JOURNAL OF BACTERIOLOGY, Feb. 1997, p. 964–967 Vol. 179, No. 3
0021-9193/97/$04.00+0

Characterization of the Intergenic Region Which Regulates the *Msp*I Restriction-Modification System

SUBHENDU SOM AND STANLEY FRIEDMAN*

Department of Pharmacology, State University of New York Health Science Center, Brooklyn, New York 11203

Received 21 March 1996/Accepted 13 November 1996

The 110-bp intergenic region between *mspIM* and *mspIR*, the genes encoding the *Msp*I modification (M.*Msp*I) and restriction (R.*Msp*I) enzymes, respectively, was fused, in both orientations, with *lacZ*. Expression of a single-copy *mspIM-lacZ* fusion is more than 400-fold stronger than expression of an *mspIR-lacZ* fusion. M.*Msp*I in *trans* represses expression of the *mspIM-lacZ* fusion by binding to the DNA but does not affect expression of the *mspIR-lacZ* fusion. Transcription start sites of the genes were identified, and a set of nonoverlapping promoters was assigned. DNase I footprinting showed that M.*Msp*I binds to a site within the intergenic region that includes only the *mspIM* regulatory elements.

Prokaryotic type II restriction-modification (RM) systems consist of a restriction endonuclease (R) and a separate modification methylase (M) encoded by two different open reading frames. The two genes are almost always located close to one another (17), although their orientations vary from system to system. Appropriate regulation of the RM genes is important for the survival of the host. Such regulation has recently been demonstrated for several RM systems, including *Bam*HI (4), *Pvu*II (14), and *Eco*RII (2, 12, 13).

*Msp*I (recognizes 5′-CCGG-3′) is one of the most widely studied RM systems. It has been cloned (5, 9, 15), and M.*Msp*I has also been overproduced in *Escherichia coli* (1). Although sequences of about 3.0 kb of DNA containing the *mspIM* and *mspIR* genes have been reported (5), the transcriptional start sites have not been determined nor have regulatory elements been defined. Since the two genes separated by an intergenic region of 110 bp are transcribed divergently from complementary strands (5), there could be (i) a common region containing overlapping promoter sequences in opposite orientations, (ii) promoter sequences for one gene within the transcribed region of the other, or (iii) total separation of the two promoters. In a previous report we proposed that expression of the *mspIM* gene is autoregulated, since M.*Msp*I binds to the DNA containing this intergenic region (12). If the promoter sequences overlapped, such binding might also interfere with the transcription of *mspIR*.

In this study we characterize the properties of the *mspI* intergenic region by (i) studying its fusions with *lacZ* in *E. coli* and the effect of M.*Msp*I in *trans* on such fusions, (ii) mapping the transcription start sites for both the *mspIM* and *mspIR* genes, and (iii) determining the binding sites for the M.*Msp*I protein.

Plasmids. A low-copy-number plasmid expressing M.*Msp*I, pACM.MspI, was constructed by cloning the 1,555-bp *Eco*RI-*Sau*3A fragment from the multicopy plasmid pM.MspI-7 (12) into the *Dra*I-*Bam*HI site of pACYC177 after end filling the *Eco*RI site. Plasmid pRS415 (11) is a pBR322 derivative fusion vector which allows expression of β-galactosidase (β-Gal) from operon fusions created by cloning a suitable promoter 5′ to the promoterless *lacZ* gene (*lacZ*$^+$). Plasmid pBW201 (16) expresses M.*Bsu*FI, an isoschizomer of M.*Msp*I.

Fusion of the *Msp*I intergenic region to the *lacZ* structural gene. A 198-bp *Sau*3AI fragment containing the intergenic region between the *mspI* genes (Fig. 1) was inserted into pRS415 linearized by *Bam*HI. Transformed *E. coli* ER1648 [Δ(*lacZ*) *mcrA1272*::Tn*10* Δ(*mcrC-mrr*)*102*::Tn*10*] colonies bearing the recombinant plasmids displayed two different phenotypes at 25°C on Luria-Bertani plates containing ampicillin and X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside): (i) moderately dark blue colonies of normal size and (ii) tiny, intensely dark blue colonies. Restriction analysis of the plasmid minipreps revealed that the moderately dark blue colonies contained plasmids (designated as pPmspR-lacZ) in which the orientation of the insert was such that expression of *lacZ* was controlled by the *mspIR* regulatory elements. The tiny, intensely dark blue colonies bore fusion plasmids (designated as pPmspM-lacZ) in which the *lacZ* gene was expressed from the *mspIM* operon. At 37°C viability of the latter type of colonies was drastically decreased. In some cases deletion occurred within the plasmid. On the other hand, these cells could

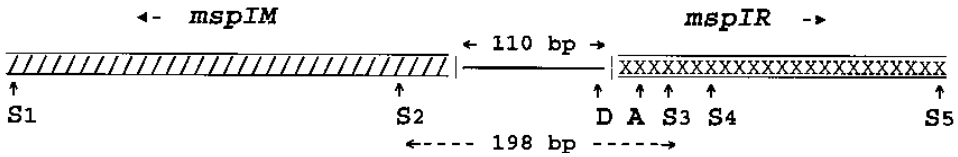

FIG. 1. Physical map of the *mspI* intergenic region. S, *Sau*3AI; D, *Dde*I; A, *Alu*I. The hatched and crossed boxes represent parts of the coding regions of the *mspIM* and *mspIR* genes, respectively. The 198-bp S2-S3 fragment was used in the fusion experiments. The S2-D fragment was used for M.*Msp*I binding and as a probe for *mspIM* RNA. The A-S3 fragment was used as primer for the primer extension reaction of *mspIR* RNA.

* Corresponding author. Mailing address: Department of Pharmacology, State University of New York Health Science Center, 450 Clarkson Ave., Brooklyn, NY 11203.

TABLE 1. Expression of β-Gal from *mspI* operon fusions[a]

Orientation of fusion	β-Gal expression (Miller units)					
	Multiple copies	Single copies	Single copy in the presence of the following plasmid[b] in *trans*:			
			pUC18	pM.MspI-7	pACM.MspI	pBW201
mspIM-lacZ$^+$	41,000	8,500	8,350	2,270	3,950	8,420
mspIR-lacZ$^+$	756[c]	19[d]	20	19.5	ND[e]	ND

[a] Cells containing the plasmid-borne multicopy fusions were grown at 30°C. λ lysogens bearing the single-copy fusions with or without other plasmids in *trans* were grown at 37°C. Assays were done as described by Miller (8).
[b] Plasmids are described in the text.
[c] Cells containing the multicopy plasmid vector pRS415 expressed 53 U.
[d] The background value obtained from ER1648 was <0.5 U.
[e] ND, not determined.

retain the plasmid pPmspM-lacZ and grow at 37°C in the presence of a functional *mspIM* gene in *trans* supplied by the compatible low-copy-number plasmid, pACM.MspI. Under similar growth conditions, expression of β-Gal in a host carrying the plasmid-borne *mspIM-lacZ*$^+$ fusion was more than 50-fold higher than that in a host carrying the *mspIR-lacZ*$^+$ fusion (Table 1). The high level of β-Gal expression is probably toxic, leading to loss of viability at 37°C. On the other hand, the low level of expression from the *mspIR-lacZ*$^+$ fusion raised some doubt about the presence of all regulatory elements within the 198-bp fragment necessary for the transcription of the *mspIR* gene. However, fusion of a much larger, 561-bp *Sau*3A fragment failed to increase expression (data not shown).

Effect of M.*Msp*I in *trans* on single-copy fusions. Both *mspIM-lacZ*$^+$ and *mspIR-lacZ*$^+$ fusions were transferred to λ by homologous recombination, and lysogens carrying a single-copy fusion were isolated as described previously (11). While lysogens carrying the *mspIR-lacZ*$^+$ fusion produced very light blue colonies on X-Gal agar plates, those containing the *mspIM-lacZ*$^+$ fusion were intensely blue in color. Lysogens carrying *mspIM-lacZ*$^+$ could be grown at 37°C and maintained without loss of viability. When lysogens were assayed for β-Gal activity, the expression from the *mspIM* promoter was found to be more than 400-fold higher than the expression from the *mspIR* promoter (Table 1). M.*Msp*I in *trans* repressed the expression from the *mspIM* promoter. Inhibition was greater (73%) when M.*Msp*I in *trans* was expressed from a multicopy plasmid. The repression is much less than the 96 to 97% repression caused by autoregulation of *ecoRIIM-lacZ*$^+$ by M.*Eco*RII (2, 12). M.*Bsu*FI, an isoschizomer of M.*Msp*I with a 56% amino acid homology (16) expressed by the plasmid pBW201, did not have any effect in *trans* on expression from the *mspIM-lacZ*$^+$ fusion. This again signifies the specificity of these promoter-protein interactions. Similar observations were made with the isoschizomers M.*Eco*RII and M.*Dcm* (13). M.*Msp*I in *trans* does not have any effect on expression initiated from the *mspIR-lacZ*$^+$ fusion, suggesting that the *mspIR* and *mspIM* regulatory elements within the intergenic region are independent.

Transcriptional start site(s) of *mspIM* and *mspIR*. The transcriptional start site(s) for the *mspIM* gene was determined by primer extension. A synthetic primer end labeled with [γ-^{32}P]ATP was annealed to 20 μg of cellular RNA extracted from ER1648(pPmspIM-lacZ) and processed for primer extension with murine leukemia virus reverse transcriptase as described by Sambrook et al. (10). Data presented in Fig. 2A reveal the presence of two bands that correspond to the 14th and 15th nucleotides upstream of the start codon of the *mspIM* gene.

We also determined the *mspIR* transcription start site by primer extension. Twenty micrograms of cellular RNA from ER1648(pPmspR-lacZ) was annealed to a ^{32}P-labeled 5′-*Sau*3AI-*Alu*I-3′ primer (see Fig. 1) and processed for primer extension as described for the methylase. Two consecutive intense bands, which correspond to nucleotides 9 and 10 upstream of the *mspIR* start codon, were detected (Fig. 2B).

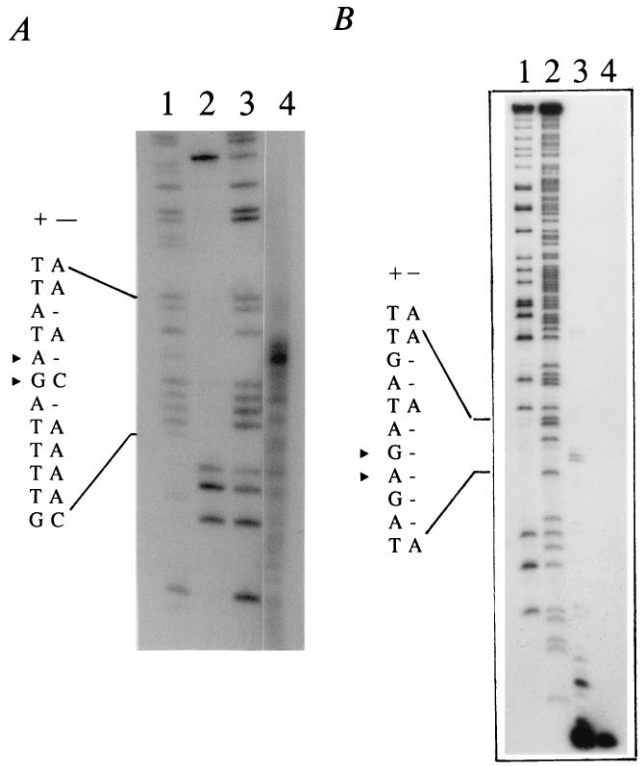

FIG. 2. Identification of the transcriptional start sites. (A) Methylase transcription start sites (arrowheads) were determined by primer extension. The 5′ end-labeled (−) strand of a specific DNA probe, GATCCAATTTACTACGA ATCAATTTCAATA, was annealed to RNA from cells containing plasmid pPmspIM-lacZ. The probe corresponded to the complement of the coding strand 43 bases downstream of the translation initiation codon. Lanes 1, 2, and 3 are standards generated by Maxam-Gilbert reactions, A > C, G, and A + G, respectively, on DNA originating at the same GATC sequence as the probe. (B) Restriction transcription start sites (arrowheads) were determined by primer extension. The 5′ end-labeled primer (see Fig. 1 and text) was annealed to RNA from cells containing pPmspIR-lacZ (lane 3) or pRS415 (lane 4). Lanes 1 and 2 represent the standards generated by Maxam-Gilbert G and G+A reactions (7), respectively. Published sequences (5) are used to fill in nucleotides shown in the plus-strand sequences.

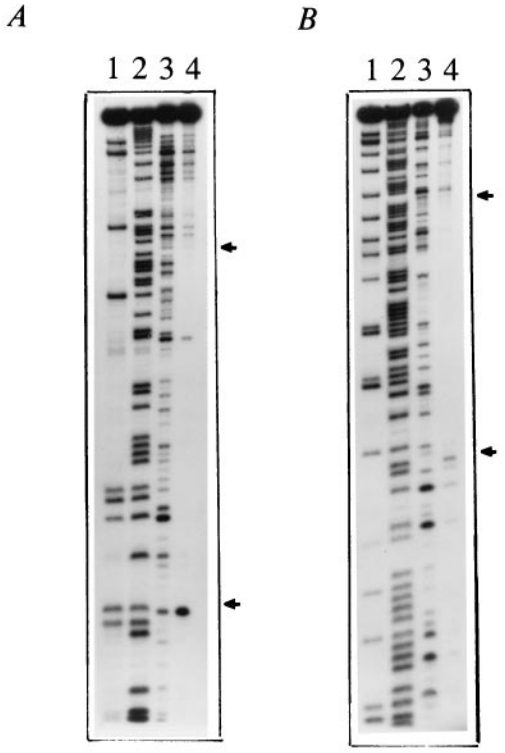

FIG. 3. DNase I footprints of M.*Msp*I bound to the *mspIM-mspIR* intergenic region. A *Sau*3AI-*Dde*I fragment containing the intergenic region was complexed with M.*Msp*I and subjected to limited digestion with DNase I (13). Autoradiograms of a 6% sequencing gel show the 5′ end labels in the template strand (A) and the nontemplate strand (B). Lanes 1, G ladder; lanes 2, G+A ladder; lanes 3, uncomplexed; lanes 4, complexed with M.*Msp*I. The regions between the arrows are protected.

DNase I footprinting of M.*Msp*I-promoter complex. The binding of the M.*Msp*I protein to *mspIM* was previously demonstrated by gel shift assay (12). Here, a *Sau*3AI-*Dde*I fragment (see Fig. 1) was subjected to complex formation with M.*Msp*I, and DNase I footprinting was performed on the protein-DNA complex as described previously (13) (Fig. 3). Sequences in both template and nontemplate strands were found to be protected.

Analysis of the data. Data from primer extension and DNase I footprinting experiments are analyzed in Fig. 4. Transcription of both genes is initiated from two consecutive nucleotides, which is not uncommon in bacteria. The distance between the 5′ ends of the two mRNAs is at least 85 nucleotides, a spacing large enough to accommodate two sets of nonoverlapping regulatory elements. The −10 and −35 promoter sequences were assigned according to the criteria of Lisser and Margalit (6). The −10 and −35 regions of the *mspIM* promoter differ from the *E. coli* consensus promoter by one base each. This explains the strength of the promoter. The 12 bases separating the transcriptional start site and the 3′ terminus of the −10 region of the methylase promoter are at the outer limit described by Harley and Reynolds (3). The −35 regions of the *mspIM* and *mspIR* genes are separated by a space of 6 nucleotides. The *mspIM* regulatory region protected from DNase I spans nucleotides −34 to +17 in the template strand and nucleotides −33 to +17 in the nontemplate strand. However, in the template strand one base at position −15 was found to be susceptible to DNase I digestion (see Fig. 3). The 50- to 51-bp region includes the −10 promoter. The entire regulatory region of *mspIR* remains unprotected. This is consistent with our finding that expression from the *mspIR-lacZ*$^+$ fusion is not affected by the presence of M.*Msp*I in *trans*. Although protein binding regions in DNA often contain inverted repeats, the protected region does not contain a long repeat sequence. However, one inverted repeat of 6 bases is found within this region (Fig. 4).

The region of 50 to 51 bp protected from DNase I digestion is unusually long. DNase I protection of large sections of DNA is usually due to DNA bending or binding of several molecules of protein. Since DNase digestion does not result in phased hypersensitivity of the protected DNA, the protection is not due to DNA bending. This pattern of protection from DNase I is similar to that found with the *ecoRIIM* promoter (13). In that case the protected region is 47 to 49 bp. In both cases protection of the template strand is offset at the 3′ end but not at the 5′ end. At present we have no evidence, other than the size of the protected region, that more than one molecule of methylase is binding with either of these enzymes. The *ecoRIIM* promoter contains an inverted repeat of 11 bp. The *mspIM* sequence has an inverted repeat of only 6 bp. Although this region of symmetry can be extended, we require data from mutants to determine if it is important for M.*Msp*I binding.

In summary, the regulatory elements for the divergently transcribed restriction and modification genes of the *msp*I RM system are located within the 110-bp intergenic region. The regulatory sequences are separate and nonoverlapping. Expression of the *mspIM* gene is autoregulated by binding of the

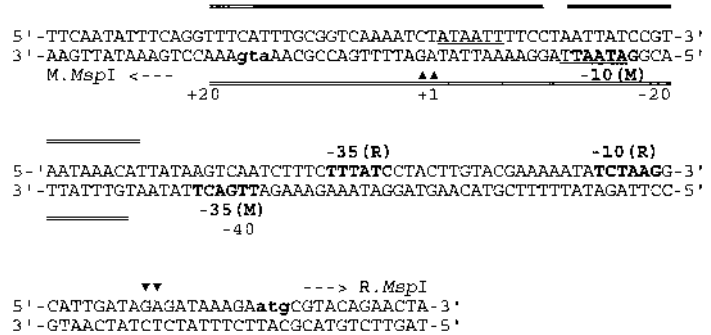

FIG. 4. Analysis of data presented in Fig. 2 and 3. The sequence of the *mspI* intergenic region presented here is from Lin et al. (5). Transcriptional start sites for the *mspIM* and *mspIR* genes are shown by upward and downward arrowheads, respectively. Initiation codons are marked with boldface lowercase letters, and the direction of transcription is indicated by broken arrows. The nucleotide positions shown are numbered relative to the *mspIM* transcription start site proximal to the initiation codon. The regions occupied by the M.*Msp*I protein are shown by a double line. The designated promoter sequences for the −10 and −35 regions are boldface, and inverted repeats are underlined.

M.*Msp*I protein to its regulatory elements within the intergenic region, but such binding does not interfere with the expression of the *mspIR* gene.

In *E. coli* a single copy of the *mspIM* promoter is 400-fold or more stronger than a single copy of the *mspIR* promoter. Low levels of activity or lack of activity of the endonuclease in *E. coli* in vivo, reported by others (5, 9), or even failure to detect its expression (15) is readily explained by our direct demonstration of the weakness of the restriction promoter. On the other hand, despite the strength of the *mspIM* promoter, little active enzyme was obtained from cells containing a multicopy plasmid expressing the *mspIM* gene from its own promoter (data not given). The discrepancy between the strength of the *mspIM* promoter when fused to *lacZ* and the poor yield of the M.*Msp*I methylase, as well as the relatively low level of repression by M.*Msp*I, could be due to the fact that these results were obtained with the system cloned in *E. coli*. The situation might be different in the natural host, *Moraxella*. Further experiments with *Moraxella* will be necessary to determine if autoregulation in the natural host for this RM system is as efficient as that of the *Eco*RII system in *E. coli.*

We thank T. A. Trautner for supplying plasmid pBW201, R. W. Simons for supplying plasmid pRS415, and E. Rawley for supplying *E. coli* ER1648. We also thank the reviewers for bringing to our attention the strong promoter in the *mspIM* gene.

REFERENCES

1. **Dubey, A. K., B. Mollet, and R. J. Roberts.** 1992. Purification and characterization of the MspI DNA methyltransferase cloned and overexpressed in *E. coli*. Nucleic Acids Res. **20:**1579–1585.
2. **Friedman, S., and S. Som.** 1993. Induction of *Eco*RII methyltransferase: evidence for autogenous control. J. Bacteriol. **175:**6293–6298.
3. **Harley, C. B., and R. P. Reynolds.** 1987. Analysis of *E. coli* promoter sequences. Nucleic Acids Res. **15:**2343–2361.
4. **Ives, C. L., P. D. Nathan, and J. E. Brooks.** 1992. Regulation of the *Bam*HI restriction-modification system by a small intergenic open reading frame, *bamHIC*, in both *Escherichia coli* and *Bacillus subtilis*. J. Bacteriol. **174:**7194–7201.
5. **Lin, P. M., C. H. Lee, and R. J. Roberts.** 1989. Cloning and characterization of the *Msp*I restriction modification system. Nucleic Acids Res. **17:**3001–3011.
6. **Lisser, S., and H. Margalit.** 1993. Compilation of *E. coli* mRNA promoter sequences. Nucleic Acids Res. **21:**1507–1516.
7. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
8. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. **Nwankwo, D. O., and G. G. Wilson.** 1988. Cloning and expression of the *Msp*I restriction and modification genes. Gene **64:**1–8.
10. **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.
11. **Simons, R. W., F. Houman, and N. Kleckner.** 1987. Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. Gene **53:**85–96.
12. **Som, S., and S. Friedman.** 1993. Autogenous regulation of the *Eco*RII methylase gene at the transcriptional level: effect of 5-azacytidine. EMBO J. **12:**4297–4303.
13. **Som, S., and S. Friedman.** 1994. Regulation of EcoRII methyltransferase: effect of mutations on gene expression and *in vitro* binding to the promoter region. Nucleic Acids Res. **22:**5347–5353.
14. **Tao, T., J. C. Bourne, and R. M. Blumenthal.** 1991. A family of regulatory genes associated with type II restriction-modification systems. J. Bacteriol. **173:**1367–1375.
15. **Walder, R. Y., C. J. Langtimm, R. Chatterjee, and J. A. Walder.** 1983. Cloning of the *MspI* modification enzyme. J. Biol. Chem. **258:**1235–1241.
16. **Walter, J., M. Noyer-Weidner, and T. A. Trautner.** 1990. The amino acid sequence of the CCGG recognizing DNA methyltransferase M.*Bsu*FI: implications for the analysis of sequence recognition by cytosine DNA methyltransferases. EMBO J. **9:**1007–1013.
17. **Wilson, G. G.** 1991. Organization of restriction-modification systems. Nucleic Acids Res. **19:**2539–2566.