METHYLATION OF CpG SEQUENCES IN EUKARYOTIC DNA

Yosef GRUENBAUM*, Reuven STEIN, Howard CEDAR and Aharon RAZIN*

**Department of Cellular Biochemistry* and *Department of Molecular Biology, The Hebrew University, Hadassah Medical Shool, PO Box 1172 Jerusalem, Israel*

Received 12 December 1980

1. Introduction

An understanding of the function of DNA methylation in eukaryotes will require a clearer, more precise, knowledge of the distribution of methylated bases in the eukaryotic genome. Although it has been known for some time that 5-methylcytosine (m^5Cyt) is the only methylated base in eukaryotes, little is known about the sequence specificity of this modification. One important fact which has emerged from base analyses is that >90% of the m^5Cyt residues are found in the sequence CpG [1]. It was therefore of interest to determine what fraction of this dinucleotide sequence is methylated. The answer to this question would not only contribute to our knowledge of the distribution of m^5Cyt but would also shed light on the factors which make up the recognition signal for eukaryotic methylation. In an attempt to answer this question we have developed a new experimental procedure for detecting m^5Cyt in CpG sequences. This dinucleotide is found to be highly methylated in animal cells and may be the major determinant in the placement of methyl groups on DNA.

2. Materials and methods

Enzymes were obtained from the following sources: restriction endonuclease *Hpa*II from Bethesda Res. Labs; its isischizomer *Msp*I and *Escherichia coli* DNA polymerase I from New England Biolabs; intestinal alkaline phosphatase from Boehringer Mannheim; micrococcal nuclease, pancreatic DNase I and venom phosphodiesterase from Sigma and spleen phosphodiesterase from Worthington Biochemical Co. Radioactive [α-^{32}P]deoxynucleoside triphosphates, [γ-^{32}P]-ATP and *S*-[*methyl*-^{3}H]adenosyl methionine were obtained from Amersham, England. DNA was prepared from *E. coli,* mouse liver, sea urchin (*Echinometria mathaei*) sperm, and *Drosophila melanogaster* larvae as in [2]. Grade I, highly polymerized calf thymus DNA was obtained from Sigma. Herpes simplex virus I DNA, λ phage DNA and rabbit liver DNA were gifts from Y. Becker, G. Glaser and J. Singer, respectively.

Partially purified DNA methylase was prepared from mouse Ehrlich ascites tumor cells as in [3]. The preparation used in these studies was purified through the phosphocellulose chromatography step (fraction III, 1.3 mg protein/ml). Methylation in vitro was carried out in a reaction mixture of 25 μl containing 0.2 mM dithiothreitol, 10 mM EDTA, 10 mM Tris–HCl (pH 7.9), 80 mM NaCl, 10 μg/ml phenylmethyl sulfonyl fluoride, 2% (v/v) glycerol, 6 μM *S*-[*methyl*-^{3}H]adenosylmethionine (15–50 Ci/mmol), 2.5 μg heat-denatured DNA and 7 μl enzyme. The reaction lasted for 30 min at 37°C and terminated by the addition of 0.2 ml sodium lauryl sulfate (25%, w/v), 250 μg calf thymus carrier DNA and brought to 2 ml with water. The mixture was incubated for 10 min at 60°C extracted with chloroform:isoamyl alcohol (24:1) and the DNA precipitated with perchloric acid. The pellet was treated with 0.5 M NaOH for 10 min at 60°C in order to remove RNA and the labelled DNA was finally precipitated with trichloroacetic acid and collected on Whatman GF/C filters and quantitated by liquid scintillation counting.

2.1. *Assay of m^5Cyt in CpG containing sequences*

The degree of methylation of CpG containing sequences has been determined by a new method described here which represents an extension of the standard nearest neighbor analysis [4,5]. DNA samples (5 μg) were nicked using either pancreatic DNase I (0.5 μg/ml for 6 min at 37°C) or sonication (5 min at

full power in an Artex sonicator). The nicked DNA was incubated for 10 min at 15°C in 50 μl containing 50 mM potassium phosphate (pH 7.2), 5 mM $MgCl_2$, 1 mM mercaptoethanol, 1.2 μM d[α-^{32}P]GTP (400 Ci/mmol) and 7 units *Escherichia coli* DNA polymerase I. The reaction was stopped by the addition of 5 mM EDTA and the unreacted dGTP removed by Sephadex G-50 chromatography. The labelled DNA was digested to deoxynucleoside 3′-monophosphates in 100 mM Tris–HCl (pH 8.5), 10 mM $CaCl_2$ using micrococcal nuclease (140 μg/ml) and spleen phosphodiesterase (7 units/ml) for 3 h at 37°C. An aliquot of the digest was applied directly to cellulose thin-layer chromatography sheets (Eastman-Kodak) and chromatographed in two dimensions as in [6]. After autoradiography the radioactive spots were quantitated by scintillation counting.

3. Results

In order to determine the extent of methylation of the CpG sequence in eukaryotic DNA we have developed a technique for measuring the nearest neighbors of m^5Cyt. m^5Cyt cannot be detected by the conventional nearest neighbor analysis technique, since in this technique only the in vitro synthesized nucleotides are analyzed. Thus wherever m^5Cyt is present in the native DNA it will be exchanged by cytosine in the newly synthesized labelled DNA. In order to determine m^5Cyt in native DNA we have inserted nicks into DNA with either DNase I or by sonication and have used these randomly placed nicks as primer for *E. coli* DNA polymerase I using d[α-^{32}P]GTP as the only available nucleotide. After digestion to nucleoside 3′-monophosphates only the nucleotide which was on the 5′-side of the nick will be labelled. Since m^5Cyt is found almost exclusively in the sequence CpG this technique should yield 5 spots upon chromatography; the corresponding 3′-mononucleotides of the 4 major bases and of m^5Cyt. The ratio of the label found in the spots corresponding to 3′-methylcytidylic acid and 3′-cytidylic acid determines that fraction of the dinucleotide CpG which is methylated at the Cyt residue. The results of such an analysis obtained with two different samples of DNA are presented in fig.1. These results clearly demonstrate that while mouse liver DNA is highly methylated at its CpG sequences, only a small fraction of the CpG sequences in sea urchin DNA are methylated. Similar analyses performed with α-^{32}P-labelled dATP dCTP and TTP confirmed data indicating that m^5Cyt is found almost exclusively in the dinucleotide CpG [1]. The underrepresentation of the CpG dinucleotide in eukaryotic DNA [5] is also apparent in fig.1.

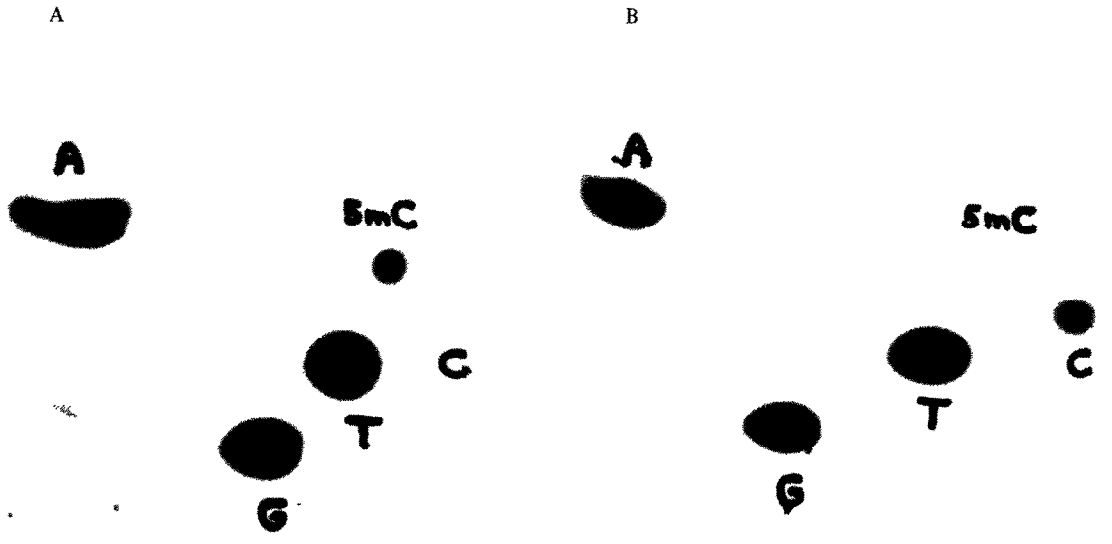

Fig.1. Identification of m^5Cyt by thin-layer chromatography: (A) mouse liver DNA; (B) sea urchin DNA were nick translated with d[α-^{32}P]GTP as in section 2. The labelled DNA was enzymatically digested to nucleoside 3′-monophosphates, separated by two-dimensional thin-layer chromatography and autoradiographed [6]. The various nucleotides are identified in the figure: (A) dAMP; (G) dGMP; (T) dTMP; (C) dCMP; (5mC) 5 methyl-dCMP.

Table 1
Methyl content of CpG sequences and its subset CCGG

Source of DNA	C^mpG (% total CpG)		CC^mGG (% total CCGG)
	A	B	
Micrococcus luteus	0	0	0
Drosophila melanogaster (larvae)	0	0	0
Sea urchin (sperm)	29	26	15
Rabbit (liver)	39	48	50
Mouse (liver)	62	71	72
Calf (thymus)	75	82	87

The methylated CpG sequences were determined as in section 2 using either DNase I (A) or sonication (B) for introducing nicks in the DNA. The degree of methylation of the sequence CCGG was determined as in [6]

Quantitative data concerning the extent of methylation of CpG sequences are summarized in table 1. Whereas mammalian DNA is highly methylated in CpG sequences (>50%) other DNA samples like *Micrococcus luteus* and *D. melanogaster* are completely unmethylated at these sites.

Two separate methods were employed for introducing random nicks into DNA, sonication and DNase I treatment. The results of both of these techniques are presented in table 1. In general the sonication technique yielded slightly higher levels of m^5Cyt. This is probably the preferred technique since the nearest neighbor analyses of other dinucleotides agreed well with the published results obtained by the conventional analysis. DNase I, on the other hand, nicks with some preferences for particular nucleotides.

The restriction enzymes *Hpa*II and *Msp*I have been used to assay methylation at the sequence CCGG in eukaryotic DNA [7]. Although both enzymes recognize the same sequence (CCGG) only *Msp*I cuts when the internal C is methylated. By treating DNA with *Msp*I and labelling the 5′-end of each molecule using polynucleotide kinase one can label the internal C of the sequence CCGG. After digestion to nucleoside 5′-monophosphates and chromatography one can determine what percentage of this site is methylated [6]. These data, for various DNA samples, are shown in table 1. There is clearly a close correlation between the amount of methylation at all CpG residues and those at the specific sequence CCGG.

Numerous investigators have studied DNA methylation in vitro using purified and semi-purified eukaryotic methylases [8–10]. It is clear from these studies that the source of the DNA template is an important factor in determining the methylase activity. A careful analysis of these data shows that the methylase activity is determined primarily by the CpG content of the DNA template. Using a DNA methylase from mouse ascites tumor cells we have tested the reactivity of various DNA templates. As shown in fig.2 the DNA methylase activity is linearly dependent on the CpG content of the DNA. Fig.2 also contains a compilation of data from 4 separate papers concerned with in vitro mehylation of DNA. Although in each case the eukaryotic methylase was partially purified from a different organism, the activity of every enzyme was proportional to the CpG content. The results obtained with SV-40 DNA are especially informative. Although both SV-40 DNA and *E. coli* DNA have about the same base composition, the CpG content of SV-40 is only 0.6% whereas that of *E. coli* is 6.7%. Indeed SV-40 is a poor template for methylation and the methylase showed 12-fold higher activity with *E. coli* DNA. It should be noted that since eukaryotic DNA is partially methylated at the CpG residues, the actual number of available CpG residues is lower than that indicated in the figure. This correction, however, would not change the curve shown in fig.2. It is clear from these results that the low in vitro methylation activity of animal DNA is not due to the presence of pre-existing methyl moieties, but is rather a consequence of the low CpG content of these templates. Several investigators have studied the activity of eukaryotic DNA methylases on synthetic polymers. Whereas the homopolymer $(dC)_n \cdot (dG)_n$ is a poor substrate, the alternating polymers $(dG \cdot dC)_n \cdot (dG \cdot dC)_n$ is the most efficient known template for DNA methylation [8–10]. Although not shown in fig.2 this template falls on the same linear line as the other DNA templates shown in the figure.

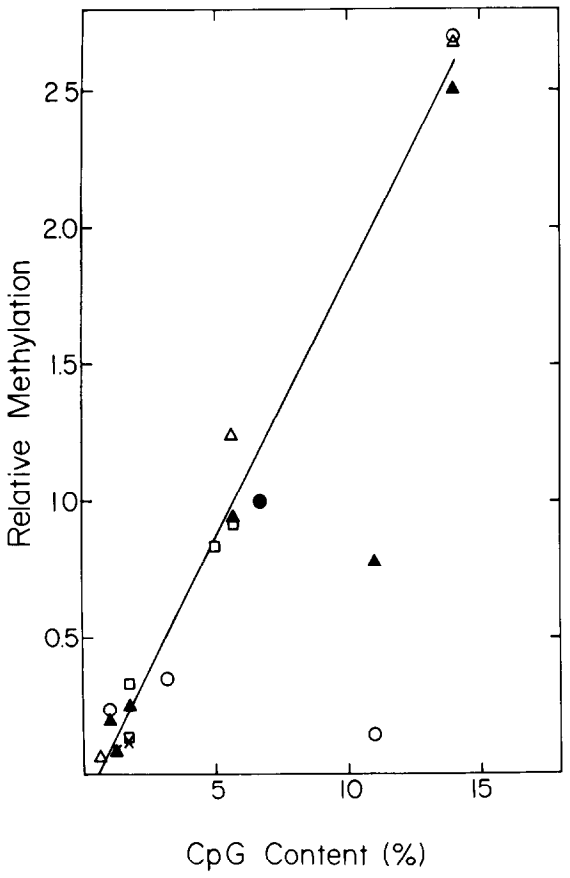

Fig.2. In vitro methylation as a function of CpG content of the DNA. This graph represents a compilation of published and our own data dealing with the in vitro methylation of DNA using eukaryotic DNA methylases. In each case methylase activity was determined for various different DNA substrates. Since, in each case, *E. coli* DNA was used as a template, this DNA is used as a point of reference (•) and the extent of methylation of other samples are expressed relative to the value obtained for *E. coli* DNA. In this way it is possible to compare data obtained in different laboratories. The following references were used as sources of data: (○) 8; (△) 9; (□) 3; (×) 10. The methylation of several DNA templates was also carried out in our laboratory using a partially purified DNA methylase from ascites tumor cells. The results of these experiments are also shown in this figure (▲) and include DNA from mouse liver, calf thymus, *D. melanogaster* larvae, *E. coli*, Herpes simplex virus I, and *M. luteus*. It should be noted that this experiment was done under conditions where the initial rate of reaction was measured. The CpG content of each DNA sample was determined from [23]: SV-40 (0.6%); rabbit (1.0%); calf (1.7%); mouse (1.2%); rat (1.2%); Vaccinia (3.2%); Aedes (5%); λ (5.7%); *E. coli* (6.7%); Herpes simplex virus (11%); *M. luteus* (14%).

4. Discussion

These data suggest that the recognition site for DNA methylation in eukaryotes is largely determined by the dinucleotide sequence CpG. It should be noted, however, that recent evidence suggests that CpG [11, 19], CpT and possibly CpA are also methylated in rare instances [12].

The enzymes *Hpa*II and *Msp*I have been found useful for analyzing the methylation state of the sequence CCGG, a small subset of the CpG sequences in the genome. Our results show a correlation between the degree of methylation at this specific site and the amount of methylation at CpG sequences in general. Thus, results obtained using these enzymes can probably be extended and generalized to all methylated cytosines. This is an important conclusion, since restriction enzymes which recognize methylated CpG sequences are extremely useful in analyzing specific genes in eukaryotic DNA.

In [13–15] satellite DNA sequences were highly methylated, suggesting that methylation may play a role in the formation of heterochromatin. The observation that the sequence CpG is in general highly methylated in mammalian cells opens the possibility that satellite sequences are highly methylated due to a high content of the dinucleotide CpG. In several instances this is indeed the case. Whereas 1% of the bases in total mouse DNA is CpG, mouse satellite contains 2.6% CpG [4]. Rat satellite DNA contains 10 CpG's in a repeating sequence of 370 basepairs. Sequence analysis has demonstrated that every CpG is indeed methylated. Several isolated calf satellites have been shown to be fully methylated at the specific sequences CCGG and GCGC [16]. That this is not the general case, however, is shown by the fact that there are other calf satellites which contain unmethylated or partially methylated CpG sequences [17,18].

Despite the high level of methylation of the dinucleotide CpG the fact remains that not all sites are methylated. In fact, restriction enzyme analysis of specific genes has shown that the level of methylation at certain sites varies from tissue to tissue. In general, gene expression seems to be correlated with a relative undermethylation at specific sites in these genes [19–22]. Using these techniques it may be possible to study this correlation not only at specific restriction sequences, but at all CpG sites.

Acknowledgement

This study was supported by the US Public Health Service, grant no. GM 20483.

References

[1] Sinsheimer, R. L. (1955) J. Biol. Chem. 215, 569–583.
[2] Pollack, Y., Stein, R., Razin, A. and Cedar, H. (1981) Proc. Natl. Acad. Sci. USA in press.
[3] Adams, R. L. P., McKay, E. L., Craig, L. M. and Burdon, R. H. (1979) Biochim. Biophys. Acta, 561, 345–357.
[4] Russell, G. J., Walker, P. M. B., Elton, R. A. and Subak-Sharpe, J. H. (1976) J. Mol. Biol. 108, 1–23.
[5] Swartz, M. N., Trautner, T. A. and Kornberg, A. (1962) J. Biol. Chem. 237, 1961–1967.
[6] Cedar, H., Solage, A., Glaser, G. and Razin, A. (1979) Nucleic Acids Res. 6, 2125–2132.
[7] Waalwijk, C. and Flavell, R. A. (1978) Nucleic Acids Res. 5, 4631–4641.
[8] Roy, P. H. and Weissbach, A. (1975) Nucleic Acids Res. 2, 1669–1684.
[9] Simon, D., Grunert, F., Acken, U. v., Doring, H. P. and Kroger, H. (1978) Nucleic Acids Res. 5, 2153–2167.
[10] Sneider, T. W., Teague, W. M. and Rogachevsky, L. M. (1975) Nucleic Acids Res. 2, 1685–1700.
[11] Sneider, T. W. (1980) Nucleic Acids Res. 8, 3829–3840.
[12] Burdon, R. H. and Adams, R. L. P. (1980) Trends Biochem. Sci. 5, 294–297.
[13] Harbers, K., Harbers, B. and Spencer, J. H. (1975) Biochem. Biophys. Res. Commun. 66, 738–746.
[14] Miller, O. J., Schnedl, W., Allen, J. and Erlanger, B. F. (1974) Nature 251, 636–637.
[15] Solage, A. and Cedar, H. (1978) Biochemistry 17, 2934–2938.
[16] Gautier, F., Bunemann, H. and Grotjahn, L. (1977) Eur. J. Biochem. 80, 175–183.
[17] Pech, M., Streeck, R. E. and Zachau. H. G. (1979) Cell 18, 883–893.
[18] Poschl, E. and Streeck, R. E. (1980) J. Mol. Biol. in press.
[19] van der Ploeg, L. H. T. and Flavell, R. A. (1980) Cell 19, 947–958.
[20] Mandel, J. L. and Chambon, P. (1979) Nucleic Acids Res. 7, 2081–2103.
[21] Sutter, D. and Doerfler, W. (1980) Proc. Natl. Acad. Sci. USA 77, 253–256.
[22] Desrosiers, R. C., Mulder, C. and Fleckenstein, B. (1979) Proc. Natl. Acad. Sci. USA 76, 3839–3843.
[23] Setlow, P. (1979) in: CRC Handbook of Biochemistry and Molecular Biology; Nucleic Acids (Fasman, G. D. ed) vol. 2, p. 32, CRC Press, Cleveland.