Unmethylated domains in vertebrate DNA

David N.Cooper, Mary H.Taggart and Adrian P.Bird

MRC Mammalian Genome Unit, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK

Received 11 December 1982; Accepted 7 January 1983

ABSTRACT

We have detected a fraction that is rich in unmethylated HpaII and HhaI sites by end-labelling HpaII fragments of chicken DNA. The fraction is not obvious when DNA fragments are stained with ethidium bromide as it amounts to less than 2% of the genome. The average frequency of sites for HpaII is over thirteen times greater in the unmethylated fraction than in total DNA. Partial digests indicate that the unmethylated sites are clustered in the genome. Similar unmethylated fractions were detected in six other vertebrates in both somatic and germ line DNA.

INTRODUCTION

5-Methylcytosine is found predominantly in the sequence CpG in multicellular organisms, but not all CpGs are methylated (1-3). With the aid of restriction endonucleases, it is possible to map the relative distributions of a subset of methylated and unmethylated CpGs (4-6), and in many organisms a pattern is discernable in the total genomic DNA. Typical of these genomes is the sea urchin, where methylated CpGs are clustered together in long domains, and unmethylated CpGs are similarly clustered (7). Thus the sea urchin genome is divided into two sequence "compartments" due to the presence or absence of 5mCpG. Equivalent methylation compartments are evident in organisms across the phylogenetic spectrum, including many invertebrates (8), slime moulds (9), a fungus (Coprinus cinereus; P.Pukkila, personal communication) and a green plant (Brassica alba; C.G. Clark and A.P. Bird, unpublished observations). By contrast the DNA of vertebrates appears heavily methylated when analysed by the same techniques (e.g. reference 10), with no obvious fraction of unmethylated DNA. In this study we show

that the apparent lack of unmethylated DNA is due to the insensitivity of DNA staining. By using a simple end-labelling technique we demonstrate that there is indeed an unmethylated fraction in vertebrates, amounting to about 1% of the genome. Apart from its lack of methylation, the fraction differs significantly from the bulk of the genome by possessing a high frequency of HpaII sites. This may be partly a consequence of the absence of the mutable base, 5mC. Our findings suggest that the co-existence of heavily methylated and unmethylated sequences in the same genome is a general phenomenon.

MATERIALS AND METHODS

(a) DNA preparations. Liver, brain, kidney and blood were obtained from a single 3-day old chick of inbred line Hy-1 (Ross Breeders Ltd., Newbridge, Midlothian). DNA was prepared by standard SDS-phenol procedures. For the preparation of blood nuclei, blood cells were suspended in 10mls of 0.25M sucrose, 0.15M NaCl, 10mM EDTA, 1mM EGTA, 1.5mM spermine, 5mM spermidine, 0.15M Tris(pH7.4). An equal volume of the same buffer containing 0.5% NP-40 was added, and the resulting nuclei were washed and suspended in lysis buffer (0.15M NaCl, 0.1M Tris pH8.5, 0.1M EDTA). Nuclei were lysed with 0.5% SDS and the DNA was purified by standard procedures. Chicken sperm were collected from ten Hy-1 individuals (Ross Breeders) and lysed in 7M Urea, 1% SDS, 0.1M β -mercaptoethanol, 0.1M EDTA before phenol/chloroform extraction to remove tightly bound chromosomal proteins. Trout and snake DNAs were a gift from K. Jones and I. Purdom (Department of Genetics, Edinburgh University). Human sperm and placental DNAs were donated by H. Cooke (this laboratory). Preparation of Xenopus laevis sperm DNA has been described (26). Herring testis DNA was prepared in this laboratory by standard procedures. Bacteriophage lambda DNA was purchased from Boehringer.

(b) End-labelling. Between 1 and 10 g of each DNA were digested with HpaII (Boehringer, New England Biolabs or Amersham) in a 20 μ l reaction mix containing 10mM Tris(pH7.5), 10mM MgCl₂, 1mM dithiothreitol. After incubation at 37 reactions were either terminated by phenol and chloroform

extraction and the ethanol precipitated, or were used directly for end-labelling. Precipitated DNA was redissolved in 50mM Tris(pH7.5), 5mM MgCl₂ and incubated with 5μCi (α-P³²)dCTP and 4 units of DNA polymerase I (Klenow fragment, Cambridge Biotechnology Labs) at 15° for 10 min. HpaII incubation mixes were used directly for end-labelling by adding 10μCi(α-P³²) dCTP (400Ci/m mole, Amersham) and 9 units DNA polymerase I (Klenow fragment). Incubation was at 15° for 15 min. Reactions were terminated by adding 200μl 2.5M ammonium acetate (pH8), 20μg tRNA and 750μl ethanol. Samples were precipitated twice more with ethanol in the presence of sodium acetate and redissolved in 10mM Tris, 1mM EDTA (pH 8.5). They were then checked for the presence of unincorporated radioactive nucleotides by precipitation with trichloroacetic acid. HhaI (BRL) digests of labelled DNA were carried out according to the supplier's instructions. In one experiment ends were labelled with (γ-P³²) ATP and polynucleotide kinase according to Maxam and Gilbert (27). Hind III (Boehringer) digests of bacteriophage lambda DNA were labelled with (α-P³²) dCTP in the presence of cold dATP and dGTP.

(c) Gel Electrophoresis. Agarose gels of 1-1.2% were run in 'E' buffer. After staining and photographing, gels were dried down under vacuum, with heating, onto Whatman DE-81 paper and exposed to X-ray film. Traces of radioactive triphosphates were occasionally detected as a diffuse smear running ahead of the DNA. Control experiments showed that nucleotides could not be confused with small labelled DNA fragments. Low molecular weight fragments were resolved on a 5% acrylamide gel. Tris-borate-magnesium buffer (28) was used in order to stabilize short duplex molecules.

RESULTS

When chicken DNA that has been digested with HpaII is fractionated by agarose gel electrophoresis and stained with ethidium bromide, most DNA fragments remain large (figure 1, c and d). No unmethylated (i.e. extensively cleaved) fraction is evident. In contrast, digestion with MspI (which also recognises CCGG, but is insensitive to CpG methylation; 6) leads

Figure 1 End-labelling of HpaII fragments of chicken DNA. Lanes (a) to (d), agarose gel of end-labelled chicken kidney DNA stained with ethidium bromide. Lanes (e) to (h), autoradiograph of the same gel. Samples (a) and (e), undigested DNA labelled with (α - P^{32})dCTP; (b) and (f), undigested DNA labelled with (α - P^{32})TTP; (c) and (g), HpaII- digested DNA labelled with (α - P^{32})dCTP; (d) and (h), HpaII- digested DNA labelled with (α - P^{32})TTP. Lane (i), bacteriophage lambda DNA digested with Hind III and end- labelled with (α - P^{32})dCTP. Lane (j) to (o), autoradiograph of DNA from various chicken tissues after HpaII digestion and end- labelling with (α - P^{32})dCTP. Sample (j), blood cell nuclei; (k), whole blood; (l), sperm, (m), liver, (n), kidney; (o), brain. Gels were 1.2% agarose. Fragment lengths are given in kilobase pairs.

to extensive cleavage of chicken DNA, confirming that methylation is responsible for poor digestion by HpaII (not shown). Equivalent results are observed with a wide range of vertebrate DNAs. DNA staining would not, however, detect a heterogeneous fraction amounting to less than 5% of the DNA. We therefore sought to increase sensitivity by end-labelling HpaII fragments of chicken DNA with P^{32} before electrophoresis. Autoradiography of the gel now gives the distribution of fragments according to their number, rather than to their weight. Figure 1(g) shows a gel autoradiograph of HpaII-digested chicken DNA after labelling with P^{32} dCTP using the large fragment of DNA polymerase I. In the presence of labelled dCTP the polymerase extends the recessed 3' end generated by HpaII by 1 nucleotide. The autoradiograph differs significantly from the ethidium bromide staining pattern of the same lane

(figure 1,c). Over half of the label is in low molecular weight fragments of less than 500 bp in length, indicating that a fraction of chicken DNA is frequently cleaved by HpaII. Specificity for HpaII ends was demonstrated by negligible labelling of undigested DNA (figure 1, a and e), and by relatively weak labelling of digested DNA when P^{32} TTP was substituted for dCTP (figure 1d and h). Limited incorporation of TTP probably occurs at internal nicks and at damaged termini.

Several other control experiments verified that the low molecular weight material was not artefactual radioactivity in this region of the gel was dependent upon the presence of HpaII-digested chicken DNA and therefore was not due to DNA fragments introduced with the HpaII or polymerase enzymes. The radioactivity was insensitive to alkaline phosphatase, as would be expected if the labelled phosphate were internal to the DNA backbone (not shown). We also demonstrated an essentially identical pattern after end- labelling with polynucleotide kinase and $[\gamma-P^{32}]$ ATP (not shown). The presence of the low molecular weight fraction following two distinct labelling procedures argues strongly that the fraction is not spurious.

A comparison of end-labelled fragments from various chicken tissues showed no major differences in the size or proportion of the low molecular weight fraction (figure 1, j- o). Additional discrete bands were evident in liver and brain samples, and we suspect that these are derived from mitochondrial DNA. That the HpaII-cut fraction as a whole was not derived from mitochondrial DNA was shown by comparing the patterns of whole blood DNA and DNA from purified blood nuclei (figure 1, j and k). The patterns were indistinguishable, thereby ruling out a cytoplasmic origin for the unmethylated fraction. We also labelled the HindIII fragments of bacteriophage lambda DNA using the large fragment of DNA polymerase I to confirm that the method does not preferentially label small DNA fragments. Equivalent levels of radioactivity in fragments from 125bp to 23,000bp indicates that end-labelling generates an unbiased labelling pattern (figure 1,i).

Since the low molecular weight DNA was poorly resolved on agarose gels, we analysed two of the labelled samples on a 5%

acrylamide gel (figure 2, a and b). The small fragments gave a broad smear below 400bp with a mean of about 120bp. This shows that the HpaII-digested fraction is a complex mixture of DNA sequences, and that the average distance between HpaII sites in this fraction is considerably less than expected, given the rarity of CpG in vertebrate DNA (11,12). The expected frequency of HpaII sites calculated from nearest neighbour frequencies for chicken DNA (1599 bp) agrees with estimates of the number average molecular weight of MspI fragments derived from ethidium bromide-stained gels (1900 bp) (21). However, the frequency of the HpaII sites in the low molecular weight fraction is over thirteen times higher than this (120 bp). From the average length of low molecular weight fragments, and the fraction of total radioactivity which they contain, we estimate that 0.5-2% of chicken blood DNA belongs to the HpaII-cleaved fraction. No significant variation between tissues was detected.

The end-labelling experiment shows that chicken DNA contains frequent pairs of unmethylated HpaII sites that are usually less than 500 bp apart. Are these unmethylated sites set among heavily methylated sequences, or do they occur within domains of unmethylated DNA? We have answered this question by analysing end-labelled fragments in partial HpaII- digests of chicken kidney DNA. If the labelled HpaII sites were part of

Figure 2 Size distribution of end-labelled HpaII fragments of chicken DNA. Lane (a), whole blood DNA; lane (b), sperm DNA; lane (c), plasmid pAT153 cut with HpaII. DNA samples were end- labelled with DNA polymerase I and (α - P^{32})dCTP. The gel was 5% polyacrylamide run in Tris-borate- magnesium buffer. Fragment lengths are in base pairs.

unmethylated domains, we expected to see the low molecular weight fragments gradually increasing in length with decreasing extent of digestion. If, on the other hand, the labelled sites were surrounded by methylated HpaII sites, we expected a sharp jump between low and high molecular weight regions of the gel with decreasing HpaII digestion. The result of this experiment (figure 3, a-d) showed that the molecular weight of most labelled DNA increased gradually as the digestion time was reduced from 120 to 15 minutes. We conclude that unmethylated HpaII sites are clustered in the genome.

We have extended the experiment to demonstrate that sequence domains which contain unmethylated HpaII sites also contain unmethylated HhaI sites. Chicken kidney DNA that had been partially digested with HpaII for 15 minutes and end-labelled (figure 3, b) was digested to completion with HhaI and fractionated on an agarose gel (figure 3, e). The autoradiograph showed a fragment pattern that closely resembled the complete HpaII pattern (compare figure 3, d and e), with a prominent low molecular weight fraction. We conclude that the domains which contain unmethylated HpaII sites also contain unmethylated HhaI sites. The bulk of chicken DNA, by contrast, is poorly digested by HhaI due to CpG methylation within its recognition sequence (GCGC). Are similar unmethylated sequences present in other heavily methylated genomes? We analysed total

Figure 3 Clustering of unmethylated HpaII and HhaI sites in chicken DNA. Kidney DNA was digested with HpaII for (a) 4 min., (b) 15 min., (c) 45 min., and (d) 120 min., prior to end-labelling with (P^{32})dCTP. Approximately equal numbers of counts were loaded onto the agarose gel, although incorporation was significantly lower at early times of digestion. Lane (e) shows a complete HhaI digest of the 15 min. sample that is shown in lane (b). The size marker was lambda DNA digested with HindIII.

Figure 4 Unmethylated DNA in a wide range of vertebrates. Lane (a), bacteriophage lambda DNA digested with HindIII and end-labelled with (α - P^{32})dCTP. Lanes (b)-(j) show various DNA samples after digestion with HpaII and end-labelling with (α - P^{32})dCTP. Sample (b) human placenta; (c) human sperm; (d) mouse liver; (e) mouse sperm; (f) snake (*Elaphe radiata*); (g) *Xenopus laevis* blood; (h) *Xenopus laevis* sperm; (i) trout testis; (j) herring testis. Fragment sizes are in kilobase pairs. The gel was 1.2% agarose.

DNA from several other vertebrates using the same simple method (figure 4). The examples include fish, an amphibian, a reptile, a bird and two mammals. In each species there was a prominent fraction of DNA that was extensively cleaved by HpaII. As in the chicken, the unmethylated sequences were detectable in both sperm and somatic tissues.

DISCUSSION

The results described in this paper show that the heavily methylated genome of the chicken contains sequence domains in which both HpaII and HhaI sites are unmethylated. Several other vertebrates show similar fractions, which generally amount to less than two per cent of the genome. We have no direct evidence that all CpGs in this fraction are unmethylated, but this is likely since in other systems HpaII and HhaI sites appear typical of the surrounding CpGs with respect to methylation (9,13,14). Thus sequences lacking methylation at HpaII and HhaI sites probably also lack methylation at other CpGs.

In all the genomes tested, the unmethylated sequences

display an unexpectedly high frequency of HpaII sites. In contrast, the bulk DNA of vertebrates is deficient in the sequence CpG (11,12) and therefore exhibits a low frequency of HpaII/MspI sites. For total chicken DNA the observed frequency is one HpaII site per 1900bp, which is close to the value predicted by nearest neighbour analysis (21). HpaII sites in the unmethylated fraction are fifteen times more abundant than this, suggesting that CpG is not deficient in these sequences. This may be a direct consequence of the absence of 5mC, as there is evidence that the CpG deficiency is caused by the mutability of 5mC (16). Lack of a CpG deficiency is not, however, sufficient to account for the high frequency of HpaII sites, since it predicts an average of one site per 479bp, compared with the observed average of one site per 120bp. To account for the observed frequency, it is necessary to assume in addition that the average G + C content of the unmethylated sequences is 60%.

Co-existence of methylated and unmethylated sequences in the same genome has been demonstrated in a wide range of eukaryotes. The presence of unmethylated sequences in the otherwise heavily methylated genomes of vertebrates suggests that methylation compartments are a general feature of eukaryotic genomes (but note that no methylated sequences have yet been observed in *Drosophila*; ref.15). Previous experiments have provided some evidence for unmethylated DNA in vertebrates. Molitor *et al* (17) found that autoradiographs of DNA fibres that had been labelled *in vivo* with (H³)methionine showed discontinuous labelling, whereas fibres labelled with (H³)thymidine were continuously labelled. They suggested that this result was due to unmethylated sequences interspersed among methylated sequences, though it is possible that fluctuations in the frequency of CpG along the DNA fibre also contributed to the discontinuities. Other evidence for an unmethylated fraction has been derived from ethidium bromide-stained agarose gels (18). Densitometric tracings of HpaII digests of mouse showed a shoulder consistent with reduced methylation. The molecular weight of the shoulder (about 3kb) is twenty-five times greater than that of the HpaII fragments detected here. A concentration

of HpaII-cut fragments at 3kb was not apparent in our experiments.

Which sequences are unmethylated? In particular, how do transcribed sequences relate to the unmethylated DNA. Experiments by Naveh-Manny and Cedar (21) have suggested that transcribed sequences in the chicken are undermethylated compared to total DNA. Analysis of individual chicken genes, however, shows a rather complex picture, with methylation patterns that vary for different genes. Most ribosomal RNA genes, for example, are unmethylated at HpaII sites in a variety of tissues, including sperm (D. Cooper and L. Errington, manuscript in preparation). The high frequency of sites for both enzymes in rDNA means that these genes belong to the unmethylated fraction identified here. Unmethylated rDNA can account for about 2% of the unmethylated fraction in chicken. A somewhat different pattern of methylation has been identified at the chicken $\alpha(2)$ I collagen gene locus (22). Sequences surrounding the 5' end of this gene are not detectably methylated, but the coding regions are heavily methylated. This pattern is observed in all the tissues tested, including those, such as sperm, that do not express the collagen gene. Since the unmethylated 5' domain is abnormally rich in HpaII sites, it must contribute to the fraction identified in our study.

Genes exhibiting analogous methylation patterns have been observed in other species (H. Cedar, personal communication). If a high proportion of genes are of this type, then it is possible that a significant fraction of unmethylated sequences are derived from 5' domains. Genes of another kind probably contribute very few sequences to the unmethylated fraction. These are genes that are heavily methylated in tissues where they are not expressed, but alter their methylation patterns in expressing tissues by loss of methylation within or near the coding region. Examples in the chicken include the ovalbumin (23), conalbumin (23) and α (24) and β (23,25) globin genes. Although these genes are undermethylated in expressing tissues, the frequency of HpaII sites is not generally high. Also, since they are heavily methylated in sperm, genes of this type would

not contribute to the prominent unmethylated fraction in sperm DNA.

In the sea urchin, at least 85% of unmethylated sequences, including several genes, remain unmethylated at all stages of the life cycle (9; D. N. Cooper, unpublished observations). Does the unmethylation identified here fraction comprise the same sequences in different cell types? A definitive answer must await experiments with cloned unmethylated DNA, but there is some evidence for conservation of the unmethylated sequences between tissues. Two chicken sequences, rDNA (D. Cooper and L. Errington, manuscript in preparation) and a domain near the $\alpha 2(1)$ collagen gene (22), are known to be unmethylated in several chicken tissues, including sperm and would give rise to HpaII fragments of low average size. Unmethylated rDNA is also observed in a range of mouse (19) and rat (20) tissues. It remains to be seen whether the sequence composition of the unmethylated fraction is mainly stable or unstable during development.

ACKNOWLEDGEMENTS

We are grateful to Ed Southern for comments on the manuscript and to R. C. Roberts for advice on statistical analysis of the data. D.N.C. was supported by an MRC postgraduate studentship.

REFERENCES

1. Doscocil J, Sorm F (1962) *Biochim. Biophys. Acta* **55**, 953-959.
2. Grippo P, Iaccarino M, Parisi E, Scarano E (1968) *J. Mol. Biol.* **36**, 195-208.
3. Cedar H, Solage A, Glaser G, Razin A (1979) *Nucl. Acids Res.* **6**, 2125-2132.
4. Bird AP, Southern EM (1978) *J. Mol. Biol.* **118**, 27-47.
5. Gautier F, Bunemann H, Grotjahn L (1977) *Eur. J. Biochem.* **80**, 175-183.
6. Waalwijk C, Flavell RA (1978) *Nucl. Acids Res.* **5**:3231-3236.
7. Bird AP, Taggart MH, Smith BA (1979) *Cell* **17**, 889-901.
8. Bird AP, Taggart MH (1980) *Nucl. Acids Res.* **8**, 1485-1497.
9. Whittaker PA, McLachlan A, Hardman N (1981) *Nucl. Acids Res.* **9**, 801-814.
10. Singer J, Roberts-Ems J, Riggs AD (1979) *Science* **203**, 1019-1021.

11. Swartz MN, Trautner TA, Kornberg A (1962) J. Biol. Chem. 237, 1961-1967.
12. Josse J, Kaiser AD, Kornberg A (1961) J. Biol. Chem. 236, 864- 875.
13. Fedoroff NV, Brown DD (1978) Cell 13, 701-716.
14. Miller JR, Cartwright EM, Brownlee GG, Fedoroff NV, Brown DD (1978) Cell 13, 717-725.
15. Urieli-Shoval S, Gruenbaum Y, Sedat J, Razin A (1982) FEBS Lett. 146, 148-152.
16. Bird AP (1980) Nucl. Acids Res. 8, 1499-1503.
17. Molitor H, Drahovsky D and Wacker A (1976) Biochim. Biophys. Acta 432, 28-36.
18. Naveh-Many T and Cedar, H. (1982) Molec. Cell Biol. 2, 758- 762.
19. Bird AP, Taggart MH, Gehring C (1981) J.Mol.Biol. 152, 1-17. 20. Kunnath L, Locker J (1982) Nucl. Acids Res. 10, 3877- 3892.
21. Naveh-Many T, Cedar H (1981) Proc. Natl. Acad. Sci. USA 78, 4246-4250.
22. McKeon C, Ohkubo H, Pastan I, de Crombrughe B (1982) Cell 29, 203-210.
23. Mandel JL, Chambon P (1979) Nucl. Acids Res. 7, 2081-2103.
24. Weintraub H, Larsen A, Groudine M (1981) Cell 24, 333-344.
25. McGhee JD, Ginder GD (1979) Nature 280, 419-420.I
26. Macleod D and Bird A (1982) Cell 29, 211- 218.
27. Maxam AM, Gilbert W (1980) Meth. in Enzymol. 65, 499-560.Biochemistry 14, 3787- 3794.
28. Maniatis T, Jeffrey A, van de Sande H (1975) Biochemistry 14, 3787-3794