Site-specific cleavage of DNA at 8- and 10-base-pair sequences

(restriction methylase/*Dpn* I/DNA methylation)

MICHAEL MCCLELLAND*, LOUIS G. KESSLER, AND MICHAEL BITTNER[†]

Department of Molecular and Population Genetics, University of Georgia, Athens, GA 30602

ABSTRACT   A method is described for cutting DNA at specific sites that are 8 and 10 base pairs long. The DNA is first treated with a specific methylase, either the restriction-modification enzyme *M. Taq* I, which converts the 4-base sequence T-C-G-A to T-C-G-^mA, or the similar enzyme *M. Cla* I, which converts the 6-base sequence A-T-C-G-A-T to A-T-C-G-^mA-T. The DNA is then cleaved with *Dpn* I, a restriction endonuclease that recognizes the sequence G-^mA-T-C. *Dpn* I is unique in that it cuts only DNA that is methylated at adenine in both strands of its recognition sequence. In DNAs that are not otherwise methylated at adenine in both strands of the sequence G-A-T-C, cleavage by *Dpn* I occurs only at the following sequences:

in the case of *M. Taq* I methylation,

$$5'\ \text{T-C-G-}^mA \text{ - T-C-G-}^mA\ 3'$$
$$3'\ ^mA\text{-G-C - T-}^mA\text{-G-C - T}\ 5';$$

in the case of *M. Cla* I methylation,

$$5'\ A \text{ - T-C-G-}^mA \text{ - T-C-G-}^mA\text{-T}\ 3'$$
$$3'\ T\text{-}^mA\text{-G-C - T-}^mA\text{-G-C - T-A}\ 5'.$$

Specific cutting and cloning at these methylase/*Dpn* I-generated sites is shown experimentally. Further, we describe how the above technique can be extended to generate *Dpn* I cleavage sites of up to 12 base pairs. In DNA that contains equal amounts of each base distributed at random, 8- and 10-base-pair recognition sequences occur, on the average, approximately once every 65,000 and 1,000,000 base pairs, respectively. Potential applications, including the development of cloning vectors and a rapid method for chromosome walking, are discussed.

Type I restriction endonucleases recognize double-stranded DNA sequences up to 7 base pairs long (1) but do not cleave site specifically. In contrast, type II and III restriction endonucleases have proved useful in molecular biology by virtue of their ability to recognize specific sequences of 4 to 6 bases in double-stranded DNA and cleave both strands at specific sites close to or in their recognition sequences (2). We describe here a method whereby methylation of DNA by the restriction-modification enzyme[‡] *M. Taq* I (6, 7) generates *Dpn* I restriction endonuclease cleavage sites (8, 9) at the 8-base-pair sequence T-C-G-A-T-C-G-A. Similarly, methylation of DNA by the modification enzyme *M. Cla* I (6) generates *Dpn* I cleavage sites at the 10-base-pair sequence A-T-C-G-A-T-C-G-A-T.

Potential applications for site-specific cleavage at 8- and 10-base-pair sequences are discussed. These include techniques that could facilitate the characterization of large genomes and the development of new cloning vehicles.

MATERIALS AND METHODS

Chemicals. 5-Bromo-4-chloro-3-indolyl β-D-galactoside and isopropyl thiogalactoside were purchased from Sigma.

[α-^{32}P]dCTP was purchased from New England Nuclear. The oligonucleotide C-C-A-T-C-G-A-T-C-G-A-T-G-G was synthesized chemically (10).

Enzymes. Purification of *M. Taq* I from *Thermus aquaticus* YT1 and *M. Cla* I from *Caryophanon latum* L and conditions for DNA methylation with *M. Taq* I and *M. Cla* I have been described (6). *Dpn* I was purchased from Bethesda Research Laboratories. Other restriction endonucleases and T4 ligase were purchased from New England BioLabs. All restriction endonuclease digestions and ligations were carried out using the vendors' recommended conditions. End-labeling conditions and autoradiography techniques have been described (11). *Escherichia coli* DNA polymerase I large fragment was purchased from New England Nuclear. DNA sequence analyses were carried out using the dideoxy technique (12).

Transformation and Selection of pMON2001 and pMON2002 Recombinants. Transformation (13) and plasmid purification procedures (14) have been described. Selection was carried out on Luria plates containing ampicillin at 20 μg/ml and chloramphenicol at 20 μg/ml. Screening was carried out on Luria plates containing tetracycline at 20 μg/ml.

Bacterial Strains and Plasmids. *E. coli* strains GM33 *dam-3* (15) and SK1592 were provided by S. Kushner. JM103 (16) was provided by S. Hollingshead. pMOB45 is a temperature-sensitive copy-number-defective plasmid containing the genes for chloramphenicol and tetracycline resistance (17).

RESULTS

Theory for Generating 8- and 10-Base-Pair *Dpn* I Cleavage Sites. *Dpn* I is a sequence-specific endonuclease, isolated from *Diplococcus pneumoniae* (8), that cuts double-stranded DNA in both strands at the sequence $\begin{matrix} 5'\ \text{G-}^mA\text{ - T-C}\ 3' \\ 3'\ \text{C}\quad \text{T-}^mA\text{-G}\ 5' \end{matrix}$ to produce flush ends (9). This enzyme is different from other known DNA endonucleases in that it requires methylation at adenine in *both* strands in order to cleave DNA (8, 9).

Theoretically, any sequence-specific methylase that recognizes a sequence overlapping the *Dpn* I recognition sequence and methylates at the correct adenine in the overlapping sequence could generate a *Dpn* I cleavage site. For instance, the *dam* product of *E. coli* methylates G-A-T-C at adenine in both strands (18), and thus DNA from *dam*$^+$ *E. coli* is cleaved by *Dpn* I (8, 9). In contrast, DNA from *dam*$^-$ mutants of *E. coli* is not cleaved by *Dpn* I (8, 9).

In general, restriction-modification enzymes recognize the same sequences as the corresponding restriction endonucle-
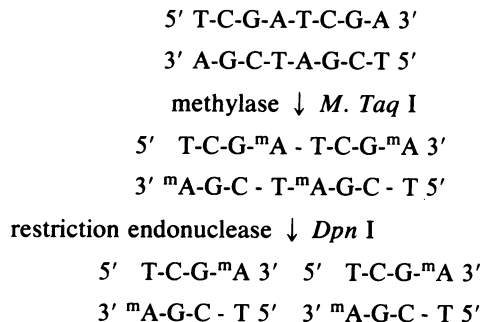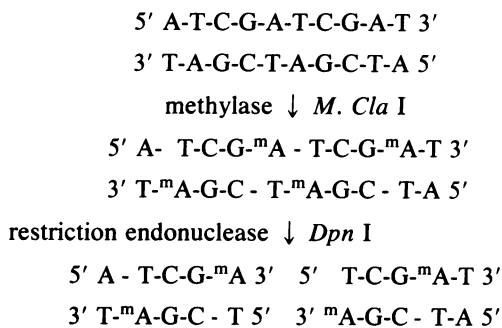
ases and protect these sequences from endonuclease cleavage by methylating DNA at adenine or cytosine (2, 4, 5). Thus, a prokaryote that contains a restriction endonuclease that recognizes a sequence overlapping G-A-T-C by 2 or more base pairs may also contain a corresponding modification enzyme that methylates at adenine in the overlapping sequence. Restriction methylase-recognition sequences (2) that overlap G-A-T-C by 2 or 3 base pairs are summarized in Table 1. Three of these modification methylases, *M. Taq* I, *M. Tth* I, and *M. Cla* I, have been isolated and their methylation specificity has been determined (6, 7).

Methylation by *M. Taq* I at a direct repeat of the *M. Taq* I recognition sequence (T-C-G-^{m}A) produces a *Dpn* I cleavage site:

5′ T-C-G-A-T-C-G-A 3′

3′ A-G-C-T-A-G-C-T 5′

methylase ↓ *M. Taq* I

5′ T-C-G-^{m}A - T-C-G-^{m}A 3′

3′ ^{m}A-G-C - T-^{m}A-G-C - T 5′

restriction endonuclease ↓ *Dpn* I

5′ T-C-G-^{m}A 3′   5′ T-C-G-^{m}A 3′

3′ ^{m}A-G-C - T 5′   3′ ^{m}A-G-C - T 5′

The 6-base-pair sequence 5′ T-C-G-A-T-C 3′, although containing the recognition sequences of both *M. Taq* I and *Dpn* I, would not be cleaved by *Dpn* I after *M. Taq* I methylation because the *Dpn* I recognition sequence would be methylated at adenine in only one strand, $\begin{array}{l}5'\ \text{T-C-G-}^m\text{A-T-C } 3'\\3'\ {}^m\text{A-G-C - T-A-G } 5'.\end{array}$

Methylation by *M. Cla* I at two partially overlapping *M. Cla* I recognition sequences produces a *Dpn* I cleavage site:

5′ A-T-C-G-A-T-C-G-A-T 3′

3′ T-A-G-C-T-A-G-C-T-A 5′

methylase ↓ *M. Cla* I

5′ A- T-C-G-^{m}A - T-C-G-^{m}A-T 3′

3′ T-^{m}A-G-C - T-^{m}A-G-C - T-A 5′

restriction endonuclease ↓ *Dpn* I

5′ A - T-C-G-^{m}A 3′   5′ T-C-G-^{m}A-T 3′

3′ T-^{m}A-G-C - T 5′   3′ ^{m}A-G-C - T-A 5′

Thus, in DNA that is otherwise unmethylated at G-A-T-C; for instance, *dam⁻ E. coli* or mammalian DNA, the 8- and 10-base-pair methylated sequences shown above would be the only *Dpn* I cleavage sites. The other potential 10- and the 12-base-pair cleavage systems (Table 1) are described in the *Discussion*.

Construction of Plasmids Containing the Sequence A-T-C-G-A-T-C-G-A-T. A derivative of pBR322 was constructed that contained the 10-base-pair sequence A-T-C-G-A-T-C-G-A-T predicted to be a substrate for *Dpn* I cleavage after *M. Cla* I or *M. Taq* I methylation.

The oligonucleotide A-T-C-G-A-T-C-G-A-T was inserted by blunt-end ligation at the *Sma* I site of bacteriophage M13 derivative mp8 (16) to produce phages mp2001 and mp2002, which contain one copy and two copies, respectively, of the oligonucleotide insert. These phage produce white plaques on JM103 in the presence of 5-bromo-4-chloro-3-indolyl β-D-galactoside and isopropyl thiogalactoside. The sequence of

the insert was confirmed by the dideoxy sequence analysis method (12).

Plasmids pMON2001 and pMON2002 were derived from pBR322 by insertion of the *Eco*RI–*Hin*dIII region of mp2001 and mp2002, respectively, in place of the *Eco*RI–*Hin*dIII region of pBR322. Both plasmids contain the ampicillin-resistance gene of pBR322. The tetracycline-resistance gene has been inactivated by the insertion.

Generation of *Dpn* I Cleavage Sites by *M. Taq* I and *M. Cla* I Methylation. Plasmids pMON2001 and pMON2002 were prepared from the *E. coli dam⁻* strain GM33 (14, 15). This strain lacks the G-^{m}A-T-C-specific *dam* methylase, which in wild-type *E. coli*, creates *Dpn* I cleavage sites. Plasmids were methylated with *M. Taq* I or *M. Cla* I (6). It was found that *M. Taq* I methylation at all sites, including sites on other DNAs such as pBR322, occurs about 3% as efficiently in the presence of pMON2001 or pMON2002 (data not shown). Destruction of the sequence A-T-C-G-A-T-C-G-A-T by *Cla* I digestion restored normal *M. Taq* I methylation efficiency at all other sites. The reason for this phenomenon is not understood, but it may be that the methylase has a higher affinity for direct repeats of its recognition sequence. Total methylation can be achieved by increasing the amount of methylase used, the incubation time, or both.

As predicted, *M. Taq* I- and *M. Cla* I-methylated pMON2001 and pMON2002 were linearized by *Dpn* I. Restriction mapping shows *Dpn* I to cut specifically at a site in the *Eco*RI–*Hin*dIII insert of both plasmids (Fig. 1). The plasmids contain 23 other occurrences of the sequence G-A-T-C, including partial overlaps with *M. Taq* I and *M. Cla* I at T-C-G-A-T-C, position 1127, and C-G-A-T-C-G, position 3735. None of these sequences is a substrate for *Dpn* I before or after methylation with *M. Taq* I or *M. Cla* I. This confirms previous observations that *Dpn* I requires methylation in both strands in order to cleave DNA (8, 9) and that under the reaction conditions used *M. Taq* I and *M. Cla* I do not produce detectable methylation at subsets of their recognition sequences (6).

Cloning at Methylase/*Dpn* I Sites. In order to demonstrate cloning at the 8-base-pair *Dpn* I cleavage site produced by *M. Taq* I, a gene for chloramphenicol resistance was inserted at the *M. Taq* I/*Dpn* I site in pMON2001; pMOB45 was cleaved with *Sma* I to generate three fragments, one of which is 3.6 kilobases long and contains the chloramphenicol-resistance gene. This mixture was blunt-end ligated to *M. Taq* I/*Dpn* I-linearized pMON2001 and transformed into SK1592 (13). Selection for ampicillin and chloramphenicol resistance and screening for tetracycline sensitivity were carried out. Ampicillin-resistant, chloramphenicol-resistant, tetracycline-sensitive colonies were picked and further screened by restriction mapping. Fragments were end labeled using large fragment DNA polymerase I prior to electrophoresis (Fig. 2). The 3.6-kilobase *Sma* I fragment of pMOB45 containing the chloramphenicol-resistance gene maps to the region between the *Eco*RI and *Sal* I sites in the amp^R, cm^R, tet^S recombinant plasmid pLK1, derived from pMON2001 after *M. Taq* I/*Dpn* I cleavage [Fig. 3 and Fig. 1 (lanes D and E)]. This corresponds to the 40-base-pair region containing the direct repeats of *Taq* I recognition sequences.

Specific cutting and ligation at the 10-base-pair *M. Cla* I/*Dpn* I site was also demonstrated. Deletions of the region from the *M. Cla* I/*Dpn* I site to the *Pvu* II site at position 2067 in plasmids pMON2001 and pMON2002 (Fig. 3) were constructed; *M. Cla* I-methylated pMON2001 and pMON2002 were each double digested with *Dpn* I/*Pvu* II and religated. The religated DNA was then redigested with *Cla* I/*Pvu* II/*Ava* I and then transformed into SK1592 (13). Cells were selected for ampicillin resistance and these were characterized by restriction digestion. Under such conditions, the majority of molecules expected to confer ampicillin resistance and

Table 1. Type II restriction-modification methylases that recognize sequences overlapping G-A-T-C by 2 or 3 base pairs

Methylase	Recognition sequence	*Dpn* I cleavage sequence	Length, base pairs	Comments (from refs. 4–6)
M. Taq I, *M. Tth* I	T-C-G-A	5' T-C-G-^{m}A - T-C-G-^{m}A 3' 3' ^{m}A-G-C - T-^{m}A-G-C - T 5'	8	Methylation at T-C-G-^{m}A
M. Cla I	A-T-C-G-A-T	5' A - T-C-G-^{m}A - T-C-G-^{m}A-T 3' 3' T-^{m}A-G-C - T-^{m}A-G-C - T-A 5'	10	Methylation at A-T-C-G-^{m}A-T
M. Hph I*	G-G-T-G-A	5' G-G-T-G-^{m}A - T-C-A-C-C 3' 3' C-C-A-C - T-^{m}A-G-T-G-G 5'	10	
M. Mbo II*	G-A-A-G-A	5' G-A-A-G -^{m}A - T-C-T-T-C 3' 3' C-T-T-C - T-^{m}A -G-A-A-G 5'	10	*Mbo* II does not cut G-A-A-G-^{m}A; thus, *M. Mbo* II may methylate G-A-A-G-^{m}A
M. Nru I*	T-C-G-C-G-A	5' T-C-G-C-G-^{m}A - T-C-G-C-G-^{m}A 3' 3' ^{m}A-G-C-G-C - T-^{m}A-G-C-G-C - T 5'	12	
M. Xba I*	T-C-T-A-G-A	5' T-C-T-A-G-^{m}A - T-C-T-A-G-^{m}A 3' 3' ^{m}A-G-A-T-C - T-^{m}A-G-A-T-C - T 5'	12	*Xba* I does not cut T-^{m}C-T-A-G-A; thus, *M. Xba* I may methylate T-^{m}C-T-A-G-A

*It is assumed that the methylase will recognize the same sequence as the corresponding restriction endonuclease. *Dpn* I cleavage sequences will be generated only if methylation occurs at the 3' adenine in the methylase recognition sequence.

remain in closed circular form were those in which the region containing the *Ava* I site at position 1424 had been deleted and in which the *M. Cla* I/*Dpn* I site was ligated to the *Pvu* II site. The five ampicillin-resistant pMON2001-derived clones and the six pMON2002-derived clones characterized by restriction analysis contained the expected deletion from the *M. Cla* I/*Dpn* I site to the *Pvu* II site (data not shown).

DISCUSSION

A method for cutting specifically at the 8-base-pair sequence T-C-G-A-T-C-G-A and the 10-base-pair sequence A-T-C-G-A-T-C-G-A-T has been found.

In DNA that contains equal amounts of each base distrib-uted at random, 8- and 10-base-pair recognition sequences occur, on the average, approximately once every 65,000 (4^8) and 1,000,000 (4^{10}) base pairs, respectively. Furthermore, the sequences T-C-G-A-T-C-G-A and A-T-C-G-A-T-C-G-A-T contain two C-G dinucleotides, which are known to occur one-third to one-fourth as frequently as expected from base composition in the DNA of higher eukaryotes and many of their viruses (19–22). Thus, the 8-base-pair sequence (T-C-G-A-T-C-G-A) may occur as rarely as once in 1,000,000 base pairs and the 10-base-pair sequence (A-T-C-G-A-T-C-G-A-T) only once in 16,000,000 base pairs in higher eukaryotic DNA. In contrast, type II restriction enzyme recognition sequences occur approximately once every 256 (4^4) to 1024 (4^6) base pairs.
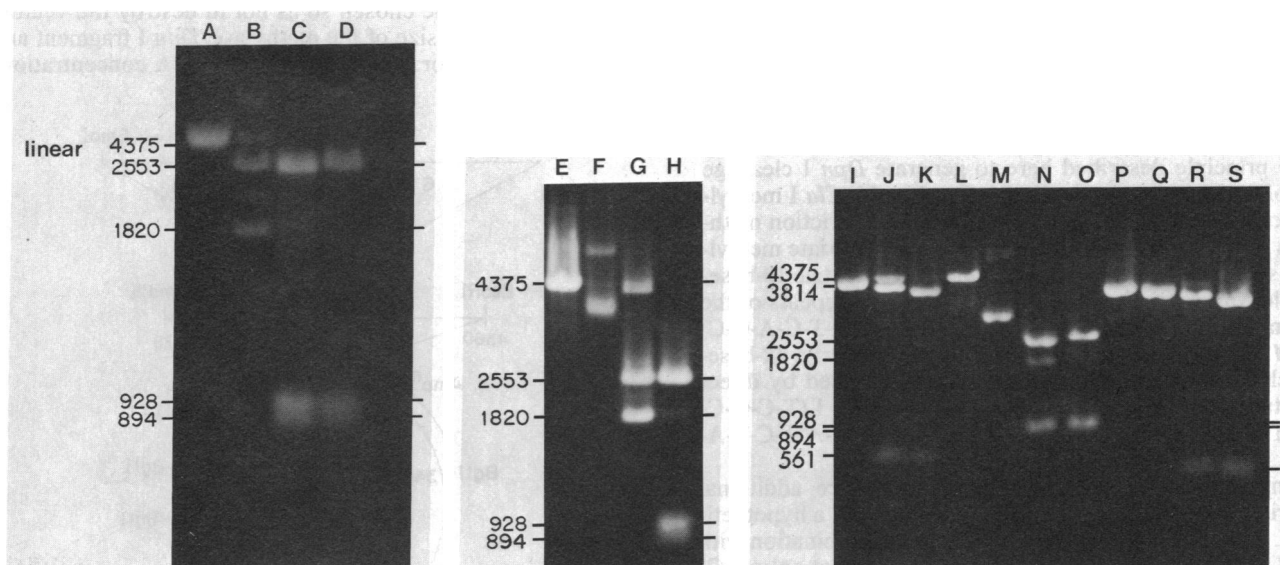

FIG. 1. Mapping of methylase/*Dpn* I cleavage sites in pMON2001 and pMON2002. Each lane contains 1 μg of DNA. Electrophoresis was through 1% agarose. Lanes: A, *M. Taq* I-methylated *Dpn* I-digested pMON2001; B, *M. Taq* I-methylated, *Bgl* I-digested pMON2001; C, *M. Taq* I-methylated *Dpn* I/*Bgl* I-digested pMON2001; D, *M. Taq* I-methylated *Dpn* I/*Bgl* I-digested pMON2002; E, *M. Cla* I-methylated, *Dpn* I-digested pMON2002; F, uncut pMON2002; G, *M. Cla* I-methylated, *Bgl* I-digested pMON2002; H, *M. Cla* I-methylated *Dpn* I/*Bgl* I-digested pMON2002; I, *M. Cla* I-methylated, *Sph* I-digested pMON2001; J, *M. Cla* I-methylated *Dpn* I/*Sph* I-digested pMON2001; K, *Sph* I/*Cla* I-digested pMON2001; L, *M. Cla* I-methylated *Dpn* I-digested pMON2001; M, uncut pMON2001; N, *M. Cla* I-methylated *Dpn* I/*Bgl* I-digested pMON2001; O, *Bgl* I/*Cla* I-digested pMON2001; P, *M. Cla* I-methylated *Dpn* I-digested pMON2002; Q, *M. Cla* I-methylated, *Sph* I-digested pMON2002; R, *M. Cla* I-methylated *Dpn* I/*Sph* I-digested pMON2002; S, *Sph* I/*Cla* I-digested pMON2002.
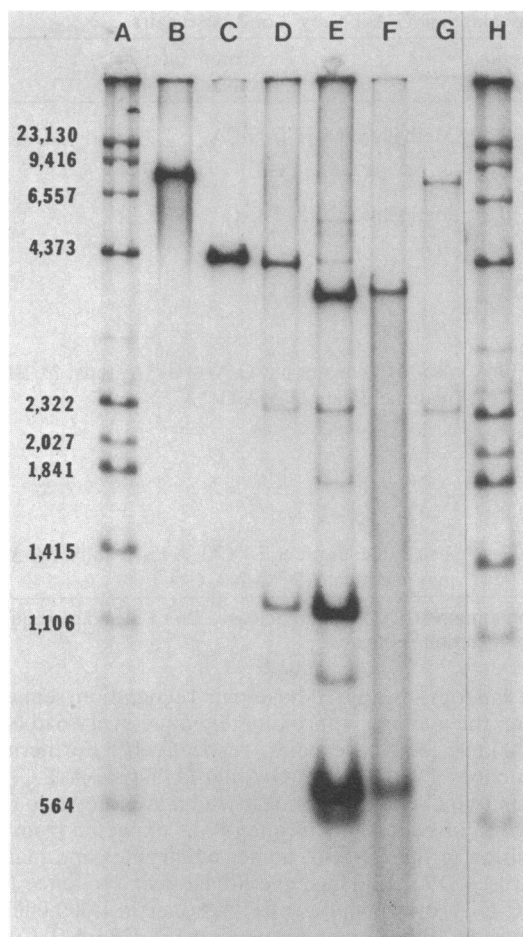
FIG. 2. Mapping of the pMON2001 derivative pLK1. A map of pLK1 is shown in Fig. 3. After restriction digestion of pLK1, fragments were end labeled by using [α-^{32}P]dCTP and electrophoresed through 1.5% agarose. DNA was visualized by autoradiography. Lanes: A and H, *Hind*III-digested phage lambda and *Hinc*II/*Pvu* II-digested pBR322 (molecular weight standards); B, *Hind*III-digested pLK1; C, *Hind*III-digested pMON2002; D, *Hind*III/*Eco*RI-digested pLK1; E, *Sal* I/*Eco*RI-digested pLK1; F, *Sal* I-digested pMON2002; G, *Sma* I/*Eco*RI-digested pMOB45.

The principle described here to generate *Dpn* I cleavage sites of 8 and 10 base pairs by *M. Taq* I or *M. Cla* I methylation could also be used with certain other restriction methylases *provided* these enzymes have the appropriate methylation specificity (Table 1). Examples of potential 10-base-pair cleavage sites for *Dpn* I include inverted repeats of the recognition sequence for *M. Hph* I (G-G-T-G-A-T-C-A-C-C) and *M. Mbo* II (G-A-A-G-A-T-C-T-T-C). Similarly, 12-base-pair cleavage sites for *Dpn* I might be generated by direct repeats of the recognition sequence for *M. Nru* I (T-C-G-C-G-A-T-C-G-C-G-A) and *M. Xba* I (T-C-T-A-G-A-T-C-T-A-G-A).

Combinations of methylases should produce additional "hybrid" sites for *Dpn* I cleavage. For instance, a hypothetical *M. Mbo* II (G-A-A-G-^{m}A) methylase in combination with *M. Cla* I would generate *Dpn* I cleavage sites not only at G-A-A-G-A-T-C-T-T-C and A-T-C-G-A-T-C-G-A-T but also at G-A-A-G-A-T-C-G-A-T and A-T-C-G-A-T-C-T-T-C. Thus the frequency of cleavage by both methylase/*Dpn* I systems simultaneously would be 4 × (1/4)^{10}, equivalent to the cleavage frequency of a restriction enzyme with a 9-base-pair recognition sequence.

A number of applications for the technique described here can be envisioned.

The sequences T-C-G-A-T-C-G-A and A-T-C-G-A-T-C-G-A-T, in addition to being very rare, contain no stop or start codons in any frame. These sequences could be inserted as a unique cloning site in a large virus or plasmid, eliminating the necessity of further engineering to remove unwanted restriction sequences. Cloning vehicles could also be developed with A-T-C-G-A-T-C-G-A-T sequences flanking the usual cloning sites. Inserts into such vehicles could then be recovered intact by *M. Taq* I/*Dpn* I or *M. Cla* I/*Dpn* I cleavage.

Cleavage at 8- or 10-base-pair recognition sequences should prove a useful step in the physical mapping of DNAs larger than 100,000 base pairs. Techniques are available for separating large DNAs (23, 24). For instance, an electrophoretic technique that can separate DNAs of up to 1,000,000 base pairs and differing in size by a few percent has recently been developed (24). This is achieved by the application of two-dimensional pulsed-field gel electrophoresis in 1% agarose gels. The method has already been used to separate yeast chromosomes (24). Subsequently, genes can be probed by using Southern blots (25).

When the pulsed-field gel electrophoretic method becomes generally available, it should be possible to detect a methylase/*Dpn* I fragment containing a gene by Southern blotting of methylase/*Dpn* I-digested total genomic DNA. λ Charon phage (26) shotgun clones of this fragment could then be used for chromosome walking in the fragment containing the gene. One advantage of using such a fragment to construct λ phage libraries would be to reduce the complexity of the DNA being probed in each step. This would reduce the amount of phage to be screened and the problems encountered with regions of repetitive DNA.

This technique may also allow chromosome walking from known genes in jumps of up to hundreds of kilobases. This would be achieved by first identifying genes in Southern blots of methylase/*Dpn* I restriction fragments of total chromosomal DNA. These fragments would then be isolated and end fragments would be subcloned. This could be achieved by ligating a linearized cloning vector to the isolated methylase/*Dpn* I fragment. This would be followed by cleavage with a restriction enzyme chosen so as not to destroy the vector while reducing the size of the methylase/*Dpn* I fragment attached to the vector. Religation at low DNA concentration
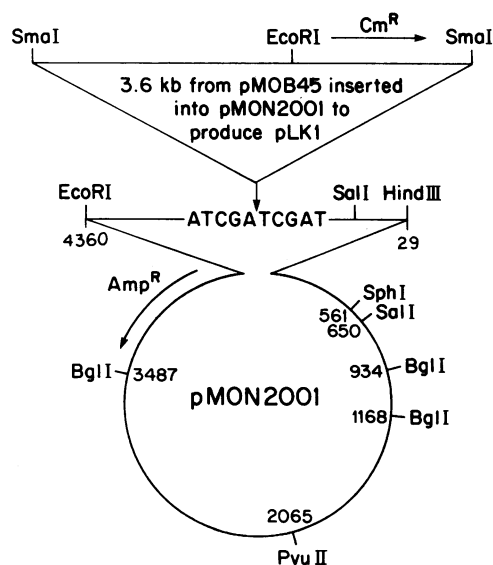

FIG. 3. Restriction map of pMON2001 and pLK1. Nucleotide positions of restriction endonuclease recognition sites are relative to the *Eco*RI site at 4360 in pBR322. kb, Kilobases; Cm^R, chloramphenicol resistance. Amp^R, ampicillin resistance.

will lead to circularization of DNA. Vectors with inserts should primarily be of two types—i.e., containing DNA extending from either end of the methylase/*Dpn* I fragment to the first restriction site in the fragment. These end fragments would in turn be used to probe a complete λ phage genomic bank for overlapping sequences. Appropriate phage clones would be used to probe Southern blots of the methylase/*Dpn* I-digested total genomic DNA. In this way fragments adjacent to the one containing the gene would be detected. This process could then be repeated with subclones of the ends of these adjacent fragments.

It may also be possible to study long-range heterogeneity in large genomes by using Southern blots (25) of methylase/*Dpn* I-digested DNA, separated by 1% agarose pulsed-field gel electrophoresis (24). Loss or gain of a methylase/*Dpn* I site and large deletions or insertions between closely related organisms may be detected over a range of many kilobases.

The techniques described here, by permitting site-specific cleavage of DNA at 8- and 10-base-pair sequences, should significantly increase our ability to study large genomes.

1. Modrich, P. (1979) *Q. Rev. Biophys.* **12**, 315–369.
2. Roberts, R. J. (1983) *Nucleic Acids Res.* **11**, r135–r168.
3. Smith, H. O. & Nathans, D. (1973) *J. Mol. Biol.* **81**, 419–423.
4. McClelland, M. (1981) *Nucleic Acids Res.* **9**, 5859–5866.
5. McClelland, M. (1983) *Nucleic Acids Res.* **11**, r169–r173.
6. McClelland, M. (1981) *Nucleic Acids Res.* **9**, 6795–6804.
7. Sato, S., Nakazowa, K. & Shuromiya, T. J. (1980) *Biochemistry* **88**, 737–747.
8. Lacks, S. & Greenberg, B. (1980) *J. Mol. Biol.* **114**, 153–168.
9. Geier, G. E. & Modrich, P. (1979) *J. Biol. Chem.* **254**, 1408–1413.
10. Adams, S. P., Kavka, K. S., Wykes, E. J., Holder, S. B. & Galluppi, G. R. (1983) *J. Am. Chem. Soc.* **105**, 661–663.
11. Lansman, R. A., Shade, R. O., Shapira, J. F. & Avise, J. C. (1981) *J. Mol. Evol.* **17**, 214–226.
12. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
13. Dagert, M. & Ehrlich, S. D. (1979) *Gene* **6**, 23–28.
14. Ish-Horowicz, D. & Burke, J. F. (1981) *Nucleic Acids Res.* **9**, 2983–2998.
15. Marinus, M. G. & Morris, N. R. (1973) *J. Bacteriol.* **114**, 1143–1150.
16. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
17. Bittner, M. & Vapnek, D. (1981) *Gene* **15**, 319–329.
18. Dreiseikelmann, B., Eichenlaub, R. & Wachernagel, W. (1979) *Biochim. Biophys. Acta* **562**, 418–428.
19. Taylor, H. J. (1979) in *Molecular Genetics*, ed. Taylor, H. J. (Academic, New York), Vol. 3, pp. 89–116.
20. Salser, W. A., Cummings, I., Lui, A., Strommer, J., Padayatty, J. & Clarkie, P. (1979) in *Cellular and Molecular Switching*, eds. Stamatoyannopoulos, G. & Nienhius, A. W. (Grune & Stratton, New York), p. 621.
21. McClelland, M. (1982) *Nucleic Acids Res.* **10**, 7865–7877.
22. McClelland, M. (1983) *J. Mol. Evol.* **19**, 346–354.
23. Fangman, W. L. (1978) *Nucleic Acids Res.* **5**, 653–665.
24. Schwartz, D. C., Saffran, W., Welsh, J., Haas, R., Goldenberg, M. & Cantor, C. R. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 189–195.
25. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–518.
26. Williams, B. & Blattner, F. (1977) in *Genetic Engineering*, eds. Setlow, J. K. & Hollander, A. (Plenum, New York), Vol. 2, p. 201.