Volume 6 Number 9 1979 Nucleic Acids Research

Absence of cytosine methylation at C-C-G-G and G-C-G-C sites in the rDNA coding regions and intervening sequences of Drosophila and the rDNA of other higher insects

Peter M.M.Rae and Robert E.Steele

Department of Biology, Yale University, New Haven, CT 06520, USA

Received 9 May 1979

ABSTRACT

Cytosine residues in C-G dinucleotides are frequently methylated in eukaryote DNA. In DNA of the dinoflagellate C. cohnii, the sequence C-MeC-G-G apparently renders Hpa II (C-C-G-G) incapable of digesting whole cell DNA in general, and rDNA in particular. Msp I, which also recognizes C-C-G-G but cleaves irrespective of methylation, degrades C. cohnii DNA and produces rDNA segments of 10.2 to 1.4 kb. We have applied this Hpa II/Msp I test to unfractionated DNA, and to rDNA and the rDNA intervening sequence of Drosophila virilis embryos and adults. There is no evidence of C-MeC-G-G sequences in either developmental stage of this species. Absence of G-MeC-G-C from coding and intervening sequences of rDNA was shown in comparisons of Hha I (G-C-G-C) cleavage patterns of unfractionated DNA and cloned (unmodified) segments of rDNA. Comparisons of Hpa II and Msp I cleavage products of DNA from the house fly, the flesh fly and a bumblebee also revealed no internal cytosine methylation in the sequence C-C-G-G. Because amounts of MeC in C-G dinucleotides vary greatly among species, from apparent nonexistence to substantial proportions, no inference may yet be drawn about the role of such base modifications in DNA.

INTRODUCTION

Methylcytosine (MeC) is the modified base commonly found in the DNA of eukaryotes, and it usually comprises about 1 to 5% of all bases in the DNA of an organism (Shapiro, 1976). Methylcytosine is distributed nonrandomly, being found at greatest frequency in the dinucleotide MeC-G (Sinsheimer, 1954; Grippo et al., 1968). In the amphibian Xenopus laevis, extrachromosomal ribosomal DNA (rDNA) amplified in pachytene and diplotene oocytes is distinguishable from rDNA in somatic cells by the fact that somatic rDNA (67.2% G+C) contains 4.5% methylcytosine while extrachromosomal rDNA is devoid of this minor base (Dawid et al., 1970); bulk DNA in somatic cells (39.6% G+C) contains 1.3% MeC. Bird and Southern (1978) investigated the distribution of methylcytosine in Xenopus somatic cell rDNA by treating DNA with restriction endonucleases that recognize sites which include the sequence C-G but are incapable of cleaving if the C is methylated. By comparing the patterns of hybridization of rRNA and cloned segments of rDNA to somatic and oocyte DNA

cleaved with such enzymes, Bird and Southern found that nearly all internal C residues in the sequences C-C-G-G, C-Py-C-G-Pu-G, G-C-G-C, and Pu-G-C-G-C-Py were methylated in somatic cell rDNA.

Taking a similar approach, Waalwijk and Flavell (1978) investigated the large intervening sequence in the rabbit β-globin gene for modification at a C-C-G-G site. Using Hap II, which cleaves C-C-G-G but not C-MeC-G-G, and Msp I, which cleaves at this sequence irrespective of cytosine methylation, they found that the C-C-G-G in the intervening sequence contained methylcytosine in about half of the globin genes in most somatic tissues (erythroid and nonerythroid), but that the site was 100% methylated in sperm DNA and about 80% methylated in brain DNA.

Drosophila DNA has been variously reported as containing no detectable methylcytosine (<5% of all C; Argyrakis and Bessman, 1963) to about 6% methylcytosine (30% of all C; Mead and Fox, 1961). Drosophila rDNA has been studied in some detail, and intervening sequences interrupt 28S rRNA coding regions in most repeat units in D. melanogaster (Glover and Hogness, 1977; White and Hogness, 1977; Wellauer and Dawid, 1977; Pellegrini et al., 1977) and D. virilis (Barnett and Rae, 1979). In the latter species, a 9.6 kb rDNA intervening sequence is present in about three-fourths of the rDNA repeat units. We undertook a study of Drosophila rDNA that parallels those of Bird and Southern and Waalwijk and Flavell in order to determine the generality of the observations they reported on the modification of cytosine in rDNA and in an intervening sequence. Drosophila rDNA is particularly appropriate for such a study because methylation can be monitored simultaneously in coding regions, spacer, and the intervening sequence. Beyond this fly, we examined the rDNA of two distantly related diptera, the house fly and the flesh fly, and of a hymenopteran. In all cases we found no evidence for cytosine methylation at C-G sites in genes or intervening sequences, and conclude that methylcytosine is not an obligatory structural and/or functional element of these components of the genome.

MATERIALS AND METHODS

The preparation of Drosophila virilis, Sarcophaga bullata and Musca domestica DNAs is described elsewhere (Barnett and Rae, 1979). Bumblebee (Psithyrus sp.) DNA was purified from individuals collected in the wild. Dinoflagellate DNA was prepared essentially as described (Rae, 1973). Gel electrophoresis, transfer of DNA to nitrocellulose filters, and the preparation and utilization of ^{32}P-labelled RNA and DNA probes were as described

(Barnett and Rae, 1979). Transfer to diazobenzyloxymethyl paper (Enzo Biochem, New York) was according to Alwine et al. (1979). Bam HI, Hha I and Msp I enzymes were from New England Biolabs; Hpa II was from Bethesda Research Laboratories. Bam HI and Hha I digestions were in 60 mM NaCl, 6.6 mM Tris-HCl, pH 7.6, 6.6 mM $MgCl_2$, 6.6 mM mercaptoethanol. Hpa II and Msp I digestions were in 6 mM NaCl, 6 mM Tris-HCl, pH 7.5, 6 mM $MgCl_2$, 6 mM mercaptoethanol, 0.1% bovine serum albumin. Reactions were at 37° for 1.5 to 2 hr with excess enzyme.

RESULTS

The effectiveness of the Hpa II/Msp I test for cytosine methylation in the sequence C-C-G-G is readily demonstrable with DNA from an organism with even a comparatively low methylcytosine content. The control for our experiments with insect DNA has been DNA from the dinoflagellate Crypthecodinium cohnii. DNA of this alga is 0.75% methylcytosine (3.5% of all C; Rae, 1973), and nearest neighbor analyses have shown that the dinucleotide MeC-G comprises 80% of all MeC-N and 10-15% of all C-G (Steele, unpublished). Figure 1 (left) shows that little, if any, C. cohnii DNA is digested by Hpa II, while Msp I cleaves the DNA very frequently. With particular regard to rDNA, Figure 1 (right) shows that Hpa II treatment does not produce segments that migrate in a 1% agarose gel farther than the position of limiting mobility (>22 kb), but that Msp I digestion produces several segments in the size range 10.2 to 1.4 kb.

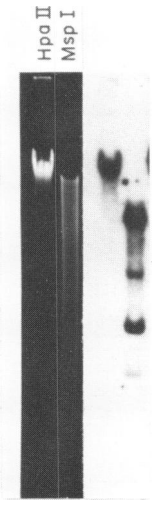

Fig. 1. DNA of the dinoflagellate Crypthecodinium cohnii treated with Hpa II or Msp I and electrophoresed in 1% agarose. Ethidium bromide staining patterns are on the left. On the right are corresponding patterns of the hybridization of ^{32}P end labelled 16S + 24S ribosomal RNA of C. cohnii to DNA transferred from the gel to a nitrocellulose filter. Controls in which phage λ DNA was included in a reaction mixture have shown that the inability of Hpa II to cleave C. cohnii DNA is not due to inhibitors in the DNA solution.

In contrast, Drosophila virilis DNA from embryos or adults is susceptible to Hpa II digestion to an extent that is indistinguishable by gel electrophoresis from that of Msp I digestion (Fig. 2, upper left). The difference between embryos and adults in the amount of stained material at the tops of the respective lanes is due to differences in the proportion of satellite DNA (which lacks the sequence C-C-G-G) between diploid and polytene nuclei (Gall

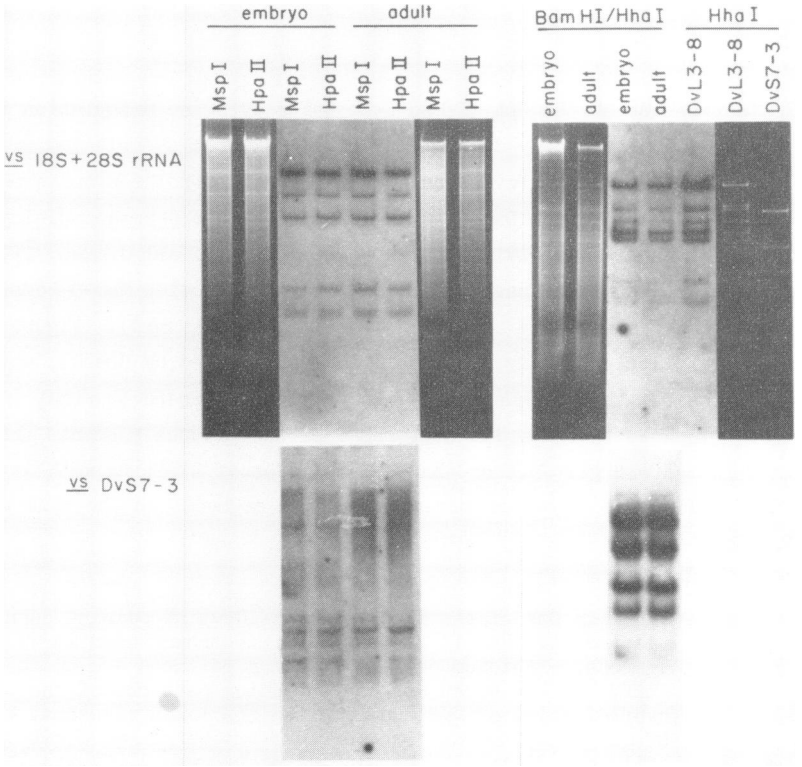

Fig. 2. DNA of Drosophila virilis embryos and adults treated with Msp I or Hpa II (left) or with Hha I (right), and electrophoresed in 1.5% agarose. After ethidium bromide staining, DNA in duplicate gels was transferred to diazotized benzyloxymethyl paper for hybridization with end labelled 18S + 28S Drosophila ribosomal RNA (upper autoradiograms) or a cloned 4.9 kb segment of the D. virilis rDNA intervening sequence (lower autoradiograms; see Fig. 3). In panels on the right, DNA from embryos and adults was cleaved with Bam HI before Hha I treatment for comparison with Hha I digests of cloned segments of D. virilis rDNA; the segments DvL3-8 (16 kb) and DvS7-3 (4.9 kb) are terminated by Bam HI sites, and were preparatively separated from the pBR313 vector for this experiment. Characterization of the plasmids pDvL3-8 and pDvS7-3 is in Barnett and Rae (1979), and maps are shown in Fig. 3.

and Atherton, 1974). By the criterion of size distribution of restriction segments, cytosine methylations in the sequence C-C-G-G in Drosophila DNA are few or nonexistent, and there are no detectable differences in this regard between embryos and adults.

Particular attention was given the ribosomal RNA genes and the rDNA intervening sequence because it is important to determine the generality of phenomena reported by Bird and Southern (1978) for Xenopus rDNA and by Waalwijk and Flavell (1978) for the major intervening sequence in the rabbit β-globin gene (INTRODUCTION). Figure 2 (upper left) shows the pattern of hybridization of 18S + 28S rRNA to DNA transferred from the stained gel shown to diazobenzyloxymethyl (DBM) paper. The distribution of Hpa II segments that contain coding sequences is the same as that of Msp I segments in DNA from both embryos and adults, so that these genes in Drosophila are not subject to cytosine methylation at C-C-G-G sites as they are in Xenopus.

In Drosophila virilis, about three-fourths of the repeating units of rDNA contain a 9.6 kb intervening sequence in the 28S rRNA coding region (Barnett and Rae, 1979). The major, 21 kb, repeat of D. virilis rDNA is diagrammed in Figure 3, in which is shown the map of two cloned Bam HI segments of rDNA that together comprise the 21 kb unit. The plasmid pDvS7-3 contains a 4.9 kb portion of rDNA derived from about the middle of the intervening sequence; the segment has sequence homology with the remainder of the intervening sequence represented in pDvL3-8, which contains the remaining 16 kb of the 21 kb repeat unit (Barnett and Rae, 1979). We mapped the Hpa II sites in DvL3-8 and DvS7-3 (Fig. 3), and used the 4.9 kb segment of pDvS7-3 as a probe

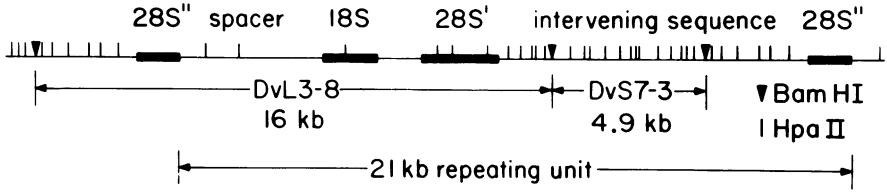

Fig. 3. A restriction map of Drosophila virilis rDNA derived from analysis of the rDNA clones pDvL3-8 and pDvS7-3. Shown is a 21 kb repeat unit in which the 28S coding region contains a 9.6 kb intervening sequence; this class comprises about three-fourths of all rDNA in this species. The rest of rDNA in D. virilis consists of units about 11 kb in length, some or all of which may contain smaller intervening sequences (Barnett and Rae, 1979). The cleavage sites for Bam HI and Hpa II are shown. Documentation of the Bam HI sites is in Barnett and Rae (1979); the Hpa II (Msp I) sites were mapped following the procedure of Smith and Birnstiel (1976).

in hybridizations to blots of Hpa II and Msp I digested genomic DNA from D. virilis embryos and adults. Figure 2 (lower left) shows the pattern of hybridization of DvS7-3 to DNA transferred from a gel similar to that shown above. As is so for rRNA hybridization patterns, Msp I and Hpa II give identical DvS7-3 hybridization patterns with DNA from both embryos and adults, and all bands can be accounted for in the Hpa II map derived from cloned (unmodified) segments of rDNA. Evidently, the internal cytosine of C-C-G-G is not methylated in the rDNA intervening sequences of either embryos or adults of D. virilis; these results contrast with those of studies on the major intervening sequence of the rabbit ß-globin gene, where the Hpa II site contains methylcytosine in more than 50% of the gene equivalents (Waalwijk and Flavell, 1978).

Hha I is another restriction endonuclease which is limited in activity by cytosine methylation (Bird and Southern, 1978). This enzyme recognizes the sequence G-C-G-C, but will not cleave if the internal cytosine is methylated. We have compared the Hha I cleavage products of genomic rDNA from D. virilis embryos and adults with one another and with the products of Hha I cleavage of cloned segments of rDNA. Because the cloned segments are terminated by Bam HI sites, the comparisons were performed using genomic DNA double digested with Bam HI and Hha I. In Figure 2 (upper right) are the ethidium bromide staining profiles of DNA from embryos and adults, cleaved with Bam HI and Hha I; also shown are the staining patterns of DvL8-7 and DvS7-3, purified from the respective plasmids and cleaved with Hha I. The autoradiogram illustrates hybridization of 18S + 28S rRNA to DNA transferred from the gel to DBM paper. There are no differences among DvL3-8, embryonic DNA and adult DNA with regard to rRNA hybridization, and because DvL3-8 is derived from an unmodified E. coli plasmid, we infer that there are no (or undetectably few) Hha I resistant G-C-G-C sequences in and around the rRNA coding regions of Drosophila virilis rDNA. Again, this result contrasts with that obtained by Bird and Southern with Xenopus rDNA.

The rDNA intervening sequence of genomic DNA from adults and embryos is also free of modifications that confer resistance to Hha I digestion. Figure 2 (lower right) shows DvS7-3 hybridization to embryo and adult DNA transferred to a nitrocellulose sheet from a gel similar to that shown above. There are no differences between the two lanes, and all bands are accounted for in the plasmid DNA lanes above.

This paucity of methylcytosine in the sequences C-C-G-G and G-C-G-C in Drosophila virilis rDNA coding and intervening sequence regions, as compared

with Xenopus rDNA and the intervening sequence of the rabbit β-globin gene, is not limited to this species of fly. We applied the Hpa II/Msp I test to DNA of the house fly (Musca domestica) and of the flesh fly (Sarcophaga bullata), and to DNA of a hymenopteran, the parasitic bumblebee Psithyrus sp., all of which have rDNA intervening sequences (Barnett and Rae, 1979; Rae, in preparation). Figure 4 shows that by the criterion of ethidium bromide staining the general cleavage patterns of each of these species' DNA by Msp I and Hpa II are indistinguishable. With particular regard to rDNA, the autoradiograms show hybridization of 18S + 28S rRNA to identical bands in Msp I and Hpa II digests of each DNA, although there are major differences among the species in the distribution of Hpa II/Msp I sites.

DISCUSSION

While cytosine modification in nuclear DNA is widespread among eukaryotes, amounts of methylcytosine vary greatly and some organisms are apparently devoid of this minor base (Shapiro, 1976; Rae and Steele, 1978). The role

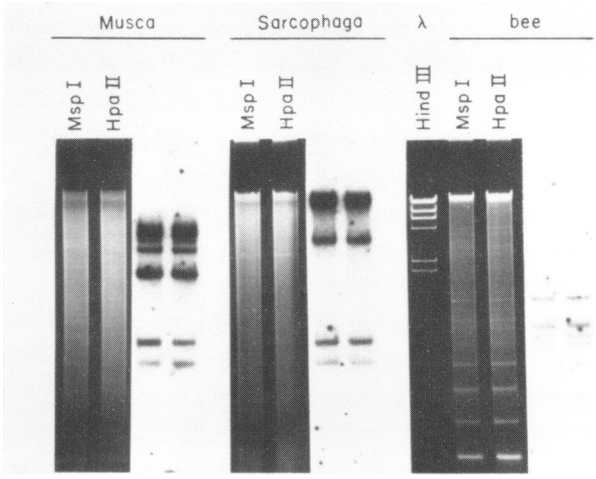

Fig. 4. DNA of Musca domestica, Sarcophaga bullata and Psithyrus sp. (bee) digested with Msp I or Hpa II and electrophoresed in 1.5% agarose. Shown are the ethidium bromide staining patterns, and the hybridization of Drosophila 18S + 28S rRNA to DNA transferred from the gel to a nitrocellulose filter. The prominent low molecular weight stained bands in the bee DNA lanes are probably derived from a G + C-rich satellite component of Psithyrus DNA (Rae, in preparation), and they are multiples of 140-160 base pairs. The λ lane displays a Hind III digest of λ C_I857Sam7 DNA; the sizes of the segments are 21.9, 8.76, 6.07, 2.10, 1.81 and 0.51 kb (Murray and Murray, 1975).

of methylcytosine in eukaryotes is presently unknown. Modification in specific genes has been investigated in only a few (although diverse) organisms, and with regard to only two genes. The repetitive genes for ribosomal RNA are highly methylated at C-C-G-G and G-C-G-C sites in chromosomal DNA of the amphibian Xenopus laevis (Bird and Southern, 1978) and in the dinoflagellate Crypthecodinium cohnii, but they are not detectably modified at such sites in the rDNA of Drosophila, other diptera, and a hymenopteran (nor are they in the amplified rDNA of Xenopus oocytes). Similarly, the extents of cytosine methylation in the sequence C-C-G-G in the intervening sequence of the β-globin gene vary between 50% and 100% among tissues of the rabbit (Waalwijk and Flavell, 1978), but the rDNA intervening sequence of Drosophila lacks internal MeC in C-C-G-G and G-C-G-C sequences. Indeed, our application of the Hpa II/Msp I test to unfractionated DNA of Drosophila and other insects has indicated, at the level of resolution afforded by ethidium bromide staining of agarose gels, that there is little, if any, cytosine methylation in C-C-G-G sequences throughout the various genomes. The conclusion that may be drawn from these data is that methylcytosine is not required for genome structure or function in eukaryotes.

Our studies of Drosophila DNA have indicated that methylcytosine is not present in the sequences C-C-G-G and G-C-G-C. In other eukaryotes, MeC occurs largely or sometimes entirely in the dinucleotide MeC-G (Sinsheimer, 1954; Grippo et al., 1968; Steele, unpublished), and from this and our data we infer that this minor base is absent from Drosophila DNA. It is a possibility that methylcytosine in this organism occurs exclusively in other N-C-G-N or N-C-N-N tetranucleotides we have not probed, but there is as yet no evidence from any source to support this.

It is an intriguing observation that extents of cytosine methylation in the sequence C-C-G-G differ among tissues in the intervening sequence of the rabbit β-globin gene (Waalwijk and Flavell, 1978). However, as other specific components of the rabbit genome have yet to be examined, it may be that such differences are not limited to the β-globin intervening sequence. In any event, speculation regarding the role of methylcytosine in coding and intervening sequences in eukaryote genomes must take into account the apparent absence or paucity of this minor base in the DNAs of higher insects and some other organisms.

ACKNOWLEDGEMENTS

This work was supported by NSF and NIH grants to PMMR, and by a USPHS Cell and Developmental Biology Training Grant.

REFERENCES

1 Alwine, J.C., Kemp, D.J., Parker, B.A., Reiser, J., Renart, J., Stark, G.R. and Wahl, G.M. (1979) Methods in Enzymology, in press.
2 Argyrakis, M.P. and Bessman, M.J. (1963) Biochim. Biophys. Acta 72, 122-124.
3 Barnett, T. and Rae, P.M.M. (1979) Cell 16, 763-775.
4 Bird, A.P. and Southern, E.M. (1978) J. Mol. Biol. 118, 27-47.
5 Dawid, I.B., Brown, D.D. and Reeder, R.H. (1970) J. Mol. Biol. 51, 341-360.
6 Gall, J.G. and Atherton, D.D. (1974) J. Mol. Biol. 85, 633-664.
7 Glover, D.M. and Hogness, D.S. (1977) Cell 10, 167-176.
8 Grippo, P., Iaccarino, M., Parisi, E. and Scarano, E. (1968) J. Mol. Biol. 36, 195-208.
9 Mead, C.G. and Fox, A.S. (1961) Federation Proc. 20, 353.
10 Murray, K. and Murray, N. (1975) J. Mol. Biol. 98, 551-564.
11 Pellegrini, M., Manning, J. and Davidson, N. (1977) Cell 10, 213-224.
12 Rae, P.M.M. (1973) Proc. Nat. Acad. Sci. USA 70, 1141-1145.
13 Rae, P.M.M. and Steele, R.E. (1978) Biosystems 10, 37-53.
14 Shapiro, H.S. (1976) in Handbook of Biochemistry and Molecular Biology, Nucleic Acids, Fasman, G.V., Ed., Vol. II, pp. 241-283, C.R.C. Press, Cleveland.
15 Sinsheimer, R.L. (1954) J. Biol. Chem. 208, 445-459.
16 Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. 3, 2387-2398.
17 Waalwijk, C. and Flavell, R.A. (1978) Nucleic Acids Res. 5, 4631-4641.
18 Wellauer, P.K. and Dawid, I.B. (1977) Cell 10, 193-212.
19 White, R.L. and Hogness, D.S. (1977) Cell 10, 177-192.