MOLECULAR AND CELLULAR BIOLOGY, July 1982, p. 758–762
0270-7306/82/070758-05$02.00/0

Vol. 2, No. 7

Topographical Distribution of 5-Methylcytosine in Animal and Plant DNA

TALLY NAVEH-MANY AND HOWARD CEDAR*

Department of Cellular Biochemistry, The Hebrew University-Hadassah Medical School, Jerusalem 91 010, Israel

Received 26 February 1982/Accepted 23 March 1982

The topographical distribution of 5-methylcytosine on animal and plant cell DNA has been examined with methyl-sensitive restriction enzymes and gel electrophoresis analysis. These DNAs digested with the enzyme *Hpa*II have a partially bimodal size distribution, indicating the existence of clusters of methylated and unmethylated CCGG sites in the DNA. By analyzing the methylation state of all CG moieties in restricted DNA fractions, it was possible to show that these genomes are, in general, arranged as clusters of relatively highly methylated and undermethylated regions. Plant DNA also contains 5-methylcytosine in the prototype sequence C-X-G. Restriction of this DNA with *Eco*RII revealed that these methyl groups are also distributed in clusters, suggesting that this is a general phenomenon. The undermethylated areas may correspond to the active fraction of the genome.

Although 5-methylcytosine is found in almost all eucaryotic organisms, its distribution on DNA is not well understood. In animal cells, over 95% of the DNA methyl moieties are found in the dinucleotide sequence cytosine-guanine (C-G) (3). In higher plants, however, additional methyl groups are located in the trinucleotide sequence C-X-G, where X can be C, A, or T (6). In both instances, the DNA appears to be symmetrically methylated at each modified site, allowing the pattern of methylation to be inherited by succeeding cell generations in a semiconservative manner (14).

Several restriction enzymes have been found useful for the study of DNA methylation. The enzyme *Hpa*II recognizes the C-G-containing sequence CCGG but cleaves the DNA only if the internal cytosine is unmethylated (18). When sea urchin DNA is digested by this enzyme, the resulting fragments, when analyzed by gel electrophoresis, are distributed in two clearly distinguishable fractions (2). These results suggest that the methyl moieties in this organism are nonrandomly placed on the DNA and appear to be arranged in long clusters of fully methylated or fully unmethylated regions. Analysis of these subpopulations with specific gene probes indicates that some expressed gene sequences are in the unmethylated compartment (1, 2). Other organisms show a similar pattern of methylation on their DNA (1, 19).

In this paper, we analyze the distribution of 5-methylcytosine in animal and higher plant cells. Although these DNAs are more highly methylated than DNAs from invertebrate organisms, a similar pattern of methyl clustering can be observed. The regions which contain stretches of unmethylated C-G residues most probably represent the active fraction of the genome.

MATERIALS AND METHODS

Determination of average molecular size. Wheat germ or mouse liver DNA was purified (14) and digested with various restriction enzymes at a ratio of 2 enzyme units per μg for 2 h in specific buffers recommended by the enzyme manufacturers (New England Biolabs and Bethesda Research Laboratories). The restricted DNA (5 to 10 μg) was subjected to electrophoresis on 0.6% agarose slab gels. Each gel was run with molecular-length marker bands obtained by restriction enzyme digestion of various plasmid and phage molecules. After 20 h of electrophoresis at 20 V at room temperature, the ethidium bromide-stained gels were photographed with positive-negative Polaroid type 665 film, and the film negative was assayed with a scanning spectrophotometer. In some cases, tritium-labeled DNA (6,000 cpm/μg) obtained from [^{3}H]thymidine-labeled mouse L-cells (16) was cleaved by restriction enzymes and analyzed by gel electrophoresis. The distribution of these digests was measured by scintillation counting of equal gel slices. The resulting graphs were computer digitalized, and the average molecular size was calculated by reference to the molecular-weight markers. Each distribution was arbitrarily divided into approximately 35 fractions and normalized to obtain the percentage of the total DNA sample in each fraction.

The theoretical distribution of size classes of DNA cleaved by specific restriction enzymes was calculated from the following considerations. Let P be the probability of a certain restriction site appearing in the DNA. Then the probability of obtaining a particular length of DNA, n, can be expressed as $n(1 - P)^{n-1}P$.

The percentage of DNA representing a specific size class (from $n = x$ to $n = y$) can then be written as:

$$\sum_{n=x}^{n=y} n(1-P)^{n-1}P \Big/ \sum_{n=1}^{\infty} n(1-P)^{n-1}P$$

which can be simplified to:

$$P^2 \sum_{n=x}^{n=y} n(1-P)^{n-1}$$

In all cases, the size fractions were arbitrarily chosen to correspond to the logarithmic distribution of sizes obtained from the gel electrophoresis analysis. P is equal to the reciprocal of the computed number average molecular size for any particular restriction digest.

RESULTS

Analytical gel electrophoresis of restricted DNA provides several types of information about the placement of a particular site on the cellular DNA. By reference to molecular size markers, one can calculate the number average molecular weight of the DNA and thereby obtain the frequency of appearance of this site. By studying the distribution pattern of these restriction fragments, one can deduce the topographical arrangement of these sites on the DNA backbone. Any deviation from a normal distribution indicates nonrandom placement of a particular site. With methylation-sensitive restriction enzymes, this method can be used to map the locations and clustering of DNA methyl moieties.

When mouse liver DNA was digested with *Hpa*II and subjected to analytical gel electrophoresis, the DNA was found to have a number average molecular size of about 7,000 nucleotide pairs. Although most of the DNA migrated as one distinct peak, a broad shoulder in the region of 2,000 base pairs was consistently observed in all DNA preparations digested to completion with *Hpa*II (Fig. 1). This shoulder was not seen when the same DNA was cleaved by enzymes such as *Msp*I and *Eco*RI, which do not recognize C-G methylations, but was evident in DNA digested with *Hha*I (data not shown), another enzyme which is sensitive to methylation in its restriction site. A similar semi-bimodal pattern was observed with DNA from several mouse and chicken tissues, including sperm. Furthermore, the distribution pattern of *Hpa*II-digested, tritium-labeled mouse L-cell DNA was found to be identical to the pattern observed for ethidium bromide-stained DNA. In all cases, restriction was shown to be complete by the fact that higher concentrations of restriction enzyme yielded identical results.

To show that this partially bimodal pattern is indeed significant, we compared these gel analyses with the expected distribution of this DNA as calculated from the average number of sites for each restriction enzyme. The computer-generated distribution patterns are also shown in Fig. 1. Although the theoretical curve for the enzymes *Eco*RI and *Msp*I fitted closely to the experimental results, there was a marked difference between the calculated and experimental distribution for *Hpa*II cleavage. The contrast between the theoretical and empirical curves points out the bimodal nature of the *Hpa*II restriction distribution.

These results strongly suggest that methylated *Hpa*II sites are distributed in a nonrandom fashion on mouse liver DNA in such a way that there are large clusters of heavily methylated regions as well as stretches of undermethylated DNA. This hypothesis was examined directly by using

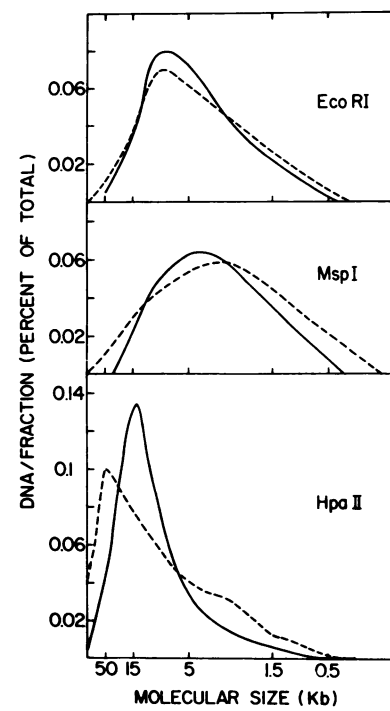

FIG. 1. Size distribution of mouse liver DNA after restriction enzyme digestion. DNA was cleaved to completion and subjected to 0.6% agarose gel electrophoresis together with approximately 20 molecular-weight markers ranging in size from 50 to 0.4 kilobases. Gel photographic negatives were stained and normalized as described in the text and are shown as a function of molecular size. (-----), Experimental results, (——), theoretical distribution as computed from the number average molecular size of each digest: *Eco*RI, 3.2 kilobases; *Msp*I, 2.3 kilobases; *Hpa*II, 6.5 kilobases. Total unrestricted DNA had a number average molecular size of 40 kilobases.

TABLE 1. Determination of methyl clusters by *Msp*I digestion of *Hpa*II-fractionated DNA[a]

DNA	Molecular size (kb[b]) of:		Average methyl cluster
	Untreated DNA	*Msp*I-digested DNA	
Total	6,300	2,100	3.2
Fraction I	21,000	3,400	6.2
Fraction II	12,000	2,800	4.3
Fraction III	3,200	1,600	2.0

[a] Mouse DNA was cleaved to completion with *Hpa*II and fractionated by sucrose gradient centrifugation with a 10 to 40% sucrose gradient containing 20 mM Tris-hydrochloride (pH 7.9), 1 M NaCl, and 10 mM EDTA (10). This method yielded a distribution similar to that obtained by gel analysis, with somewhat less resolution. Fractions were collected, and the DNA was recovered by ethanol precipitation. Each DNA fraction was treated with *Msp*I and analyzed by analytical gel electrophoresis with 0.6% agarose slab gels. The number average molecular size for both untreated DNA and *Msp*I-cleaved DNA was then determined by computer analysis. The ratio of the size of the untreated DNA to that of the *Msp*I-treated DNA represents the average number methylated CCGG clusters for each fraction.

[b] kb, Kilobases.

*Hpa*II and *Msp*I digestion to measure the extent of methylation at CCGG sites. DNA was digested with *Hpa*II and preparatively size fractionated on sucrose gradients. Several size fractions were then collected, digested with *Msp*I, and analyzed by analytical gel electrophoresis to determine the number average size of the DNA. The extent of methylation of each fraction could be determined by comparing the average size of the DNA before and after *Msp*I cleavage. The large DNA produced by *Hpa*II digestion contained multiple clustered methyl moieties, whereas the smaller DNA fragments were significantly undermethylated (Table 1). These results demonstrate that the bimodal pattern obtained by *Hpa*II restriction is due to the unique distribution of methylated sites and is not the result of clustering of CCGG sites.

Although these experiments convincingly show that *Hpa*II methylations are clustered on the DNA, these particular methyl moieties represent only about 6% of the C-G methylations in cellular DNA. To analyze the distribution of all methyl moieties, total DNA was labeled by nick translation with [α-^{32}P]dGTP as the sole nucleotide source. When this DNA was subjected to nearest-neighbor analysis, the C-G dinucleotide was found to be 67% methylated. This labeled DNA was digested to completion with *Hpa*II and fractionated by gel electrophoresis. Several fractions were extracted from the gel and analyzed by nearest-neighbor analysis. Large *Hpa*II fragments were heavily methylated, whereas smaller DNA pieces were quantitatively undermethylated (Table 2). This observation strongly indicated that all C-G methylations are nonrandomly arranged on the DNA in highly methylated and relatively undermethylated compartments.

When wheat germ DNA was digested to completion with *Hpa*II, results similar to those shown in Fig. 1 were obtained. Whereas animal cell DNA contains methyl groups almost exclusively at the dinucleotide sequence C-G, plant DNA is highly methylated, and additional methylation sites are observed. These plant-specific methyl moieties are located in the prototype trinucleotide sequence C-X-G, where X can be either adenine, thymine, or cytosine (6). Methylation at these sites can be measured with specific restriction enzymes which are sensitive to methylation in the C-X-G-containing restriction site. Thus, the enzymes *Pvu*II (CAGCTG), *Pst*I (CTGCAG), and *Eco*RII (CC$\overset{A}{\underset{T}{}}$GG) cut plant DNA very poorly owing to extensive methylation at these sites. Although the role of these unusual methylations is not understood, it was of interest to determine their distribution. *Eco*RII digests wheat germ DNA in a distinct bimodal pattern reminiscent of the distribution

TABLE 2. C-G methylation of *Hpa*II-fractionated mouse DNA[a]

DNA	C-G methylation (%)
Total *Hpa*II restricted	67
Fraction I	69
Fraction II	53
Fraction III	42

[a] Total mouse DNA was nick translated in the presence of [α-^{32}P]dGTP as the sole nucleoside triphosphate to an intentionally low specific activity of 50,000 cpm/μg to obtain relatively unsheared DNA. The resulting DNA was cleaved to completion by the restriction enzyme *Hpa*II and fractionated by electrophoresis with 0.6% low-melting agarose. Several gel fractions were cut, extracted (13), and subsequently examined for C-G methylation by nearest-neighbor analysis (7). The results are presented as the percentage of methylated cytosine in all C-G residues. Total unrestricted labeled DNA was also subjected to gel electrophoresis, extraction, and nearest-neighbor analysis. *Hpa*II-cleaved DNA was divided into three large fractions. Fraction I contained the largest 70% of the DNA (>7 kilobases), Fraction II contained the next largest 20% of the DNA (3 to 7 kilobases), and Fraction III contained the smallest 10% of the DNA (<3 kilobases). The results shown in the table represent the average data obtained from six separate experiments.

obtained for C-G methylations (Fig. 2). *Bst*NI, an isoschizomer of *Eco*RII which is unaffected by cytosine methylation, cleaves this DNA, yielding an apparently normal distribution with a number average size of 530 base pairs. The *Eco*RII digestion has a shoulder in this same size range, indicating that a fairly large fraction of the plant DNA has clustered regions of undermethylation. DNA from tobacco and cauliflower gave similar results.

DISCUSSION

Previous results have demonstrated that methylated CCGG sites are found in a clustered organization in some invertebrate organisms (1, 19). The data shown in this paper extend these findings to include all methylated C-G residues in animal cell DNA and methylated C-X-G moieties in plant DNA. This suggests that the compartmentalization into methylated and unmethylated DNA regions may be a general phenomenon.

Bird et al. originally pointed out that one could study the arrangement of DNA methyl moieties by looking at the size distribution of DNA digested to completion with *Hpa*II (2). A normal distribution would suggest that the methyl groups are placed randomly on the DNA, whereas deviation from this curve might indicate clustering of methylation. In the case of sea urchin DNA, *Hpa*II cleavage yields two distinct size classes of DNA, suggesting compartmentalization of methylated and unmethylated regions. The same technology used on animal cell DNA yielded qualitatively similar results. Owing to several factors, gel analysis of this DNA revealed a quantitatively small undermethylated compartment. Animal cell DNA is 70 to 80% methylated at CCGG sites, and *Msp*I sites are much less frequent in mouse DNA (every 2,100 base pairs) (12), making gel analysis difficult, since clusters of even 10 methylated sites would yield DNA too large to be resolved by our gel systems and close to the size of unrestricted DNA. In this regard, it should be noted that we attempted analysis on 0.25% agarose gels. Although clustering was nicely visualized, we chose not to use these gels since, for unknown reasons, average molecular-weight determination was found to be extremely inaccurate.

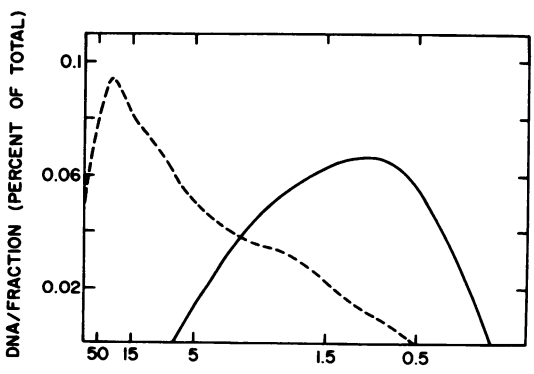

FIG. 2. Size distribution of wheat germ DNA after restriction enzyme digestion. DNA was cleaved, subjected to 0.6% agarose gel electrophoresis, scanned, and normalized as described in the text. (-----), Cleavage distribution obtained from *Eco*RII (number average size, 6 kilobases); (——), result of *Bst*NI cleavage (number average size, 0.5 kilobases).

The clustering of methylated *Msp*I sites is probably representative of all methylated C-G residues, as shown by the fact that fragments cleaved by *Hpa*II were relatively unmethylated at other sites. The most likely explanation for the clustered distribution of methyl moieties is that undermethylated regions represent the active portion of the genome, whereas inactive DNA is heavily methylated. Many examples point to a correlation between gene expression and undermethylation. Several individual genes were found to be highly methylated at restriction sites in germ line DNA but relatively undermethylated at specific sites in somatic cells which express the gene, but not in those where the gene is silent (9, 15, 18). Other evidence shows that all expressed genes in a particular cell are about 30% as methylated as the average cellular DNA (12). In sea urchin DNA, all active genes which were examined were found to be present in the unmethylated compartment (1). Since animal DNA is 70 to 80% methylated, on the average, every fourth C-G will be unmethylated. If there exist enough areas which are relatively undermethylated or overmethylated, this will obviously appear as clustering in a gel distribution analysis. The extent of clustering cannot be determined accurately owing to the limitation of the gel electrophoresis system.

Although much is known about the inheritance of a particular methylation pattern, very little has been learned about the mechanisms involved in altering this pattern. In particular, specific demethylation must occur during differentiation, and de novo methylation is known to occur during viral transformation by certain viruses (4, 17). In this regard, it is interesting to note that demethylation can be induced artificially by the use of certain drugs, such as 5-azacytidine. This compound is thought to act during DNA replication by replacing cytosine with a compound which cannot subsequently undergo methylation (8). This drug causes the induction of otherwise inactive genes, such as the chicken endogenous viral genome (5) and the

X-chromosome-associated hypoxanthine phosphoribosyl transferase gene (11), presumably by causing demethylation. This undermethylation must be partially specific to certain regions of the genome, since other genes do not seem to undergo extensive demethylation (5). This suggests that this drug-induced undermethylation may occur processively in long clusters. If this same process were to occur under normal circumstances, it might explain the general clustering of methyl moieties in eucaryotic DNA.

ACKNOWLEDGMENTS

We acknowledge the advice of A. Razin and the assistance of C. Gopin in the preparation of this manuscript.

This research was supported by Public Health Service grants GM-20483 from the National Institutes of Health and by grant 2333/80 from the U.S.-Israel Binational Science Foundation.

LITERATURE CITED

1. **Bird, A. P., and M. H. Taggart.** 1980. Variable patterns of total DNA and rDNA methylation in animals. Nucleic Acids Res. **8**:1485–1497.
2. **Bird, A. P., M. H. Taggart, and B. A. Smith.** 1979. Methylated and unmethylated compartments in the sea urchin genome. Cell **17**:889–901.
3. **Bonen, L., T. Y. Huh, and T. W. Gray.** 1980. Can partial methylation explain the complex fragment pattern observed when plant mitochondrial DNA is cleaved with restriction endonucleases? FEBS Lett. **111**:340–346.
4. **Desrosiers, R. C., C. Mulder, and B. Fleckenstein.** 1979. Methylation of *Herpesvirus saimiri* DNA in lymphoid tumor cell lines. Proc. Natl. Acad. Sci. U.S.A. **76**:3839–3843.
5. **Groudine, M., R. Eisenman, and H. Weintraub.** 1981. Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. Nature (London) **292**:311–317.
6. **Gruenbaum, Y., T. Naveh-Many, H. Cedar, and A. Razin.** 1981. Sequence specificity of methylation in higher plant DNA. Nature (London) **292**:860–862.
7. **Gruenbaum, Y., R. Stein, H. Cedar, and A. Razin.** 1981. Methylation of CpG sequences in eukaryotic DNA. FEBS Lett. **124**:67–71.
8. **Jones, P., and S. Taylor.** 1980. Cellular differentiation, cytidine analogs and DNA. Cell **20**:85–93.
9. **Mandel, J. L., and P. Chambon.** 1979. DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. Nucleic Acids Res. **7**:2081–2090.
10. **Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, J. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis.** 1980. The isolation of structural genes from libraries of eukaryotic DNA. Cell **15**:687–701.
11. **Mohandas, T., R. S. Sparkes, and L. J. Shapiro.** 1981. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science **211**:393–396.
12. **Naveh-Many, T., and H. Cedar.** 1981. Active gene sequences are undermethylated. Proc. Natl. Acad. Sci. U.S.A. **78**:4246–4250.
13. **Parker, R. C., and B. Seed.** 1980. Two-dimensional agarose gel electrophoresis "Sea Plaque" agarose dimension. Methods Enzymol. **65**:358–363.
14. **Pollack, Y., R. Stein, A. Razin, and H. Cedar.** 1980. Methylation of foreign DNA sequences in eukaryotic cells. Proc. Natl. Acad. Sci. U.S.A. **77**:6463–6467.
15. **Shen, C. K. J., and T. Maniatis.** 1980. Tissue-specific DNA methylation in a cluster of rabbit β-like globin genes. Proc. Natl. Acad. Sci. U.S.A. **77**:6634–6638.
16. **Solage, A., and H. Cedar.** 1978. Organization of 5-methylcytosine in chromosomal DNA. Biochemistry **17**:2934–2938.
17. **Sutter, D., and W. Doerfler.** 1980. Methylation of integrated adenovirus type 12 DNA sequences in transformed cells is inversely correlated with viral gene expression. Proc. Natl. Acad. Sci. U.S.A. **77**:253–256.
18. **van der Ploeg, L. H. T., and R. A. Flavell.** 1980. DNA methylation in the human γ5β-globin locus in erythroid and nonerythroid tissues. Cell **19**:947–958.
19. **Whittaker, P., and N. Hardman.** 1980. Methylation of nuclear DNA in *Physarum polycephalum*. Biochem. J. **191**:859–862.