 © 1990 Oxford University Press

Cleavage at the twelve-base-pair sequence 5′-TCTAGATCTAGA-3′ using M·*Xba*I (TCTAG^{m6}A) methylation and *Dpn*I (G^{m6}A/TC) cleavage

Yogesh Patel, Elizabeth Van Cott[1], Geoffrey G.Wilson[1] and Michael McClelland
California Institute of Biological Research, 11099 North Torrey Pines Road, La Jolla, CA 92037 and [1]New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA

Received September 26, 1989; Revised and Accepted December 19, 1989

ABSTRACT

The DNA methylase M·*Xba*I was isolated from an *E. coli* recombinant clone. We deduce that the enzyme methylates at the sequence 5′-TCTAG^{m6}A-3′. In combination with the methylation-dependent restriction endonuclease, *Dpn*I (5′-G^{m6}A/TC-3′), DNA cleavage occurs at the sequence 5′-TCTAGA/TCTAGA-3′. This twelve-base-pair site should occur once every 16,000,000 base pairs in a random sequence of DNA. The exceptional rarity of the M·*Xba*I/*Dpn*I sequence makes it an ideal candidate for transpositional integration of a unique cleavage site into bacterial genomes. Retrotransposition into mammalian genomes is also an attractive possibility.

INTRODUCTION

It is difficult to generate chromosome fragments in the 'megabase' range because there are few enzymatic cleavage methods that generate DNA fragments of this size. To date, only a handful of endonucleases are known that will generate fragments averaging up to 1,000,000 base pairs, depending on the source of the DNA to be cleaved. These endonucleases include the restriction enzymes with the longest known recognition sequences *Not*I (5′-GCGGCCGC-3′) and *Sfi*I (5′-GGCCN$_5$GGCC-3′) (1).

To expand the repertoire of highly specific DNA cutting tools, we have developed a cleavage method (2, 3) in which sequence-specific adenine methylases are used to generate rare recognition sites (8 to 14 base pairs) for the methylation-dependent restriction endonuclease *Dpn*I (5′-G^{m6}A/TC-3′) (4, 5). Recently, we demonstrated selective cleavage of a bacterial genome at a ten-base-pair site by M·*Cla*I/*Dpn*I (3). We report here another methylase/*Dpn*I combination that utilizes cloned M·*Xba*I (5′-TCTAG^{m6}A-3′) (6) and cloned *Dpn*I (7) to cleave the twelve-base-pair sequence, 5′-TCTAGATCTAGA-3′.

METHODS

Enzymes

*Xba*I, *Hind*III, *Sau*3AI, *Mbo*I and M·*Hpa*II were purchased from New England Biolabs, Beverly, MA. M·*Xba*I was prepared from an *E. coli* clone (6), as described in the results.

*Dpn*I was prepared at 4°C by a modification of the previous purification procedure (7), from an *E. coli* clone, kindly provided by S. Lacks, Brookhaven Natl. Lab (7). 40 grams of cells grown in 9 liters of LB were suspended in 200 ml of 50 mM Tris-HCl (pH 7.5), 100 mM KaCl, 10 mM 2-mercaptoethanol, 1 mM Phenylmethylsulfonyl fluoride (Extract buffer). The cells in suspension were broken in a French press and centrifuged at 10,000 g for 30 minutes. Streptomycin sulfate was added to the supernatant to make a final concentration of 1% (weight/volume). This solution was spun at 10,000 g for 30 minutes and the resulting supernatant was dialysed extensively against Extract buffer (pH 7.0). The dialysate was loaded onto a 3×30 cm phosphocellulose P-11 column (Whatman) and eluted by 500 ml gradient of 0.1 M to 1.0 M KCl. 2 µl of column fractions were assayed with 1 µl of bacteriophage lambda *dam*$^+$ in 20 µl of KGB (8, 9) for one hour at 37°C. *Dpn*I eluted at about 400 mM KCl. The peak fractions, diluted two-fold with Extract buffer (pH 7.5), were loaded onto a 1×6 cm Heparin Sepharose column (Pharmacia/LKB) and eluted with a 300 ml gradient of 50 mM to 500 mM KCl. *Dpn*I eluted at about 20 mM KCl. The peak fractions were pooled and loaded directly onto a 1×6 cm hydroxyapatite column. Elution with a gradient of 10 mM to 500 mM potassium phosphate gave purified *Dpn*I at 200 mM phosphate.

Strains

E. coli DS1310 *dam*$^-$Sm^R, was kindly supplied by D.Smith.

DNAs

Bacteriophage lambda DNA prepared from a *dam*$^-$*E. coli* and the phosphorylated oligonucleotide 5′-pGGTCTAGACC-3′ were purchased from New England Biolabs. Bacteriophage T7 and Adenovirus-2 DNA were from Sigma Chemicals, St. Louis, MO.

DNA methylation and cleavage

Methylation of Adenovirus-2 DNA was performed in magnesium-free 1X KGB (100 mM Potassium glutamate, 50 mM Tris-acetate (pH 7.6), 1 mM 2-mercaptoethanol, 50 mg/ml BSA) (8, 9) containing 50 µM S-Adenosyl-methionine. Five units of M·*Xba*I were added and incubated at 30°C for one hour. 5 more units of M·*Xba*I and 50 µM S-Adenosyl-methionine were added and

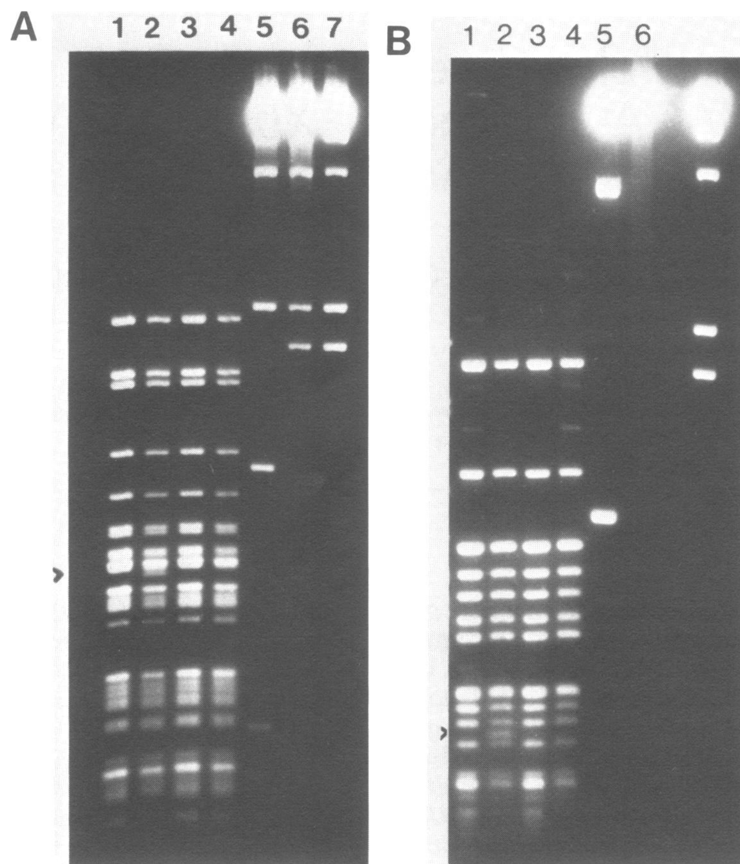


Figure 1. Methyl-specificity of M·*Xba*I tested by cross-protection versus *Mbo*I and *Sau*3AI.
A. 1.4% Agarose TBE gel electrophoresis. Lanes 1−7 contain 2 μg of bacteriophage lambda DNA (*dam*$^-$) that contains the sequence TCTAGATC at position 24,508. Lane 1. Bacteriophage lambda DNA, digested with *Mbo*I. Lane 2. Methylated with M·*Xba*I then digested with *Mbo*I. Lane 3. Digested with *Sau*3AI. Lane 4. Methylated with M·*Xba*I then digested with *Sau*3AI. Lane 5. Digested with *Hin*dIII+*Xba*I. lane 6. Methylated with M·*Xba*I then digested with *Hin*dIII+*Xba*I. Lane 7. Digested with *Hin*dIII.
B. 1.4% Agarose TBE gel electrophoresis. Lanes 1−6 contain 2 μg of Adenovirus 2 DNA that contains the sequence TCTAGATCTAGA at position 30,459. Lane 1. Adenovirus 2 DNA, digested with *Mbo*I. Lane 2. Methylated with M·*Xba*I then digested with *Mbo*I. Lane 3. Digested with *Sau*3AI. Lane 4. Methylated with M·*Xba*I then digested with *Sau*3AI. Lane 5. Digested with *Xba*I. Lane 6. Methylated with M·*Xba*I then digested with *Xba*I. Lane 7. Bacteriophage lambda DNA, digested with *Hin*dIII.

the mixture incubated for one more hour. The half-life of the enzyme was about 30 minutes under these conditions. M·*Xba*I was then inactivated by heating at 65°C for 30 minutes and magnesium acetate was added to make the buffer 10 mM Mg^{2+}. Incubation was continued with 10 units of *Dpn*I for 5 hours at 30°C.

Gel electrophoresis

Electrophoresis was performed in a 1% agarose gel with TBE buffer (90 mM Tris-borate (pH 8.0) and 2 mM EDTA).

RESULTS AND DISCUSSION

M·*Xba*I was prepared at 4°C from an *E. coli* clone pXbaI-M (6; Elizabeth Van Cott and Geoff Wilson, unpublished). Initially, the plasmid was transferred to a *dam*$^-$ derivative of *E. coli* LE392 to ensure that the M·*Xba*I methylase would not be contaiminated by M·*Dam* ($G^{m6}ATC$). 30 g of cells from 9 liters of LB were suspended in 200 ml of 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM 2-mercaptoethanol, 1 mM Phenylmethylsulfonyl fluoride (Extract buffer at 4°C). The suspension was broken in a French press and centrifuged at 10,000 g for 30 minutes. The supernatant was made 1% with streptomycin sulfate, spun at 10,000 g for 30 minutes and the supernatant dilaysed extensively against Extract buffer (pH 7.0). The dialysate was loaded on a phosphocellulose 3×30 cm P-11 column (Whatman) and eluted with a gradient of 0.1 M to 1.0 M KCl. The column was assayed with a polymerized *Xba*I linker $(5'\text{-pGGTCTAGACC-3'})_n$ and [^{3}H-methyl]-S-adenosyl-methionine using M·*Hpa*II ($C^{m}CGG$) as a control (10). M·*Xba*I eluted at about 450 mM KCl. The peak fractions were diluted two fold with Extract buffer (pH 7.5) and loaded on to a 1×6 cm Heparin sepharose (Pharmacia/LKB) column and eluted with a gradient of 50 mM to 500 mM KCl. M·*Xba*I eluted at about 200 mM KCl. The peak fractions were pooled and loaded directly on a 1×6 cm Hydroxyappatite column and eluted with a gradient of 10 mM to 500 mM potassium phosphate. M·*Xba*I eluted at about 200 mM phosphate. The peak was dialyzed against 50% glycerol, 50 mM KCl, 20 mM Tris-Cl, pH 7.5 and stored at −20°C.

One unit was the amount of M·*Xba*I that, in one hour at 30°C, protected 1 μl of Adenovirus−2 DNA in 30 μl of 1×KGB from *Xba*I cleavage. Approximately 100,000 units were recovered.

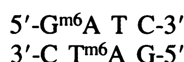
Specificity of M·*Xba*I and cleavage at 5′-TCTAGATCTAGA-3′ by M·*Xba*I + *Dpn*I

The specificity of M·*Xba*I was determined to be 5′-TC-TAG^{m6}A-3′ by a number of criteria. First, overlapping sites for M·*Xba*I with *Mbo*I and *Sau*3AI (5′-GATC-3′) at

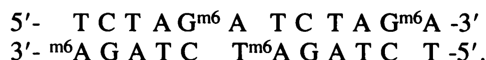
5′-TCTAGATC-3′
3′-AGATCTAG-5′

were inspected. *Mbo*I cuts 5′-GAT^{m5}C-3′ or 5′-GAT^{m4}C-3′, but not 5′-G^{m6}ATC-3′; whereas *Sau*3AI cuts 5′-G^{m6}ATC-3′, but not 5′-GAT^{m5}C-3′ or 5′-GAT^{m4}C-3′ (11). Figures 1A and B indicate that methylation by M·*Xba*I blocked *Mbo*I cleavage (lane 2), but did not block *Sau*3AI cleavagσ (lanσ 4ã wherσ thσ 5′-TCTAGA-3′ and 5′GATC-3′ recognition sequences overlap at position 24,508 in bacteriophage lambda and at position 30,459 in Adenovirus-2 DNA. These lanes demonstrate that M·*Xba*I is not a ^{m5}C or ^{m4}C specific enzyme. Blocking of *Mbo*I but not *Sau*3AI cleavage indicates ^{m6}A specificity for M·*Xba*I, or a specificity hitherto unknown.

Second, *Dpn*I requires adenine methylation in both strands at


5′-G^{m6}A T C-3′
3′-C T^{m6}A G-5′

for double-stranded cleavage to occur (4, 5). Methylation of a tandem repeat of the M·*Xba*I site should introduce ^{m6}A at


5′- T C T A G^{m6}A T C T A G^{m6}A -3′
3′- ^{m6}A G A T C T^{m6}A G A T C T -5′.

This tandem *Xba*I site creates a *Dpn*I site in the center of the twelve-base-pair sequence. DNA that has ^{m6}A only at the sequence 5′-TCTAG^{m6}A-3′ should be cleaved by *Dpn*I only at this twelve-base-pair site. A search of the GenBank database of 33 million base pairs revealed two such sequences. One site occurs in the Adenovirus-2 genome at position 30459. We have used this site to test the M·*Xba*I/*Dpn*I cleavage since Ad-2 DNA is otherwise free of ^{m6}A. Figure 2 lanes 6, 7 and 8 indicate that the M·*Xba*I methylase protected against *Xba*I cleavage and also *enabled Dpn*I to cleave at a site that maps to 30459 (Figure 2, lanes 3, 4 and 5).

Twenty M·*Xba*I protection units did not allow complete *Dpn*I cleavage at position 30459 (Figure 2, lane 3). The reason for this incomplete digestion may be that the *Dpn*I digest is a much more sensitive assay for complete methylation than *Xba*I digestion. Since hemi-methylation is sufficient to block *Xba*I, a 90% methylation would block 99% of *Xba*I sites $(1-(0.1)^2)$, but would allow cleavage by *Dpn*I at only 81% of sites $(0.9)^2$. Futhermore, there was a large amount of a non-specific DNA binding protein in our methylase preparation which may have inhibited the methylation when large amounts of enzyme were used.

The specificity of the M·*Xba*I/*Dpn*I cleavage was checked against substrates that do not contain the sequence 5′-TCTAGATCTAGA-3′. M·*Xba*I/*Dpn*I treatment of bacteriophage T7 39,936 bp) and lambda (48,502 bp) DNAs did not result in any detectable cleavage, even at the partial matches 5′-TGCAGATCTGGA-3′ and 5′-TCTAGATCCCTC-3′ (data not shown). Furthermore, the sequences 5′-TCTCGATCCA-GC-3′ and 5′-TGAAGATCTGGA-3′ in adenovirus-2 were not cleaved.

Effect of prior ^{m5}C methylation on M·*Xba*I/*Dpn*I

Many genomes, including the human genome, are modified at ^{m5}C, primarily at 5′-^{m5}CG-3′, 5′-^{m5}CC-3′ (11, 12) and ^{m5}CWG

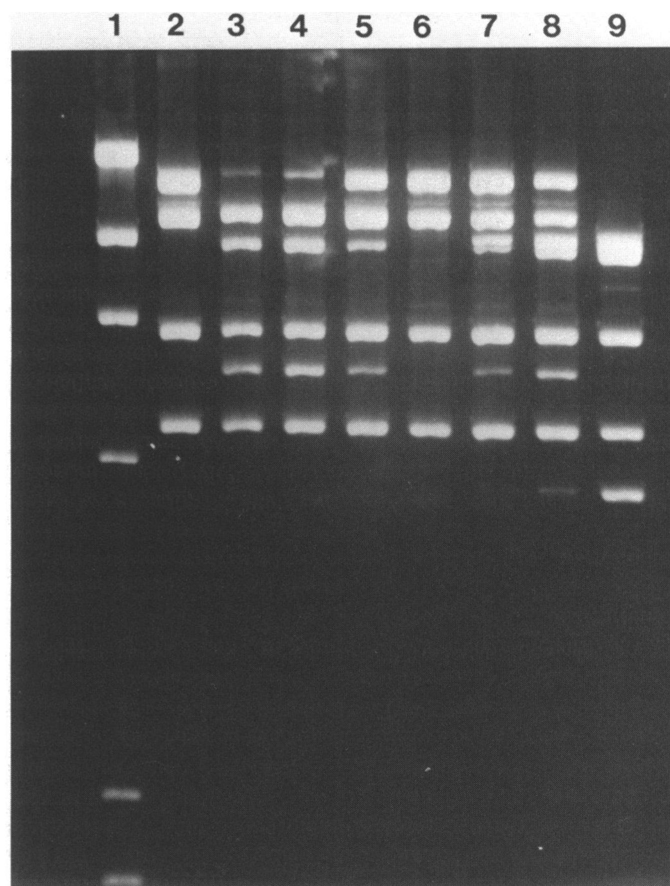


Figure 2. Cleavage at TCTAGATCTAGA. 1% Agarose TBE gel electrophoresis. Each lane contains 2 μg of Adenovirus 2 DNA that contains the sequence TCTAGATCTAGA at position 30,459. Lane 1. Lambda DNA marker, digested with *Hin*dIII. Lane 2. Ad-2 DNA digested with *Dpn*I+*Bam*HI. Lane 3. Methylated with 20 units of M·*Xba*I then digested with *Dpn*I+*Bam*HI. Lane 4. Methylated with 10 units of M·*Xba*I then digested with *Dpn*I+*Bam*HI. Lane 5. Methylated with 5 units of M·*Xba*I then digested with *Dpn*I+*Bam*HI. Lane 6. Methylated with 20 units of M·*Xba*I then digested with *Xba*I+*Bam*HI. Lane 7. Methylated with 10 units of M·*Xba*I then digested with *Xba*I+*Bam*HI. Lane 8. Methylated with 5 units of M·*Xba*I then digested with *Xba*I+*Bam*HI. Lane 9. Digested with *Xba*I+*Bam*HI.

(13). This cytosine modification blocks the action of most methylases and endonucleases at overlapping sites (14). For example, *Not*I does not cleave when its recognition sequence is methylated at 5′-GCGGC^{m5}CGC-3′ (14; Qiang *et al*, submitted). 5′-^{m5}CG-3′ modification therefore causes incomplete *Not*I cleavage of the human genome because such methylation is variable and incomplete in human DNA (15). Neither M·*Xba*I nor *Dpn*I recognition sequences overlap 5′-CG-3′ or 5′-CC-3′ in the TCTAGATCTAGA sequence. The M·*Xba*I/*Dpn*I cleavage strategy could therefore be useful for cleaving genomes that are methylated at 5′-^{m5}CG-3′.

Frequency of TCTAGATCTAGA

In a random sequence of DNA, in which all four bases are equally represented, the twelve-base-pair 5′-TCTAGATCTAGA sequence should occur, on average, once every 4^{12} base pairs or once every 16,000,000 base pairs. Bacterial genomes are typically less than 10^7 base pairs in size. 5′-TCTAGATCTAGA-3′ should be absent from most such genomes. Furthermore, the tetranucleotide sequence 5′-CTAG-3′

Table 1. Methylase/DpnI strategy

Enzyme	Recognition Specificity	Methylase/*Dpn*I Site	G^{m6}ATC Specific	Cloned
A: Suitable Methylases				
		8 base pairs		
M·CviBIII	TCG^{m6}A	TC*GATC*GA	yes	yes
M·*Taq*I	”	”	yes	yes
M·*Tth*I	”	”	yes	no
M·*Cla*I	ATCGAT	ATC*GATC*GAT	yes	no
M·*Mbo*II	GAAG^{m6}A	GAA*GATC*TTC	yes	no
		12 base pairs		
M·*Xba*I	TCTAG^{m6}A	TCTA*GATC*TAGA	yes	yes
		14 base pairs		
M·*Eco*B	TG^{m6}AN$_8$TGCT	AGCAN$_5$T*GATC*AN$_5$TGCT	yes	yes
B: Potentially suitable methylases				
		10 base pairs		
M·*Ban*III	ATCGAT	ATC*GATC*GAT	?	no
M·*Bsp*106I	ATCGAT	ATC*GATC*GAT	?	no
M·*Bsp*XI	”	”	?	no
M·*Bsc*I	”	”	?	no
M·*Ncu*I	GAAGA	GAA*GATC*TTC	probably	no
M·*Tce*I	GAAGA	GAA*GATC*TTC	?	no
M·*Mam*I	GATN$_4$TC	GATN$_3$GATCN$_3$ATC	?	no
		11 base pairs		
M·*Mme*I	GTYGGA	GTYG*GATC*CRAC	?	no
		12 base pairs		
M·*Eco*R124	GAAN$_6$RTCG	GAAN$_4$C*GATC*GN$_4$TTC	probably	yes
M·*Eco*R124/3	GAAN$_7$RTCG	GAAN$_5$C*GATC*GN$_5$TTC	probably	yes
M·*Taq*II	GACCGA	GACC*GATC*GGTC	?	no
M·*Rsp*XI	TCATGA	TCAT*GATC*ATGA	?	no
M·*Acc*III	TCCGGA	TCCG*GATC*CGGA	probably	no
M·*Nru*I	TCGCGA	TCGC*GATC*GCGA	?	no

Systems that have been reduced to practice in PFE: M·*CviBIII/Dpn*I, M·*Taq*I/*Dpn*I, M·*Cla*I/*Dpn*I (3) M·*Xba*I/*Dpn*I (Hanish and McClelland, in preparation). Systems that have been tested and proven not to be G^{m6}A specific: M·*Hph*I (T^{m5}CACC), M·*Ngo*VIII (T^{m5}CACC), M·*Bsp*HI (TCATA), M·*Bsp*MII (TCCGGA), M·*Mro*I (TCCGGA), M·*Kpn*21 (TCCGGA), M·*Ama*I (TCGCGA) M·*Sal*DI (TCGCGA), M·*Sbo*13 (TCGCGA) and M·*Eco*DXXI (TCAN$_7$^{m6}ATTC). M·*Eco*B, M·*Eco*R124, and M·EcoR124/3 are all type-I restriction methylases. See (3), (14) and (21) for references.

is rare in many eubacterial DNAs (18, 19, 20). For instance, the *Xba*I site 5′-TCTAGA-3′ occurs less than once every 100,000 base pairs in the *E. coli* genome. The M·*Xba*I/*Dpn*I site should occur every $(100{,}000)^2$ base pairs or once every 10^{10} base pairs!

The frequency of the M·*Xba*I/*Dpn*I site in the A+T rich human genome, calculated from dinucleotide frequencies (16, 17), should be once every 10,500,000 base pairs, on average.

Potential applications

The M·*Xba*I/*Dpn*I sequence can be introduced into a bacterial genome by transposition or by piggy-backing on a homologous recombination event. The 12-base-pair site will only occur where it is integrated. Preliminary experiments have been performed with transposons containing the M·*Xba*I/*Dpn*I site and a selectable marker. We are able to cleave bacterial genomes as few as one to six times by the introduction of one or more transposons into bacterial chromosomes (Hanish and McClelland, manuscript in preparation). Mapping *Not*I restriction sites relative to single integrations has been achieved by M·*Xba*I/*Dpn*I cleavage, partial digestion with *Not*I, PFE, Southern blotting and indirect end labelling with each end of the transposon.

M·*Xba*I/*Dpn*I transposon integrations will also help to produce physical/genetic maps of bacterial genomes. We propose to integrate two M·*Xba*I/*Dpn*I transposons into a 'type' strain. Subsequent mutants generated by a third integration of an M·*Xba*I/*Dpn*I transposon with a different selectable marker could be mapped relative to the original two integrations by M·*Xba*I/*Dpn*I cleavage and PFE. This 'clock' style of physical/genetic mapping may allow genetic mapping in species or strains where this is presently difficult or impossible. Furthermore, since the M·*Xba*I gene is cloned, it should be possible to construct a transposon in which the methylase gene is carried along with the M·*Xba*I/*Dpn*I site. Such a construct would bypass the need for *in vitro* methylation.

Methods described for bacteria may also work for mammalian genomes.

Pulsed field electrophoresis (22) and subsequent improvements (for example, 23, 24) have permitted the separation in agarose gels of DNA molecules up to 10,000,000 base pairs in length. A battery of highly specific eight- to twelve-base-pair cleavage methods should aid in the mapping of a number of genomes. The current status of methylase/*Dpn*I combinations is summarized in Table 1.

ACKNOWLEDGEMENTS

We thank Ira Schildkraut and Don Comb for their kind cooperation. We thank Mike Nelson for helpful discussions and critical review of the manuscript. M.M. was a Lucille P.Markey Biomedical Scholar. Yogesh Patel was a Srinivasa Ramanujan fellow. This work was funded in part by the U.S. Department of Energy, the National Science Foundation and the National

Institutes of Health, New England Biolabs and the Lucille P.Markey Charitable Trust.

REFERENCES

1. Qiang,B.-Q. and Schildkraut,I. (1987) *Methods in Enzymol.* **155**, 15–21.
2. McClelland,M., Kessler,L.G. and Bittner,M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 983–987.
3. Weil,M.D. and McClelland,M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 51–55.
4. Lacks,S. and Greenberg,B. (1975) *J. Biol. Chem.* **250**, 4060–4066.
5. Lacks,S. and Greenberg,B. (1977) *J. Mol. Biol.* **114**, 153–168.
6. Van Cott,E.M. and Wilson,G.G. (1988) *Gene* **74**, 55–59.
7. de la Campa,A.G., Springhorn,S.S., Purushottam,K. and Lacks,S.A. (1988) *J. Biol. Chem.* **263**, 14696–14702.
8. McClelland,M., Hanish,J., Nelson,M. and Patel,Y. (1988) *Nucl. Acids Res.* **16**, 364.
9. Hanish,J. and McClelland,M. (1988) *Gene Anal. Tech.* **5**, 105–107.
10. Nelson,M. and McClelland,M. (1987) *Methods in Enzymology* **155**, 32–41.
11. Bird,A.P. (1980) *Nucl. Acids Res.* **8**, 1499–1504.
12. Swartz,M.N., Trautner,T.A. and Kornberg,A. (1962) *J. Biol. Chem.* **237**, 1961.
13. Woodcock,D.M., Crowther,P.J., Jefferson,S. and Diver,W.P. (1989) *Gene* **74**, 151–152.
14. Nelson,M. and McClelland,M. (1989) *Nucl. Acids Res.* **17**, 389–415.
15. Waalwijk,C. and Flavell,R.A. (1978) *Nucl. Acids Res.* **5**, 4631–4641.
16. Setlow,P. (1976) In: *CRC Handbook of Biochemistry and Molecular Biology*, 3rd Edition. Fasman,G.D. (ed). CRC Press, Cleveland, Ohio, pp. 313–318.
17. McClelland,M. (1985) *J. Mol. Evol.* **21**, 317–322.
18. McClelland,M. and Nelson,M. (1987) *Gene Amplification and Analysis* **5**, 257–282.
19. Phillips,G.J., Arnold,J. and Ivarie,R. (1988) *Nucl. Acids Res.* **15**, 2627–2638.
20. McClelland,M., Jones,R., Patel,Y. and Nelson,M. (1987) *Nucl. Acids Res.* **15**, 5985–6005.
21. Wilson,G.G. (1988) *Gene* **74**, 281–289.
22. Schwartz,D.C. and Cantor,C. (1984) *Cell* **37**, 67–75.
23. Orbach,M.J., Vollrath,D., Davis,R.W. and Yanofsky,C. (1988) *Mol. Cell. Biol.* **8**, 1469–1473.
24. Turmel,C. and Lalande,M. (1988) *Nucl. Acids Res.* **16**, 4727.