pvuIIC* included the *c*I repressor from phage λ, the analogous repressor from *Bacillus* phage φ105 (15, 17), and an apparent repressor of sporulation coded for by a *B. subtilis* chromosomal gene (18). The latter two also matched well to a consensus sequence from the four C proteins and could be considered as members of the *pvuIIC* family. The *pvuIIC* product does not appear to belong to the LysR family (22), the OmpR family (summarized in reference 39), the XylS family (38), or the Mnt family (6).

The genetic relationship among the C-producing RMS2s is unclear. The C proteins are far more similar to one another than are their associated MTase or REase proteins (Table 3). Possibly C genes have been independently recruited into several RMS2s. This view would be consistent with evidence that MTase and REase genes may move independently of one another in forming new RMS2s (25, 42).

Regulation of the *Pvu*II RMS2. Regulatory studies of the *Pvu*II genes are just beginning, but there are already several basic observations to explain. Our working hypothesis is that the role of *pvuIIC* is to mediate temporal regulation. When the *Pvu*II genes enter a new cell, transcription of *pvuIIM* would take place immediately, but transcription of *pvuIIR* would occur solely or predominantly from the *pvuIIC* promoter. Eventually, *pvuIIC* product would accumulate and would increase transcription either from its own promoter or from a promoter within *pvuIIC*. The behavior of the nonrestricting *Cla* mutants tends to favor the possibility of a transcript initiating within *pvuIIC*. Both *Cla* mutations result in *pvuIIC* frameshifts and might be expected to have a polar effect on *pvuIIR* transcription. In fact, these mutations are complemented by p*Pvu*CYC2.6, suggesting that *pvuIIR* transcription may initiate downstream of the *Cla* I site (see Fig. 1).

This model is consistent with the known facts and makes testable predictions. One such prediction is that when the intact *Pvu*II RMS2 is introduced into cells containing p*Pvu*CYC2.6, active REase should accumulate more rapidly than in cells lacking p*Pvu*CYC2.6 because of the preexisting pool of C protein.

While our data could be explained by transcriptional regulation alone, there may also be translation-level controls. All but one of the C genes found to date overlap their respective REase ORFs by several codons (Fig. 5). Polycistronic transcripts, carrying both the C and REase ORFs, are a major in vivo transcription product of the *Sma*I genes (21). Such polycistronic mRNAs might initiate REase translation inefficiently because of the overlap with C, and REase translation may be more efficient from transcripts initiated within C.

The *pvuIIC* operator. Wharton and Ptashne (47) demonstrated the importance of the solvent-exposed amino acids on helix 3 for DNA sequence recognition. They replaced just those exposed amino acids in phage 434 repressor (1) with the analogous amino acids from phage P22 repressor and showed that the altered repressor had the same sequence specificity as P22 repressor. On the basis of the alignments shown in Fig. 2 and 3, the predicted solvent-exposed amino acids of helix 3 in the *pvuIIC* and *smaIC* proteins are identical. It will be interesting to determine whether the *pvuIIC* and *smaIC* proteins recognize the same base sequence. We searched for a common 10- to 20-bp sequence within the *Pvu*II and *Sma*I RMS2s. The *pvuIIC* candidate, 5′-GATCAGCTAAGCTTTC-3′, is within the *pvuIIC* gene and contains two symmetrically disposed occurrences of AGCT (Fig. 1A), the core of the *Pvu*II recognition sequence (CAGCTG; 19). The sequence present in *Sma*I, which lies outside the C gene, is 5′-GATCGTTAAAGCTTTC-3′. These sequences both contain a GATC substrate site for the Dam MTase, and Dam MTase is produced by *P. vulgaris* and *S. marcescens* as well as by *E. coli* (2, 9). The *Esp* mutations of *pvuIIC* occur within that common sequence (Fig. 1A), adding or deleting 3 bp (TAA). One would expect operator mutations to be *cis* dominant, while the *Esp* mutants are both recessive (Table 2). The true nature of the putative operator has yet to be determined by overexpressing *pvuIIC* and using the protein in footprinting studies.

ACKNOWLEDGMENTS

We greatly appreciate the provision of unpublished data, helpful suggestions, and critical reviews of the manuscript from Ashok S. Bhagwat and Joan C. Dunbar of Wayne State University; Joan E. Brooks, David Landry, Elisabeth A. Raleigh, and Jack S. Benner III of New England BioLabs, Inc.; and Arvydas A. Janulaitis and Viktoras Butkus of Fermentas (Lithuania).

This report is based on work supported by the National Science Foundation under grant DMB-8818673.

REFERENCES

1. **Anderson, J. E., M. Ptashne, and S. C. Harrison.** 1985. A phage repressor-operator complex at 7Å resolution. Nature (London) **316**:596–601.
2. **Barbeyron, T., K. Kean, and P. Forterre.** 1984. DNA adenine methylation of GATC sequences appeared recently in the *Escherichia coli* lineage. J. Bacteriol. **160**:586–590.

2a. **Benner, J. S., III (New England BioLabs, Inc.).** Personal communication.

3. **Blumenthal, R. M.** 1986. The *Pvu* II restriction-modification system: cloning, characterization, and use in revealing an *E. coli*

barrier to certain methylases or methylated DNAs, p. 225–245. *In* J. G. Chirikjian (ed.), Gene amplification and analysis, vol 5. Elsevier, New York.

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. **Bougueleret, L., M. Schwarzstein, A. Tsugita, and M. Zabeau.** 1984. Characterization of the genes coding for the *EcoR* V restriction and modification system of *Escherichia coli*. Nucleic Acids Res. **12**:3659–3676.
6. **Bowie, J. U., and R. T. Sauer.** 1990. TraY proteins of F and related episomes are members of the Arc and Mnt repressor family. J. Mol. Biol. **211**:5–6.
7. **Brennan, R. G., and B. W. Matthews.** 1989. The helix-turn-helix DNA binding motif. J. Biol. Chem. **264**:1903–1906.
8. **Brennan, R. G., and B. W. Matthews.** 1989. Structural basis of DNA-protein recognition. Trends Biochem. Sci. **14**:286–290.

8a. **Brooks, J. E.** Personal communication.

9. **Brooks, J. E., R. M. Blumenthal, and T. R. Gingeras.** 1983. The isolation and characterization of the *Escherichia coli* DNA adenine methylase *dam* gene. Nucleic Acids Res. **11**:837–851.
10. **Butkus, V., S. Klimasauskas, L. Petrauskiene, Z. Maneliene, A. Lebionka, and A. A. Janulaitis.** 1987. Interaction of *Alu* I, *Cfr6* I and *Pvu* II restriction-modification enzymes with substrates containing either N4-methylcytosine or 5-methylcytosine. Biochim. Biophys. Acta **909**:201–207.
11. **Cashel, M., and K. E. Rudd.** 1987. The stringent response, p. 1410–1438. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
12. **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.
13. **Christman, M. F., R. W. Morgan, F. S. Jacobson, and B. N. Ames.** 1985. Positive control of a regulon for defenses against oxidative stress and some heat-shock proteins in *Salmonella typhimurium*. Cell **41**:753–762.
14. **Chung, C. T., and R. H. Miller.** 1988. A rapid and convenient method for the preparation and storage of competent bacterial cells. Nucleic Acids Res. **16**:3580.
15. **Cully, D. F., and A. J. Garro.** 1985. Nucleotide sequence of the immunity region of *Bacillus subtilis* bacteriophage ϕ105: identification of the repressor gene and its mRNA and protein products. Gene **38**:153–164.
16. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. A manual for genetic engineering: advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Dhaese, P., J. Seurinck, B. DeSmet, and M. van Montagu.** 1985. Nucleotide sequence and mutational analysis of an immunity repressor gene from *Bacillus subtilis* temperate phage ϕ105. Nucleic Acids Res. **13**:5441–5455.

17a. **Dunbar, J. C.** Personal communication.

18. **Gaur, N. K., E. Dubnau, and I. Smith.** 1986. Characterization of a cloned *Bacillus subtilis* gene that inhibits sporulation in multiple copies. J. Bacteriol. **168**:860–869.
19. **Gingeras, T. R., L. Greenough, I. Schildkraut, and R. J. Roberts.** 1981. Two new restriction endonucleases from *Proteus vulgaris*. Nucleic Acids Res. **9**:4525–4536.
20. **Gold, L., and G. Stormo.** 1987. Translational initiation, p. 1302–1307. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
21. **Heidmann, S., W. Seifert, C. Kessler, and H. Domdey.** 1989. Cloning, characterization and heterologous expression of the *Sma* I restriction-modification system. Nucleic Acids Res. **17**: 9783–9796.
22. **Henikoff, S., G. W. Haughn, J. M. Calvo, and J. C. Wallace.** 1988. A large family of bacterial activator proteins. Proc. Natl. Acad. Sci. USA **85**:6602–6606.
23. **Ingraham, J.** 1987. Effect of temperature, pH, water activity, and pressure on growth, p. 1543–1554. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.

23a. **Janulaitis, A., and V. Butkus.** Personal communication.

24. **Jessee, J., and F. Bloom.** 1988. DH5αMCR: new competent cells for cloning methylated DNA. Focus **10**:69.
25. **Kaszubska, W., C. Aiken, C. D. O'Connor, and R. I. Gumport.** 1989. Purification, cloning and sequence analysis of *Rsr* I DNA methyltransferase: lack of homology between two enzymes, *Rsr* I and *Eco*R I, that methylate the same nucleotide in identical recognition sequences. Nucleic Acids Res. **17**:10403–10425.
26. **Kruger, D. H., and T. A. Bickle.** 1983. Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. Microbiol. Rev. **47**:345–360.
27. **Lewis, M., A. Jeffrey, J. Wang, R. Ladner, M. Ptashne, and C. O. Pabo.** 1982. Structure of the operator-binding domain of bacteriophage λ repressor: implications for DNA recognition and gene regulation. Cold Spring Harbor Symp. Quant. Biol. **47**:435–440.
28. **Modrich, P., and R. J. Roberts.** 1982. The type-II restriction and modification enzymes, p. 109–154. *In* S. M. Linn and R. J. Roberts (ed.), Nucleases. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Muto, A., and S. Osawa.** 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. Proc. Natl. Acad. Sci. USA **84**:166–169.
30. **Nathan, P. D., and J. E. Brooks.** 1988. Characterization of clones of the *Bam*H I methyltransferase gene. Gene **74**:35–36.
31. **Neidhardt, F. C., and R. A. van Bogelen.** 1987. Heat shock response, p. 1334–1345. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
32. **Normore, W. M.** 1973. Guanine-plus-cytosine (GC) composition of the DNA of bacteria, fungi, algae and protozoa, p. 585–740. *In* A. I. Laskin and H. A. Lechevalier (ed.), Handbook of microbiology, vol. 2. CRC Press, Cleveland.
33. **Pabo, C. O., W. Krovatin, A. Jeffrey, and R. T. Sauer.** 1982. The N-terminal arms of λ repressor wrap around the operator DNA. Nature (London) **298**:441–443.
34. **Pabo, C. O., and M. Lewis.** 1982. The operator-binding domain of λ repressor: structure and DNA recognition. Nature (London) **298**:443–447.
35. **Pearson, W. R.** 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. **183**:63–98.
36. **Price, C., and T. A. Bickle.** 1986. A possible role for DNA restriction in bacterial evolution. Microbiol. Sci. **3**:296–299.
37. **Raleigh, E. A., and G. Wilson.** 1986. *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. USA **83**:9070–9074.
38. **Ramos, J. L., F. Rojo, L. Zhou, and K. N. Timmis.** 1990. A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and the *Escherichia coli* AraC activators. Nucleic Acids Res. **18**:2149–2152.
39. **Roberts, R. J.** 1990. Restriction enzymes and their isoschizomers. Nucleic Acids Res. **18**:r2331–r2365.
40. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
41. **Smith, H. O., and S. V. Kelly.** 1984. Methylases of the type II restriction-modification systems, p. 39–71. *In* A. Razin, H. Cedar, and A. D. Riggs (ed.), DNA methylation: biochemistry and biological significance. Springer-Verlag, New York.
42. **Stephenson, F. H., B. T. Ballard, H. W. Boyer, J. M. Rosenberg, and P. J. Greene.** 1989. Comparison of the nucleotide and amino acid sequences of the *Rsr* I and *Eco*R I restriction endonucleases. Gene **85**:1–13.

43. **Szybalski, W., R. M. Blumenthal, J. E. Brooks, S. Hattman, and E. A. Raleigh.** 1988. Nomenclature for bacterial genes coding for class-II restriction endonucleases and modification methyltransferases. Gene **74**:279–280.

43a. **Tao, T., and R. M. Blumenthal.** Submitted for publication.

44. **Tao, T., J. Walter, K. J. Brennan, M. M. Cotterman, and R. M. Blumenthal.** 1989. Sequence, internal homology and high-level expression of the gene for a DNA-(cytosine *N*4)-methyltransferase, M · *Pvu* II. Nucleic Acids Res. **17**:4161–4175.

45. **Walder, R. Y., J. A. Walder, and J. E. Donelson.** 1984. The organization and complete nucleotide sequence of the *Pst* I restriction-modification system. J. Biol. Chem. **259**:8015–8026.

46. **Walker, G. C.** 1987. The SOS response of *Escherichia coli*, p. 1346–1357. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.

47. **Wharton, R., and M. Ptashne.** 1985. Changing the binding specificity of a repressor by redesigning an α-helix. Nature (London) **316**:601–605.

48. **Wilson, G. G.** 1988. Cloned restriction-modification systems—a review. Gene **74**:281–289.

49. **Winans, S. C., P. R. Ebert, S. E. Stachel, M. P. Gordon, and E. W. Nester.** 1986. A gene essential for *Agrobacterium* virulence is homologous to a family of positive regulatory loci. Proc. Natl. Acad. Sci. USA **83**:8278–8282.

50. **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.