© 1993 Oxford University Press

Nucleic Acids Research, 1993, Vol. 21, No. 4 905–911

The M·*Alu*I DNA-(cytosine *C5*)-methyltransferase has an unusually large, partially dispensable, variable region

Bohong Zhang, Tao Tao[1], Geoffrey G.Wilson and Robert M.Blumenthal[1]*
New England BioLabs, Inc., Beverly, MA 01915-5599 and [1]Department of Microbiology, Medical College of Ohio, Toledo, OH 43699-0008, USA

Received November 16, 1992; Revised and Accepted January 9, 1993 EMBL accession no. Z11841

ABSTRACT

The DNA methyltransferase of the *Alu*I restriction-modification system, from *Arthrobacter luteus*, converts cytosine to 5-methylcytosine in the sequence AGCT. The gene for this methyltransferase, *aluIM*, was cloned into *Escherichia coli* and sequenced. A 525-codon open reading frame was found, consistent with deletion evidence, and the deduced amino acid sequence revealed all ten conserved regions common to 5-methylcytosine methyltransferases. The *aluIM* sequence predicts a protein of M_r 59.0k, in agreement with the observed M_r, making M·*Alu*I the largest known methyltransferase from a type II restriction-modification system. M·*Alu*I also contains the largest known variable region of any monospecific DNA methyltransferase, larger than that of most multispecific methyltransferases. In other DNA methyltransferases the variable region has been implicated as the sequence-specific target recognition domain. An in-frame deletion that removes a third of this putative target-recognition region leaves the *Alu* I methyltransferase still fully active.

INTRODUCTION

The biological transmethylation of DNA from *S*-adenosyl-L-methionine (AdoMet) is carried out by two families of DNA methyltransferases (MTases). One family modifies the exocyclic amino groups of adenine or cytosine, generating *N6*-methyladenine or *N4*-methylcytosine respectively. These amino MTases have limited sequence homology to one another, although two conserved elements have been identified: FXGXG is common to virtually all enzymes that utilize AdoMet (1), and DPPY elements (actually D/N/S-PP-Y/F) are apparently unique to the amino MTases (see refs 2–4). The second family of MTases generate 5-methylcytosine (5mC) via a covalent enzyme-cytosine intermediate (5). All of the over 30 sequenced members of the 5mC MTases contain ten conserved sequence motifs that have been found whether the MTase originated in bacteria or in mammals (6–8). The existence of multispecific 5mC MTases allowed the sequence-specific target recognition domain (TRD) to be identified as the largest variable region, which lies between conserved elements VIII and IX (7, 9–10). This variable region has also been shown to be the TRD in monospecific 5mC MTases (11, 12). The degree of conservation among 5mC MTases makes this group of enzymes interesting and instructive subjects for studies on structure, function and evolution, and for this reason we have continued to isolate and characterize the genes for these enzymes.

*Alu*I is a type II restriction-modification system from the bacterium *Arthrobacter luteus*. The endonuclease (REase) cleaves to give blunt ends (AG^CT), while the MTase generates AG^{5m}CT (13–14). We report here the cloning and subcloning of the *aluIM* gene, its sequence determination, and unusual properties of the variable region in the deduced protein sequence.

MATERIALS AND METHODS

Strains and plasmids

The *Escherichia coli* K-12 strains RR1 (15), K802 (16) and JM107MA2 (17) were used. All of them are defective for the Mcr phenotype, which is essential for stable expression of MTases such as M·*Alu*I (17–20; reviewed in 21). Cloning vectors pUC8 and pUC19 (22) were used. All other plasmids referred to were products of this work.

Enzymes

Restriction endonucleases unless otherwise indicated, mung bean nuclease, and the CircumVent™ thermocycle sequencing kit were purchased from New England BioLabs (Beverly, MA). Calf intestinal alkaline phosphatase (CIAP), *Mae*II and *Mae*III were obtained from Boehringer Mannheim (Indianapolis, IN). The T7 Sequenase kit was purchased from USB Corp. (Cleveland, OH). Manufacturers' recommendations were followed throughout.

Library construction and selection of initial *Alu*I-resistant clones

Total cell DNA from *Arthrobacter luteus* (ATCC #21606) was partially digested with *Pst*I, *BamH*I, or *Bgl*II. The digests were ligated into pUC19 that had been cut with the same enzymes and treated with CIAP. The ligation products were used to transform *E.coli* strain RR1, and transformants were pooled. Plasmid DNA from these pools was digested with *Alu*I endonuclease (8u/μg

* To whom correspondence should be addressed

DNA for 1h), and then used to transform RR1. The resulting secondary transformants were pooled, and plasmid DNA was extracted and purified. These pools were again digested with excess *Alu*I and used to transform RR1. Plasmid DNA from 1ml cultures of individual transformants was then screened for resistance to *Alu*I digestion.

Subcloning

An *Alu*I-resistant plasmid clone derived from the previous step was partially digested with *Sau*3AI to an average size of 4kb. The digested material was purified by phenol-chloroform extraction and passage through a spun gel filtration column. About 0.6 μg of this DNA was ligated into 0.5 μg of *Bam*HI-digested and CIAP-treated pUC8. Three-quarters of the ligated material was used to transform *E.coli* strain JM107MA2, yielding 2122 transformants. The transformants were pooled, grown overnight in 150ml LB plus carbenicillin, and plasmid DNA was isolated by alkaline lysis. Five μg of this DNA was digested with excess *Alu*I and 3/4 was used to transform JM107MA2 after spun column purification. A total of 112 transformants were obtained, and 12 were picked for individual screening for resistance to *Alu*I digestion.

Restriction and deletion mapping of *aluIM*

Restriction maps of the subclones were determined by digestion with several endonucleases and determination of the fragment sizes on tris-acetate-EDTA (TAE) agarose gels of 0.8–3.0% concentration (see 23). Where pairs of unique restriction sites were identified, deletions were made between them by cleavage, making the ends blunt with Klenow DNA polymerase if necessary, and religation at low DNA concentrations. The resulting constructs were confirmed by restriction map analysis, and tested for resistance to digestion by *Alu*I endonuclease.

To make the in-frame deletion within the methylase gene, 25 μg of p*AluM*2.7 was digested with 16u of *BamH*I and 20u of *Xho*I for 2h at 37° in 1× One-Phor-All buffer (Pharmacia). After ethanol precipitation, 5 μg aliquots of the digested DNA were incubated with twofold serial dilutions of mung bean

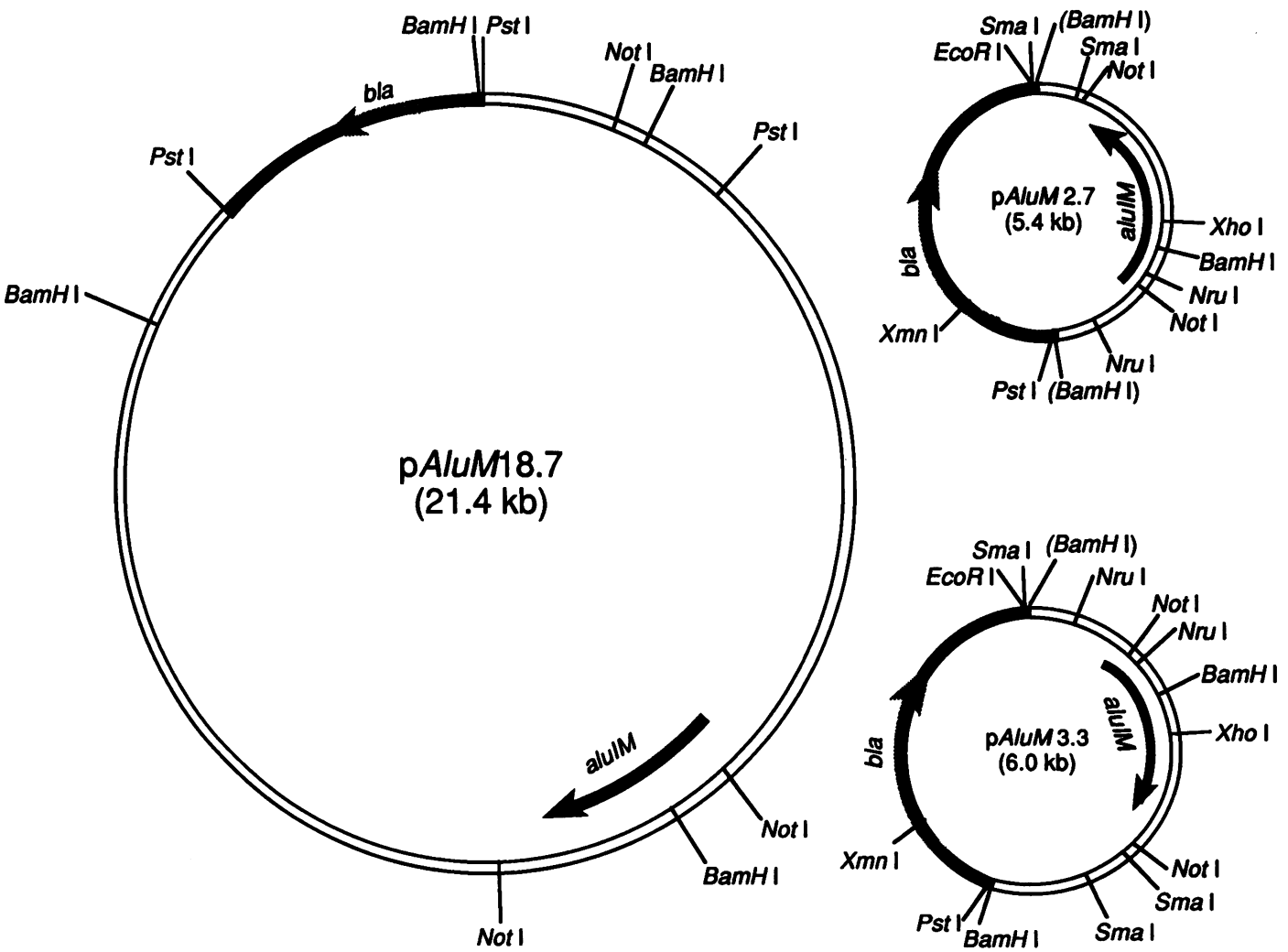

Figure 1. Plasmid clones of *aluIM*. The three plasmid clones with intact *aluIM* that were generated in the course of this study are shown to scale. Plasmid p*AluM*14.0 is the original clone, isolated from a *Pst* I-digested library of *A. luteus* chromosomal DNA and inserted into vector pUC19. The two smaller plasmids were isolated by partial digestion of p*AluM*14.0 with *Sau3A* I and ligation into *BamH* I-digested pUC8 vector. The arrows within the circular maps indicate the *aluIM* ORF. Vector sequence is indicated by filled regions, and the β-lactamase (*bla*) gene is indicated.

nuclease ranging from 1u to 0.0625u in 10 μl reactions. After 30 min at 30°, the reactions were stopped by adding sodium dodecyl sulfate to 0.01% and ethanol-precipitating the DNA. Each resuspended DNA preparation was incubated with 2u of T4 DNA ligase in 40 μl for 4 h at 16°. Prior to transformation, the ligated DNAs were digested with *Mlu*I, the site for which would have been deleted in the desired constructs.

Sequence determination and analysis

The *aluIM*$^+$ subclones were used in double-stranded form as templates for dideoxy sequence determination (24–26). Two M13 (vector-specific) primers, and 40 synthetic primers specific to the insert were used after deblocking and gel filtration, but without further purification (27). The oligonucleotides were obtained from Genosys (Woodlands,TX), or from New England BioLabs. The deduced amino acid sequence of *aluIM* gene was compared with the PIR and Genbank databases using FASTA (28), BLAST (29), and Clustal V (30) programs.

RESULTS AND DISCUSSION

Cloning the *aluIM* gene

Arthrobacter luteus DNA was partially digested with *Bam*HI, *Pst*I, or *Bgl*II, and cloned into pUC19 to generate three libraries. These libraries were then digested with R·*Alu*I in order to positively select for *aluIM* clones. This approach, introduced by Szomolanyi *et al.* (31) and now widely used, is based on the following rationale. When a plasmid carries an expressible MTase gene into a host cell, the plasmid carrying that gene is protectively methylated and becomes resistant to *in vitro* digestion with the cognate REase. As pUC19 contains 16 *Alu*I sites, this selection was expected to be robust. The digested plasmid mixtures were used to transform *Escherichia coli* strain RR1, and the process was repeated for a second cycle.

Twenty-eight colonies were picked for screening from each of the three libraries, and plasmid minipreparations were made. All 84 plasmids were partially resistant to *Alu*I digestion. The partial resistance is probably due to poor expression of *aluIM* in *E.coli*, which would not be surprising given the large phylogenetic distance between *Arthrobacter* and *Escherichia*. Partial resistance still conferred a strong selective advantage in this protocol. Plasmid p*AluM*14.0, containing a *Pst*I insert, was chosen for further work (Fig. 1). The p*AluM*14.0 insert was originally believed to be about 14 kb, but subsequent work suggests a size close to 18.7 kb (not shown).

Subcloning p*AluM*14.0

To reduce the size of the fragment carrying *aluIM*, p*AluM*14.0 was partially digested with serially diluted amounts of *Sau3A*I and the reactions yielding a 4kb average fragment size were

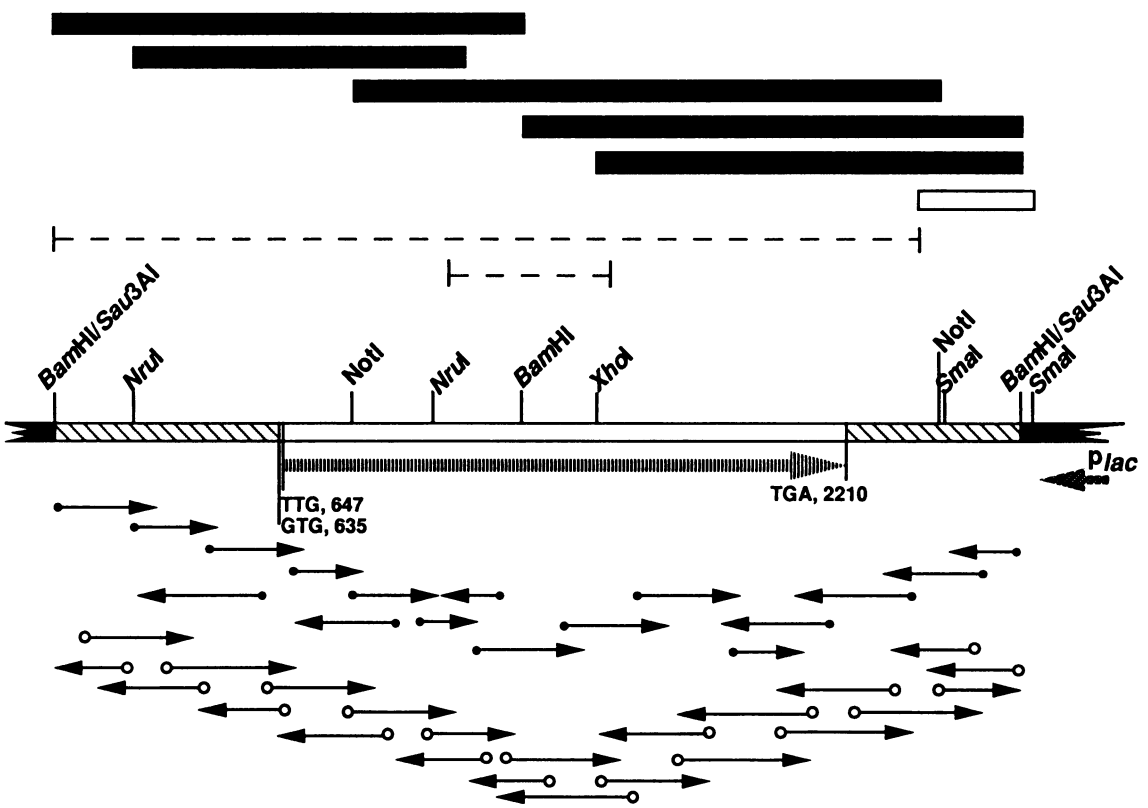

Figure 2. Characterization of the *aluIM* gene. A restriction map of the insert from plasmid p*AluM*2.7 is shown. Above the map, the boundaries of deletions that yielded *aluIM*$^+$ (open bars) and *aluIM*$^-$ plasmids (filled bars) are shown. Some of the deletions shown here were actually generated in plasmid p*AluM*3.3 (see Fig. 1). The broken lines indicate the maximum and minimum range of *aluIM* as predicted by these deletions. Synthetic oligonucleotide primers were used as shown under the restriction map. Each arrow represents the direction and length of sequence determined by using one primer. Arrows beginning with closed circles represent primers used with plasmid p*AluM*2.7 and its derivatives; arrows beginning with open circles represent primers synthesized based on the p*AluM*2.7 sequence and used to sequence p*AluM3.3*. In some cases (not shown), these primers were also used to sequence across the junction of a deletion construct. The boundaries of the *aluIM* ORF are indicated, including both possible initiation codons.

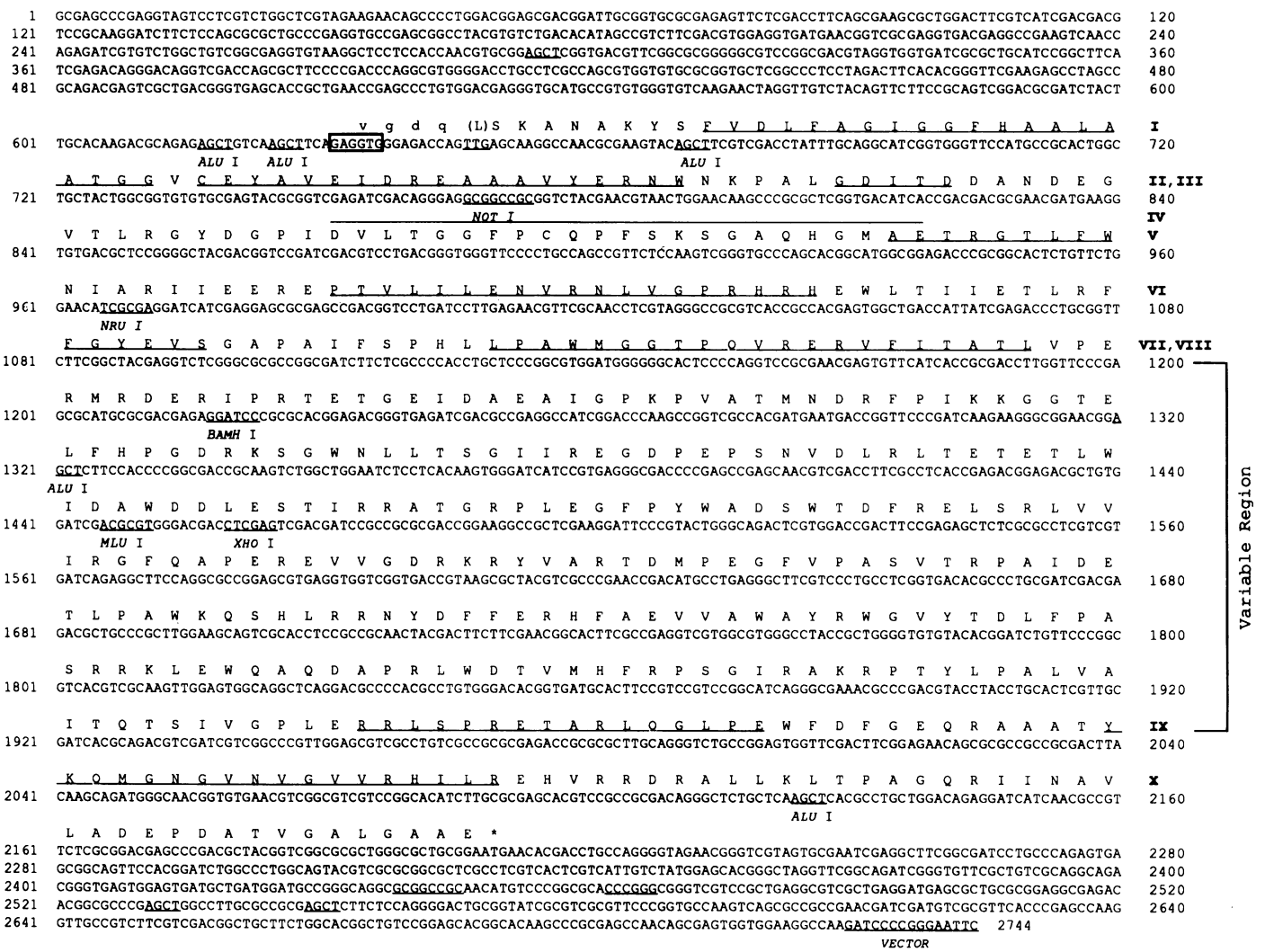

Figure 3. Sequence of the *aluIM* region. This sequence has been deposited to GenBank (accession number Z11841). The sites for selected restriction endonucleases have been indicated. The translation of *aluIM* is shown above the DNA sequence. Lower case is used at the start to indicate that either of two initiation codons (GTG or TTG) may be used to code for *N-formyl* methionine. The boxed bases in this region are complementary to the 3′ end of 16S rRNA from *Arthrobacter globiformis* (GenBank #M23411; the information for *A. luteus* is not available). The ten conserved motifs found in all known 5mC MTases (6–8) are indicated in underlined amino acids and roman numerals. Motifs IV and V overlap by two amino acids. The variable region, which in other MTases has been shown to play a role in DNA sequence recognition, lies between motifs VIII and IX as described in the text.

pooled and purified. This material was ligated into *Bam*HI-cut and CIAP-treated pUC8, and used to transform *E.coli* JM107MA2. The resulting plasmid pool was digested with excess *Alu*I as described above, and plasmid minipreparations from individual transformants were screened for resistance to R·*Alu*I. Of seven resistant clones, two were chosen for further analysis and were named p*AluM*2.7 and p*AluM*3.3 for their roughly 2.7 and 3.3kb insert sizes (Fig. 1). These two clones carry *aluIM* in opposite orientation to one another, and *aluIM* expression is probably not dependent on promoters in the vector as both plasmids are fully resistant to digestion by R·*Alu*I. However, when these two plasmids were used for *in vitro* expression with an *E.coli* S30 extract no extra bands corresponding to the MTase were detected (not shown), suggesting that the expression of M·*Alu*I is relatively weak in *E.coli*.

None of the clones, including p*AluM*14.0, gave any evidence of R·*Alu* I activity. There are several possible explanations for this. First, in cloning the *BamH*I restriction-modification system Brooks *et al.* concurrently cloned a second MTase gene, unlinked to the *BamH*I genes, which had the same sequence specificity as the *BamH*I MTase (32). A second possibility, unlikely given the location of *aluIM* well within the p*AluM*14.0 insert (Fig. 1; see below), was that the 14.0 kb primary clone didn't contain the intact *aluIR* gene. However, recent work supports one or more of the following explanations, as the *aluIR* gene is present in the initial *BamH*I and *Pst*I clones, but is expressed poorly (G.G.W., unpublished observation). The *aluIR* expression sequences may not be recognized in *E.coli* , the codon usage may lead to poor translation, the expressed protein may be unstable, or *aluIR* expression may rely on a regulatory protein

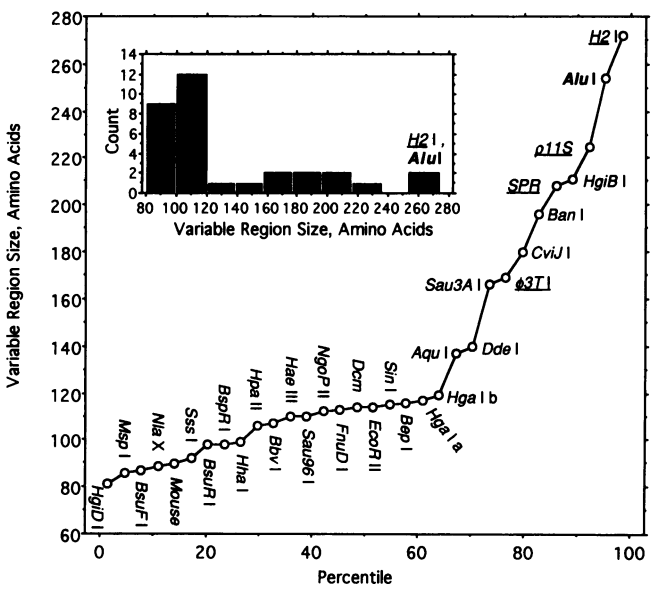

Figure 4. Distribution of variable region sizes among 32 5mC MTases. The variable region size, in amino acids, is plotted *versus* its percentile ranking for each MTase. The inset is a histogram showing the population distribution for the same MTases. The underlined MTases are phage-coded multispecific enzymes. Two MTases require additional comment. First, *Aqu* I is expressed as two distinct polypeptides that associate to form an active MTase; the break between them is within the variable region (50). *CviJ* I was reported to have a very small variable region (51), but the assignment of motifs used in this analysis is based on a substantially larger database (see ref 52).

such as has been found to be the case for several other REase genes (33–35), and this regulatory protein may not be expressed in *E.coli*.

Deletion mapping of gene boundaries and base sequence of *aluIM*

Deletions in p*AluM*2.7 were prepared and characterized as described (Materials and Methods), using plasmid resistance to digestion by R·*Alu*I as an indication of the *aluIM* status of that plasmid. These deletions are illustrated in Fig. 2 (top), and revealed that *aluIM* occupies at least 0.5kb (between the *Nru*I and *Xho*I sites), and at most 2.4kb (the entire insert up to the inner *Sma*I site).

The DNA sequences of both strands of the *aluIM* insert were determined from both of two independent *aluM*$^+$ (*Alu*I nuclease-resistant) plasmid DNAs, p*AluIM*2.7 and p*AluIM*3.3, in order to minimize the possibility of sequencing mutant or rearranged variants of the gene. The primers used are indicated in the lower portion of Fig. 2. The conserved amino acid sequence motifs found among 5mC MTases allowed us to discern the location and orientation of the *aluIM* gene from even preliminary sequence results. We expected an open reading frame (ORF) of about 500–520 codons, based on the published size of the M·*Alu* I protein (56,000 ± 1,000 MW; 36). A single ORF of the expected size was indeed found, and its boundaries agreed with the deletion mapping results (Fig. 3). This ORF could be 521 or 525 codons in length using, respectively, GTG or TTG as the initiation codon. The TTG seems to have a better Shine-Dalgarno element based on both the *E.coli* consensus (37) and on the 16S rRNA sequence

Table 1. Susceptibility of the *Alu* I MTase plasmid p*AluM*2.7 to digestion by various REases[a,b]

Specificity[c]	REase	Result	Cleavage sites[d]
AGCT	***Alu* I**	**R**	**23**
ACGT	*Mae* II	S	22
CATG	*Nla* III	S	16
CCGG	*Hpa* II	S	32
	Msp I	S	32
CCNGG	*ScrF* I	S	28
CGCG	*BstU* I	S	55
CTAG	*Bfa* I	S	8
CTNAG	*Dde* I	S	10
GATC	*Sau3A* I	S	40
GANTC	*Hinf* I	S	12
GCGC	*Hha* I	S	54
GCNGC	*Fnu4H* I	S	46
GGCC	*Hae* III	S	25
GTAC	*Rsa* I	S	9
GTNAC	*Mae* III	S	21
TCGA	*Taq* I	S	29
TTAA	*Mse* I	S	13

a—All digests used 1 μg of plasmid DNA, incubated with 2u of the indicated REase for 2h under the conditions recommended by the enzyme supplier.
b—An 'S' indicates cleavage (susceptibility) of the DNA, while 'R' indicates that the DNA was resistant to that REase.
c—Underlined sequences correspond to specificities of the known multispecific MTases *SPR*, *H2I*, *ρ11s*, and *φ3T*.
d—The number of potential cleavage sites for the REase in p*AluM*2.7.

from *Arthrobacter globiformis* (Fig. 3) (38; the corresponding sequence from *A. luteus* is not yet available). While they are rare, a TTG initiation codon has been found in another *Arthrobacter* gene and was recognized in *E.coli* (39).

Two substrate sequences for M·*Alu* I occur 4bp apart just upstream of the *aluIM* ORF (Fig. 3). The location of substrate sequences in potential regulatory regions has also been seen in other restriction-modification systems (40–43), though there is no evidence that they actually play a regulatory role (reviewed in ref. 44). Another closely-spaced pair of *Alu*I sites precedes an ORF that begins with CTG, downstream of *aluIM* (bottom of Fig. 3; the *Alu*I sites are underlined).

Comparison to other 5mC MTases

The deduced amino acid sequence for M·*Alu*I predicts a protein with molecular weight 59.0 k and a denatured isoelectric point of 6.0. The isoelectric point is within the range of known 5mC MTases, but the size is unusually large. As indicated in Fig. 3, all ten conserved motifs common to 5mC MTases are present in *aluIM*. Their spacing is unremarkable with the exception of an unusually large gap between conserved motifs VIII and IX, the so-called 'variable region.'

This unusually large gap is perhaps the most interesting result to come from the *aluIM* sequence itself. In several 5mC MTases, the variable region has been shown to include the target-recognizing domain (TRD), responsible for recognition of the target DNA sequence (7, 9–12). The multispecific *Bacillus* phage MTases each contain up to four TRDs of about 40 amino acids each, and mutations that affect different specificities define the respective TRDs (10, 45). Furthermore, recombinants have been made between these phage MTases, and the resulting chimeras have novel combinations of specificities based on the particular TRDs they have acquired (46, 47). Comparable

evidence now exists that the TRD resides within the variable region of monospecific 5mC MTases as well (11, 12).

The multispecific MTases have been noted to have larger variable regions than monospecific MTases, which is not surprising because the multispecific MTases contain multiple TRDs (reviewed in 44). We compared 32 known 5mC MTase sequences for the distance between motifs VIII and IX, and *aluIM* has the second-largest (Fig. 4). In fact, *aluIM* has a larger variable region than three of the multispecific phage MTases (*ϕ3T*, *ϱllS*, and *SPR*). Other apparently monospecific MTases, such as *HgiB*I and *Ban*I, also have unexpectedly large variable regions. The distribution shown in Fig. 4 suggests that the MTases fall into two populations based on the size of the variable region, on either side of a 120-amino acid boundary. Two-thirds (21/32) of the characterized 5mC MTases fall into the smaller group. All of the multispecific MTases and eight of the monospecific enzymes fall into the larger group, and the increments between their variable regions are larger.

One possible explanation for the large variable regions in the monospecific MTases such as *Alu*I is that they are really not monospecific. It is noteworthy that two of the known multispecific MTases, *H2I* and *r11s*, modify sequences that include the *Alu*I substrate AGCT. To test the possibility that *Alu*I MTase is multispecific, we digested plasmid p*AluM*2.7 with each of 17 different REases having 4–5-base specificities (Tab. 1). Among these REases are several that would have detected any of the specificities of the known multispecific MTases *SPR*, *H2I*, *ϱ11s*, and *ϕ3T*. For example, all of the known multispecific MTases protect DNA from the R·*Hae*III; M·*Alu*I did not. M·*Alu*I appears to be monospecific, as self-modification of p*AluM*2.7 did not protect the DNA from any REase tested aside from R·*Alu*I itself. Of course, we cannot yet rule out *Alu*I methylations that do not affect the 17 REases we tested, and we are attempting to detect other modifications by methylating R·*Alu* I-digested DNA with [^{3}H-CH$_3$] AdoMet.

The *Alu* I MTase retains activity when a third of the variable region is deleted

One of the properties of the multispecific phage MTases, as described above, is that deletions within the variable region can leave some of the activities intact. Individual TRDs are approximately 40 amino acids in length (9, 48). The variable region of M·*Alu* I is large enough to accommodate up to six such TRDs. We decided to test whether a substantial deletion within the variable region of *aluIM* could be made without eliminating MTase activity. To do this, we took advantage of two restriction sites within the variable region which are unique in the plasmid p*AluM*2.7 (see Figs. 1, 3): *Bam*HI and *Xho*I. Cleaving p*AluM*2.7 with both enzymes, removing the 5′ extensions with mung bean nuclease, and religating the blunted ends, would yield an in-frame deletion of 82 codons. The resulting construct was confirmed as having lost the *Bam*HI and *Xho*I sites, along with the *Mlu* I site between them. We sequenced across the deletion junction in p*AluM*2.7ΔBX from both sides. The result confirmed that the deletion was precise, generating the expected change from GAG′AGG′ATC′CCG···GAC′CTC′GAG′TCG to GAG′AG|G′TCG. Despite the removal of the amino-proximal third (82/254 codons) of the variable region, the resulting plasmid p*AluM*2.7ΔBX was still fully resistant to digestion by R.*Alu* I.

What is the significance of such a large variable region in a monospecific MTase? One possible explanation may come from the multispecific *ϱ11S* MTase. Two of its TRDs are nonfunctional (9), and differ from functional TRDs by only a few amino acids (48). Perhaps *aluIM* is simply an evolutionary intermediate: a multispecific MTase gene that has become associated with a REase gene, lost activity for all but one of its TRDs in the absence of positive selection, and which may yet lose its nonfunctional TRDs by deletion. In this case the only difference between the phage MTases and the larger monospecific MTases such as M·*Alu* I may be that the former group has been under selective pressure to retain several active TRDs, so as to provide the broadest possible defense against host restriction-modification systems; while the latter group has been under selective pressure to have only one active TRD, perhaps to maximize protection of the host against the single associated restriction enzyme or to minimize utilization of AdoMet.

Alternatively, once-functional but now-inactive TRDs may provide the ability to acquire new DNA recognition specificities without losing existing specificities, much as gene duplication allows evolutionary experimentation with one of the gene copies (49). According to this view, MTases such as M·*Alu* I would play a greater role in the emergence of new restriction-modification systems than MTases with smaller variable regions. It will be interesting to see if the variable region of *aluIM* contains reactivatable TRDs for additional recognition specificities.

ACKNOWLEDGEMENTS

We thank Ms. Sharon Master (Medical College of Ohio) for help in compiling the data for Fig. 4, and Drs. Ashok S. Bhagwat (Wayne State University) and Joan E. Brooks (New England BioLabs, Inc.), for their suggestions and comments on the manuscript. This report is based on work supported by the National Science Foundation under grants DMB-8818673 and MCB-9205248 to R.M.B.; and by New England BioLabs, Inc., research funds to G.G.W.

REFERENCES

1. Ingrosso, D., Fowler, A.V., Bleibaum, J. and Clarke, S. (1989) *J. Biol. Chem.* **264**, 20131–20139.
2. Chandrasegaran, S., and Smith, H.O. (1988) in Sarma, R.H., and Sarma, M.H. (eds.), *Structure and Expression Vol. 1: from Proteins to Ribosomes.* Adenine Press, Guilderland, NY, pp. 149–156.
3. Klimasauskas, S., Timinskas, A., Menkevicius, S., Butkiene, D., Butkus, V., and Janulaitis, A. (1989) *Nucl. Acids Res.* **17**, 9823–9832.
4. Lauster, R. (1989) *J. Mol. Biol.* **206**, 313–321.
5. Wu, J.C., and Santi, D.V. (1987) *J. Biol. Chem.* **262**, 4778–4786.
6. Posfai, J., Bhagwat, A.S., Posfai, G., and Roberts, R.J. (1989) *Nucl. Acids Res.* **17**, 2421–2435.
7. Lauster, R., Trautner, T.A., and Noyer-Weidner, M. (1989) *J. Mol. Biol.* **206**, 305–312.
8. Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988) *J. Molec. Biol.* **202**, 971–983.
9. Gunthert, U., Lauster, R., and Reiners, L. (1986) *Eur. J. Biochem.* **159**, 485–492.
10. Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B., and Trautner, T.A. (1988) *EMBO J.* 7, 2601–2609.
11. Klimasauskas, S., Nelson, J.L., and Roberts, R.J. (1991) *Nucl. Acids Res.* **19**, 6183–6190.
12. Mi, S., and Roberts, R.J. (1992) *Nucl. Acids Res.* 20, 4811–4816.
13. Roberts, R.J., Myers, P.A., Morrison, A., and Murray, K. (1976) J. Mol. Biol. **102**, 157–165.
14. Smith, H.O., and Kelley, S.V. (1984) in *DNA Methylation: Biochemistry and Biological Significance*, Cedar, H., Riggs, A.D., and Razin, A., ed., pp. 39–71, Springer Verlag, New York.
15. Bolivar, F., and Backman, K. (1979) *Methods Enzymol.* **68**, 245–267.

16. Raleigh, E.A., Murray, N.E., Revel, H., Blumenthal, R.M., Westaway, D., Reith, A.D., Rigby, P.W.J., Elhai, J. and Hanahan, D. (1988) *Nucleic Acids Res.* **16**, 1563–1575.
17. Blumenthal, R.M., Gregory, S.A., and Cooperider, J. S. (1985) *J. Bacteriol.* **164**, 501–509.
18. Raleigh, E.A., and Wilson, G. (1986) *Proc. Natl. Acad. Sci. USA* 83, 9070–9074.
19. Kiss, A., Posfai, G., Keller, C.C., Venetianer, P., and Roberts, R.J. (1985) *Nucleic Acids Res.* 13, 6403–6421.
20. Noyer-Weidner, M., Diaz, R., and Reiners, L. (1986) *Mol. Gen. Genet.* 205, 469–475.
21. Raleigh, E.A. (1992) *Molec. Microbiol.* **6**, 1079–1086.
22. Yannisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103–119.
23. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in Molecular Cloning: a Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory, N.Y.
24. Hattori, M., and Sakaki, Y. (1986) *Anal. Biochem.* **152**, 232–238.
25. Bartlett, J.A., Gaillard, R.K., Jr., and Joklik, W.K. (1986) *BioTechniques* **4**, 208–210.
26. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
27. Sanchez-Pescador, R., and Urdea, M.S. (1984) *DNA* **3**, 339–343.
28. Pearson, W. R. (1990) *Methods Enzymol.* **183**, 63–98.
29. Altschul, S.F., Miller, W., Myers, E.W., and Lipman, D.J. (1990) *J. Mol. Biol.* **215**, 403–410.
30. Higgins,D.G. Bleasby,A.J. and Fuchs,R. (1991) *CABIOS* .
31. Szomolanyi, E., Kiss, A., and Venetianer, P. (1980) *Gene* **10**, 219–225.
32. Brooks, J.E., Nathan, P.D., Landry, D., Sznyter, L.A., Waite-Rees, P., Ives, C.L., Moran, L.S., Slatko, B.E., and Benner, J.S. (1991) *Nucleic Acids Res.* **19**, 841–850.
33. Tao, T., Bourne, J.C., and Blumenthal, R.M. (1991) *J. Bacteriol.* **173**, 1367–1375.
34. Brooks, J.E., Nathan, P.D., Landry, D., Sznyter, L.A., Waite-Rees, P., Ives, C.L., Moran, L.S., Slatko, B.E., and Benner, J.S. (1991) *Nucleic Acids Res.* 19, 841–850.
35. Tao, T., and Blumenthal, R.M. (1992) *J. Bacteriol.* **174**, 3395–3398.
36. Yoon, H., Suh, H., Han, M.H., and Yoo, O.J. (1985) *Korean Biochem. J.* **18**, 82–87.
37. Gold, L., and Stormo, G. (1987) in *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., Schaechter, M., and Umbarger, H.E., ed., pp. 1302–1307, American Society for Microbiology, Washington, D.C.
38. Woese, C. (1990) GenBank accession #M23411.
39. Mauch, L., Bichler, V., and Brandsch, R. (1990) *Mol. Gen. Genet.* **221**, 427–434.
40. Theriault, G., Roy, P.H., Howard, K.A., Benner, J. S., Brooks, J.E., Waters, A.F., and Gingeras, T.R. (1985) *Nucleic Acids Res.* **13**, 8441–8461.
41. Looney, M.C., Moran, L.S., Jack, W.E., Feehery, G.R., Benner, J.S., Slatko, B.E., and Wilson, G.G. (1989) *Gene* **80**, 193–208.
42. Bocklage, H., Heeger, K., and Muller-Hill, B. (1991) *Nucleic Acids Res.* **19**, 1007–1013.
43. Slatko, B.E., Benner, J.S., Jager-Quinton, T., Moran, L.S., Simcox, T.G., et al. (1987) *Nucleic Acids Res.* 15, 9781–9796.
44. Wilson, G.G., and Murray, N.E. (1991) *Annu. Rev. Genet.* **25**, 585–627.
45. Behrens, B., Noyer-Weidner, M., Pawlek, B., Lauster, R., Balganesh, T.S., and Trautner, T.A. (1987) *EMBO J.* 6, 1137–1142.
46. Balganesh, T.S., Reiners, L., Lauster, R., Noyer-Weidner, M., Wilke, K., and Trautner, T.A. (1987) *EMBO J.* **6**, 3543–3549.
47. Trautner, T.A., Balganesh, T.S., and Pawlek, B. (1988) *Nucleic Acids Res.* **16**, 6649–6658.
48. Lange, C., Jugel, A., Walter, J., Noyer-Weidner, M., and Trautner, T.A. (1991) *Nature* **352**, 645–648.
49. Ohno, S. (1970) *Evolution by Gene Duplication*, Springer-Verlag, New York.
50. Karreman, C., and de Waard, A. (1990) *J. Bacteriol.* **172**, 266–272.
51. Shields, S.L., Burbank, D.E., Grabherr, R., and Van Etten, J.L. (1990) *Virol.* **176**, 16–24.
52. Wilson, G.G. (1992) *Methods Enzymol.* **216**, 259–279.