Proc. Natl. Acad. Sci. USA
Vol. 74, No. 7, pp. 2725–2728, July 1977
Biochemistry

Distribution of 5-methylcytosine in chromatin

(mass spectrometry/eukaryotic DNA methylation/chromatin structure)

AHARON RAZIN AND HOWARD CEDAR

Departments of Cellular Biochemistry and Molecular Biology, The Hebrew University-Hadassah Medical School, Jerusalem, Israel 91000

Communicated by Robert L. Sinsheimer, April 18, 1977

ABSTRACT The content of 5-methylcytosine in eukaryotic DNA was measured by mass spectrometry. Almost equal amounts of methylated cytosine were found in the DNA of various tissues of the chicken. When chromatin or nuclei were digested with micrococcal nuclease, 50% of the DNA was found to be nuclease resistant. In contrast to this, over 75% of the 5-methylcytosine was protected from nuclease digestion by chromatin proteins. These results suggest that 5-methylcytosine is nonrandomly distributed with respect to the nucleoproteins.

Although DNA methylation in bacteria has been fairly well studied, the biological significance of most of these methyl groups is still obscure. Only in some cases is the function of methylated bases known. Thus, methylation is known to be involved in restriction modification (1) and may have a role in bacterial DNA replication (2, 3). In the bacteriophage ϕX174, methylation of viral DNA plays a role in the final steps of virus maturation (4). A single 5-methylcytosine (m^5Cyt) residue in the phage DNA seems to serve as a recognition site for a specific endonuclease that is necessary to process the newly replicated DNA into viral DNA of one genome length (5).

In eukaryotes, only cytosine occurs in a methylated form and the biological role of this base has not been well-studied. One aspect which could shed light on the function of this modified base in eukaryotes concerns its localization with respect to particular DNA sequences and its distribution with respect to the chromatin proteins. Recently, investigations in several laboratories have indicated that chromatin is composed of a basic repeating structure, the nucleosome, which contains a complete set of histones and a DNA core of 200 base pairs (6–8). When chromatin is digested to the limit with micrococcal nuclease (spleen endonuclease, nucleate 3′-oligonucleotidohydrolase, EC 3.1.4.7), 50% of the total chromatin DNA remains strongly associated with protein and resistant to further digestion (9). To gain some insight into the distribution of m^5Cyt along the eukaryotic chromosome, we have studied the content of m^5Cyt in total DNA and in DNA sequences protected from nuclease digestion by chromatin proteins. This analysis has been done with standard techniques for isolating "open" and "covered" regions of chromatin DNA (9) and a newly developed, highly sensitive mass spectrometric method for the detection and quantitation of m^5Cyt in DNA (10).

MATERIALS AND METHODS

Materials. Poly-D-lysine hydrobromide was obtained from Sigma. It had an average molecular weight of 70,000. Purified micrococcal nuclease and RNase free of DNase were purchased from Worthington Biochemical Corp. Pronase B was from Calbiochem, proteinase K from Merck and Co., Inc., and trifluoroacetic acid (>99% purity) from Pierce Chemical Co.

Nuclease Digestion of Chromatin and Nuclei. Nuclei were obtained from calf thymus and chicken tissues by homogenizing and washing with sucrose-Triton buffers as described by Axel *et al.* (11). Chromatin was obtained from intact nuclei by a stepwise reduction in ionic strength followed by shearing and was digested with micrococcal nuclease in 1 mM Tris·HCl, pH 7.9/0.1 mM $CaCl_2$ as described (12). Digestion of nuclei was carried out on suspensions of nuclei washed twice in 1 mM Tris·HCl, pH 7.9/0.1 mM $CaCl_2$/0.25 M sucrose and resuspended in the same buffer at a DNA concentration of 1 mg/ml and a nuclease concentration of 10 μg/ml. The kinetics of DNA digestion by micrococcal nuclease was assayed by measuring the optical density at 260 nm of material soluble in 1 M NaCl/1 M $HClO_4$. Nuclease reactions were terminated by the addition of 5 mM NaEDTA, pH 7.0. The DNA resistant to nuclease treatment and total DNA were purified by treatment with proteinase K, phenol and chloroform extractions, and RNase as described by Axel *et al.* (12). "Open DNA" was obtained from chromatin as described (13). The nuclease-sensitive regions were protected by titration with poly-D-lysine and the covered regions were exposed by digestion of the chromatin proteins by Pronase (inactive towards poly-D-lysine). Open-region DNA was then obtained after micrococcal nuclease digestion of this complex followed by removal of the poly-D-lysine and purification of the DNA.

Determination of m^5Cyt. The content of m^5Cyt in various DNA samples was measured by using high-resolution mass spectrometry (10). Samples of DNA or deoxyribonucleoside monophosphates were dried over concentrated sulfuric acid and solid KOH under reduced pressure and then treated with trifluoroacetic acid in sealed ampules at 180° for 30 min. The trifluoroacetic acid was removed by vacuum evaporation and the sample was dissolved in H_2O. An equivalent of 1–20 μg of material was injected into the gold cup of the direct inlet and heated to 100° to evaporate the H_2O. After the temperature was brought to 20°, the sample was inserted into the direct inlet of the ion source. The amount of m^5Cyt was determined by using the molecular ion peak of thymine as an internal standard. The peak matching unit was set so that the reference thymine peak at m/e 126 was placed in one channel and the mass ratio was adjusted so that the m/e 125 peak fell into the second channel. The two channels were scanned alternately by using the peak matcher to flip between them at a slow speed, and the results of the scans were fed into a fast-response chart recorder. The temperature programmer was then used to raise the temperature automatically so that the bases could be distilled off. The peak heights were plotted as a function of temperature and the area under the curves was measured (Fig. 1). The m^5Cyt/Cyt molar ratio was calculated by using the base composition data obtained by native and melting spectral analysis (14). Our measurements were all performed on DNA that was freed of RNA by treatment with RNase. Treatment of these samples with NaOH to remove RNA yielded similar results, indicating that methylated RNA bases did not contribute to the m^5Cyt content detected in our experiments.

Abbreviation: m^5Cyt, 5-methylcytosine.

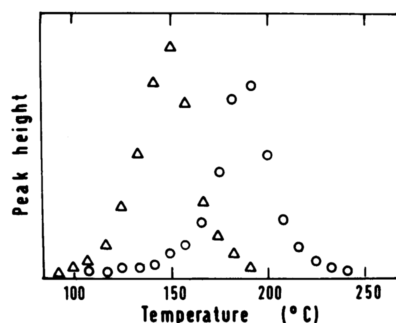

FIG. 1. Molecular ion yield of thymine and m5Cyt. Trifluoroacetic acid-treated purified calf thymus DNA was inserted as described in *Materials and Methods*. The sample was heated in the range of 80–250° for 500 sec by using the automatic temperature control. The peak heights of the molecular ions of m5Cyt and thymine were recorded as a function of temperature by fast consecutive selection of the corresponding ions. Thymine (△); m5Cyt (○). The scale used for m5Cyt is 10 times more sensitive than that used for the determination of thymine.

Table 1. Distribution of m5Cyt in chromatin (mol % of total Cyt)

Type of chromatin	Fraction of DNA			Covered: total
	Open	Covered	Total	
Native chicken erythrocyte	2.4 ± 0.2	4.5 ± 0.2	3.2 ± 0.1	1.4
Reconstituted chicken erythrocyte		3.3	3.3	1.0
Native calf thymus		5.4	4.1	1.3

Fractions of DNA were prepared and analyzed by mass spectrometry as described in *Materials and Methods*. Each result was obtained from the average of at least two measurements on each of two different DNA preparations. In the case of native erythrocyte chromatin, over 10 measurements were made on each fraction. The results are presented as the mean ± the standard error of the mean. The values presented were obtained from the molar ratio of m5Cyt/Thy after correcting for the base composition. For each fraction, the G+C base composition was approximately 42%.

RESULTS

Measurement of m5Cyt. The technique of mass spectrometry for the quantitative determination of m5Cyt in DNA is more sensitive and more accurate than other conventional methods. The data in Fig. 1 show the results of a typical analysis of the m5Cyt content of calf thymus DNA. As the temperature in the direct inlet was increased, both thymine and m5Cyt molecular ions distilled off. The m5Cyt molecular ion typically distilled at a higher temperature than thymine. In all experiments, the temperature was raised until no more of either molecular ion was detected. Because the temperature increase was linear, the total quantity of each compound was proportional to the area obtained by plotting the peak heights versus the temperature. After determining the ratio m5Cyt/Thy by this technique, the m5Cyt/Cyt ratio was calculated by considering the base composition of the DNA. By this technique, with 1–10 μg of DNA, as little as 1 m5Cyt residue per 10 kilobases can be detected (10).

Distribution of m5Cyt in Chromatin. By using micrococcal nuclease as a probe, we can divide chromatin DNA into two classes. "Covered DNA" is resistant to nuclease digestion presumably because it is protected by chromatin proteins. Open DNA is titratable with divalent cations or poly-D-lysine and is susceptible to micrococcal nuclease digestion (13). To determine the distribution of m5Cyt in chromatin, we prepared total, covered, and open DNA and subjected them to mass spectrometric analysis. As shown in Table 1, covered chicken erythrocyte DNA contained 4.5 mol of m5Cyt per 100 mol of Cyt. This value is 1.4-fold higher than that found in total DNA. Consistent with this higher content of m5Cyt found in covered DNA, open DNA contained less m5Cyt than did total DNA. The content of m5Cyt found in covered and open DNA added up to the value found for total DNA, assuming 50% digestion of chromatin DNA. As shown in Table 2, a higher value of m5Cyt could also be demonstrated in covered regions prepared from fresh nuclei treated *in situ* with micrococcal nuclease. These results indicated that this enrichment did not result from a rearrangement of chromatin proteins during the preparation of chromatin or the subsequent shearing procedure. We observed this enrichment in calf thymus chromatin (Table 1) as well as in nuclei of many chicken tissues (Table 2), which suggests that it is a general phenomenon in eukaryotic chromatin.

When the proteins of chromatin are dissociated from the DNA by treatment with high concentrations of salt and urea and then reassociated with the DNA by gradual dialysis of the mixture against low salt solutions (13), reconstituted chromatin is obtained. Several workers have conclusively demonstrated that this treatment results in a randomization of the major portion of chromatin proteins with respect to the DNA (13, 15), despite the evidence that reconstituted chromatin retains many of the structural and biological properties of native chromatin (16, 17). When covered DNA was prepared from reconstituted chromatin, no enrichment of m5Cyt was observed. This provides further evidence that reconstitution causes a randomization of chromatin proteins. The fact that covered regions of DNA prepared from reconstituted chromatin were not enriched with methyl groups provides an important control, suggesting that our results with native covered DNA are not due to artifacts of preparation of covered regions or to a specific resistance of certain DNA sequences to nuclease digestion.

When nuclei were prepared from various tissues of the chicken two basic observations were made (Table 2). First, the content of m5Cyt in DNA was similar, if not identical, in all tissues tested. This result is in good agreement with previous observations. Second, every tissue had the same distribution of m5Cyt between open and covered regions of chromatin DNA. Our results indicate that 70–85% of the methyl groups were localized in covered regions. To confirm that open-region DNA contains less m5Cyt than covered DNA, we isolated the free nucleotides after micrococcal nuclease digestion of erythrocyte nuclei. This fraction, which is equivalent to open-region DNA, contained approximately 2 mol of m5Cyt per 100 mol of Cyt,

Table 2. Distribution of m5Cyt in DNA from nuclei of various chicken tissues (mol % of total Cyt)

Tissue	Fraction of DNA		Covered: total
	Covered	Total	
Erythrocyte	4.5	2.8	1.6
Kidney	3.8	2.5	1.5
Spleen	3.6	2.8	1.3
Brain	4.0	2.5	1.6
Oviduct	4.0	2.4	1.7
Liver	4.2	2.7	1.6

Fractions of DNA were prepared and analyzed by mass spectrometry as described in *Materials and Methods*.

consistent with the molar ratio obtained for open erythrocyte DNA (see Table 1). Similar results have been obtained with L-cell nuclei, both by using the mass spectrometric technique and by analyzing the m^{5}Cyt content of covered region DNA isolated from cells labeled *in vivo* with [^{3}H]methionine (A. Solage and H. Cedar, unpublished data).

It should be noted that the values of m^{5}Cyt content for both chicken and calf DNA obtained in this study are considerably lower than those reported in the literature. Various workers have reported a m^{5}Cyt/Cyt molar ratio of 0.058–0.081 for calf thymus DNA (18) and 0.043 for chicken embryo DNA (19). In all of these reports, DNA was prepared without the use of either a protein-denaturing agent such as phenol or chloroform/isoamyl alcohol or ethanol to precipitate the DNA. In fact, most of these workers prepared DNA by salt precipitation. The m^{5}Cyt content of commercial DNA [obtained from Sigma and prepared by the method of Zamenhof (20)] when assayed by mass spectrometry was found to have a m^{5}Cyt/Cyt molar ratio of 0.065 (10), a value consistent with previously reported data. After ethanol precipitation, this DNA had a molar ratio of 0.041, indicating that it is this step in our purification procedure that is probably responsible for the lower values of m^{5}Cyt obtained with purified DNA. These results could imply that salt-purified DNA contains contaminants that interfere with the measurement of m^{5}Cyt. On the other hand, this may suggest that ethanol precipitation or extraction procedures using phenol or chloroform result in purified DNA lacking certain sequences rich in m^{5}Cyt content. One possibility is that our purification procedures, in fact, eliminate certain satellite sequences. Support for this has been reported by Rae *et al.* (21), who showed that certain specific satellite sequences of drosophila were lost during phenol extraction. In an attempt to clarify this point, we have tried to isolate DNA from ethanol supernatants. Thus far, we have not detected any DNA or m^{5}Cyt-containing compounds in the supernatants. These observarions warrant a careful reexamination of the m^{5}Cyt content of other DNA, especially the DNA of certain plant species which have been reported to be highly methylated (18).

DISCUSSION

Progress in understanding the function of m^{5}Cyt in eukaryotic DNA has been hindered by the lack of a suitable technique for measuring the m^{5}Cyt content of DNA. The use of radioactive precursors is limited to those tissues that can be labeled; the amount of incorporation is generally low, requiring several generations of growth to label all of the methyl groups. The use of chromatographic techniques coupled with UV spectroscopy has shortcomings in that large amounts of material are required for detecting the small amounts of m^{5}Cyt present in most DNA samples. Mass spectrometry is a technique that is both accurate and sensitive, being able to detect as little as 1 m^{5}Cyt residue per 10,000 nucleotide bases (10).

We have successfully applied this technique to the measurement of m^{5}Cyt in different fractions of chromatin DNA. By using previously developed methods (13), we have prepared DNA protected from micrococcal nuclease digestion (covered DNA) and DNA that is sensitive to nuclease treatment and titratable with poly-D-lysine (open DNA). About 75% of the methyl groups present in chromatin DNA were found to be clustered in the covered regions. That this was not an artifact of chromatin preparation was shown by the fact that the same results were obtained with treatment of nuclear chromatin *in situ* (Table 2).

These natural methyl groups could serve as a marker for the investigation of chromatin structure. Several workers have shown that, when chromatin is reconstituted, the majority of the chromatin proteins return to random locations on the DNA (13, 15). All of these studies have used artificial or synthetic markers to label regions of chromatin DNA. By using m^{5}Cyt as a natural marker, we have clearly shown that reconstitution causes a randomization of protein location in relation to these methyl groups. These results do not imply, however, that all chromatin protein sites are randomized; some specific protein–DNA interactions might still be preserved, especially since reconstituted chromatin has been shown to retain some biological function.

Although early observations on the distribution of chromatin proteins indicated that the proteins associated with nucleosomes were distributed randomly with respect to specific DNA sequences, recent evidence has suggested that this is not the case. In particular, Weintraub and Groudine have demonstrated that the globin genes in avian erythrocytes are selectively sensitive to digestion with pancreatic deoxyribonuclease I (22). These same genes are associated with a different protein conformation in other cells that do not express the globin genes. Garel and Axel (23) have found that this phenomenon also occurs with the avian ovalbumin gene. In our laboratory, we have shown that transcriptionally active viral sequences integrated into the DNA of mammalian cells are specifically sensitive to DNase I digestion. Nonactive viral genes are not preferentially digested (A. Panet and H. Cedar, unpublished data). Thus, it may be a general phenomenon that actively transcribed sequences are in a different protein conformation than those that are not. Our results with m^{5}Cyt also indicate that the protein conformation and localization on DNA is not random. The regions or sequences of DNA containing m^{5}Cyt seem to be in a form that renders them more resistant to micrococcal nuclease action.

Several workers have pointed out that most, if not all, m^{5}Cyt is located in satellite DNA sequences (24–26). Thus, the finding that m^{5}Cyt is nonrandomly distributed with respect to chromatin proteins may actually be due to a nonrandom distribution of satellite sequences. It has been demonstrated that satellite DNA is found mostly in the highly condensed heterochromatin (27). However, it should be pointed out that highly repetitious sequences in nuclei are not preferentially sensitive or resistant to micrococcal nuclease (28). In addition, it has been shown that nucleosomes are distributed randomly about the *Eco*RI sites of bovine satellite I DNA (29).

Mass spectrometry has also enabled us to analyze the quantity and distribution of m^{5}Cyt in DNA of various tissues of the chicken. Within the limits of our experimental accuracy, all of the tissues studied had the same amount of m^{5}Cyt and the same distribution of m^{5}Cyt within the chromatin fractions. These results are in agreement with the reported observations of Vanyushin (30, 31) that the amount of methylated cytosine does not vary considerably in tissues of many species that were studied. On the other hand, the degree of methylation does seem to vary during development. That satellite DNA sequences are methylated to different extents during each stage of development of *Drosophila melanogaster* has been demonstrated (A. Razin, S. Urieli, Z. Weinberg, I. Dawid and J. Sedat, unpublished data).

We are grateful to Dr. J. Deutsch and Mr. D. Linder for their invaluable assistance in operating the mass spectrometer, and to Prof. J. Mager for his thoughtful reading of the manuscript and useful suggestions. This work was supported by U.S. Public Health Service Grant GM20483 and U.S.-Israel Binational Research Foundation Grants 733 and 1140.

The costs of publication of this article were defrayed in part by the

payment of page charges from funds made available to support the research which is the subject of the article. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

1. Arber, W. & Linn, S. (1969) *Annu. Rev. Biochem.* **38**, 467–500.
2. Lark, C. (1968) *J. Mol. Biol.* **31**, 389–399.
3. Billen, D. (1968) *J. Mol. Biol.* **31**, 477–486.
4. Razin, A., Goren, D. & Friedman, J. (1975) *Nucleic Acids Res.* **2**, 1967–1974.
5. Friedman, J. & Razin, A. (1976) *Nucleic Acids Res.* **3**, 2665–2675.
6. Olins, A. & Olins, D. (1974) *Science* **181**, 330–332.
7. Noll, M. (1974) *Nature* **251**, 249–251.
8. Kornberg, R. D. (1974) *Science* **184**, 868–871.
9. Clark, R. J. & Felsenfeld, G. (1971) *Nature New Biol.* **229**, 101–106.
10. Deutsch, J., Razin, A. & Sedat, J. (1976) *Anal. Biochem.* **72**, 586–592.
11. Axel, R., Cedar, H. & Felsenfeld, G. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 2029–2032.
12. Axel, R., Melchior, W., Sollner-Webb, B. & Felsenfeld, G. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 4101–4105.
13. Bloch, S. & Cedar, H. (1976) *Nucleic Acids Res.* **3**, 1507–1519.
14. Hirschmann, S. Z. & Felsenfeld, G. (1966) *J. Mol. Biol.* **16**, 347–358.
15. Zimmerman, S. B. & Levin, C. J. (1975) *Biochemistry* **14**, 1671–1677.
16. Paul, J., Gilmour, R. S., Affra, N., Birnie, G., Harrison, P., Hell, A., Humphries, S., Windass, J. & Young, B. (1973) *Cold Spring Harbour Symp. Quant. Biol.* **38**, 885–890.
17. Barrett, T., Maryanka, D., Hamlyn, P. H. & Gould, H. J. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 5057–5061.
18. Hall, R. J. (1971) *The Modified Nucleosides in Nucleic Acids* (Columbia University Press, New York).
19. Smith, J. D. & Stoker, M. G. P. (1951) *Br. J. Exp. Pathol.* **32**, 433–441.
20. Zamenhof, S. (1957) in *Methods in Enzymology*, eds. Colowick, S. P. & Kaplan, N. O. (Academic Press, New York), Vol. 3, pp. 696–704.
21. Rae, P. M. M., Barnett, T. R. & Babbitt, D. G. (1976) *Biochim. Biophys. Acta* **432**, 154–160.
22. Weintraub, H. & Groudine, M. (1976) *Science* **193**, 848–856.
23. Garel, A. & Axel, R. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3966–3970.
24. Salomon, R., Kaye, A. M. & Herzberg, M. (1969) *J. Mol. Biol.* **43**, 581–592.
25. Miller, O. J., Schnedl, W., Allen, J. & Erlanger, B. F. (1974) *Nature* **251**, 636–637.
26. Harbers, K., Harbers, B. & Spencer, J. H. (1975) *Biochem. Biophys. Res. Commun.* **66**, 738–746.
27. Pardue, M. L. & Gall, J. G. (1970) *Science* **168**, 1356–1358.
28. Lacy, E. & Axel, R. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3978–3982.
29. Lipchitz, L. & Axel, R. (1976) *Cell* **9**, 355–364.
30. Vanyushin, B. F., Tkacheva, S. G. & Belozersky, A. N. (1970) *Nature* **225**, 948–949.
31. Vanyushin, B. F., Mazin, A. L., Vasilyev, V. K. & Belozersky, A. N. (1975) *Biochim. Biophys. Acta* **299**, 397–403.