The 5-methylcytosine content of highly repeated sequences in human DNA

Miguel A.Gama-Sosa*, Richard Y.-H.Wang*, Kenneth C.Kuo+, Charles W.Gehrke+ and Melanie Ehrlich*

*Department of Biochemistry, Tulane Medical School, New Orleans, LA 70112, and +Department of Biochemistry, Experiment Station Laboratories, University of Missouri, Columbia, MO 65201, USA

Received 8 March 1983; Revised and Accepted 18 April 1983

ABSTRACT

Previously, we found much tissue- or cell-specificity in the levels of 5-methylcytosine (m^5C) in the total human genome as well as in DNA fractions resolved by reassociation kinetics. We now report that there were even greater differences in the m^5C content of the highly repeated, tandem EcoRI family of DNA sequences from different human organs or cell populations. The ratio of m^5C levels in this DNA fraction from brain, placenta, and sperm was 2.0:1.2:1.0. At a HhaI site in this repeat family, sperm DNA was 5-10 fold less methylated than somatic DNAs. In contrast, the highly repeated Alu family, which is ~5% of the genome, had almost the same high m^5C content in brain and placenta despite marked tissue-specific differences in m^5C levels of the single copy sequences with which these repeats are interspersed. These data show that very different degrees of change in methylation levels of various highly repeated DNA sequences accompany differentiation.

INTRODUCTION

5-Methylcytosine (m^5C) is found in vertebrate DNA as a minor base located predominantly in m^5CG sequences with the methyl group added after DNA replication (1). The amount of this base in the genome depends on the type of organism as well as the tissue of origin, but highly repetitive sequences are commonly enriched in m^5C (1,2; Gama-Sosa et al., submitted for publication). Similar tissue-specific differences in the extent of DNA methylation were generally seen in highly and moderately repeated DNA sequence classes and in the single copy class of DNA sequences from a variety of human organs (2). However, several examples of unequal changes in methylation in highly repetitive or satellite DNA fractions compared to that in the analogous unfractionated mammalian DNAs have been reported (2-4).

Much evidence indicates that one of the functions of vertebrate DNA methylation is to participate in negatively controlling transcription (1). Most of these studies involved specific, highly expressed vertebrate or viral genes (1,5-12). However, these findings do not explain the considerable tissue-specific differences of up to 3×10^7 m^5C residues per diploid cell,

which have been found in mammals (2,13,14; Gama-Sosa et al., submitted for publication).

In order to further our understanding of the significance of large scale tissue-specific variations in DNA methylation and of the commonly observed enrichment in methylation of highly repeated, vertebrate DNAs, we have further analyzed the intragenomic distribution of m^5C in human DNA. We have examined the minor base composition of two dissimilar classes of highly repeated DNA sequences by high performance liquid chromatography (HPLC) of DNA enzymatically degraded to deoxynucleosides. These repeated sequences are the Alu family, which is present as ~5 x 10^5 copies interspersed into the haploid, human genome (15) and the Eco family of tandemly repeated sequences of the alphoid type (16), which is found in ~1.5 x 10^5 copies of the dimeric unit per haploid human genome (17, 18).

MATERIALS AND METHODS

Purification of human DNA and preparation of subfractions.

To isolate the Alu family sequences, DNA purified as previously described (2) was sheared to ~2 kilobase pairs (kb) by sonication, denatured, reassociated at a Cot of 68 or 0.05 molar sec, treated with 15 U of nuclease S1 (Boehringer-Mannheim) per µg of DNA, and then chromatographed on hydroxyapatite to obtain the double-stranded fraction. (19). This family of DNA sequences was then isolated by gel electrophoresis as a discrete 0.30 kb band which was recovered from the gel, purified (20), and digested with AluI. Electrophoresis in a 2.5% agarose gel gave doublet, ~0.17 + 0.13 kb bands (19), which were subsequently purified. The Eco family was isolated from total DNA by electrophoresis after digestion with EcoRI (21).

Analysis of the extent of methylation.

For analysis of the nucleoside content, DNA fractions were quantitatively hydrolyzed to deoxynucleosides and the major and minor base composition determined by HPLC on a reversed phase column by a modification (Gehrke et al., in preparation) of our previous method (22) using a Supelcosil LC-18DB column (Supelco). From 2 to 5 µg of hydrolyzed DNA was chromatographed in the presence of 8-bromoguanosine as an internal marker.

For analysis of methylation at HhaI sites in the Eco family repeats, DNA was digested for 5 h with HhaI (5 U/µg of DNA) or, as a control, with EcoRI under standard conditions. Internal controls for each type of DNA demonstrated that no inhibitors of HhaI activity were present in any of the samples. The resulting DNA fragments were electrophoresed and blot hybrid-

ized as described by Southern (23) except that the hybridization conditions were incubation at 42°C in 6 X SSC, 50% formamide, 5 X Denhardt's solution, 200 μg/ml denatured salmon sperm DNA, 0.1% Sarkosyl, 3 mM EDTA, 18 mM Tris-HCl, pH 7.4, and the DNA in the gel was partially depurinated (24). The probe for hybridization was from a recombinant pBR322 DNA, pB(EcoRI-2°)6 (a gift from Joseph Maio; 25), containing an insert of the human Eco dimeric repeat at the EcoRI site of the plasmid. It was nick-translated with E. coli DNA polymerase I (26) to a specific activity of 1-2 x 10^8 cpm/μg.

RESULTS

Levels of m^5C in Alu family sequences

As is true of the Cot ≤0.05 reassociation fraction from which it was derived (2), the Alu family was m^5C-rich compared to the total human genome or to nuclear DNA (Table 1). The Alu family of brain DNA and placental DNA had 1.54 and 1.61 mol% m^5C, respectively, a difference of only 4%, which is within experimental error. In contrast, unfractionated human brain DNA has 29% more m^5C than does the analogous placental DNA (2). The Alu family sequences of placental and brain DNAs were on the average about 2.1 and 1.6 times, respectively, as methylated as those of the total corresponding genomes. This enrichment in m^5C content was not due to commensurate increases in the G + C content since the C contents of the Alu family and unfractionated human DNA were 23.0 ± 0.3 and 19.6 ± 0.1 mol%, respectively, a difference of only ~17%. It can, however, be explained by an unusually high percentage of CG-containing sequences (27,28) as will be discussed below.

Levels of m^5C in other DNA subfractions obtained by reassociation kinetics

Besides examining the deoxynucleoside composition of 0.3 kb Alu family sequences, we determined the composition of fragments of lower molecular weight (LMW) than that of the Alu family and those of higher molecular weight (HMW) in the S1-resistant, Cot ≤0.05 reassociated fraction from which the Alu repeat family was isolated. In the cases of both brain and placenta, the HMW and LMW subfractions of highly repeated DNA were markedly hypermethylated (Table 1). Unlike the Alu family sequences, these two sets of repeated sequences showed clear tissue specificity in their DNA methylation since the analogous brain and placental sequences differed even more in their m^5C content than did the total Cot ≤0.05 DNA fractions from these two organs (Table 1).

Similar results for the Alu family were obtained when instead of a Cot ≤0.05 fraction, a Cot ≤68 fraction was used as the source of Alu family

TABLE 1. Mean m^5C levels in various human DNA fractions[a]

Source of DNA	Mol% m^5C							
		Cot ≤0.05		Cot ≤0.05 subfractions			*Eco* family	
	Total	HAP[b] column fraction	S1[b] digest fraction	*Alu*[c] family	HMW[d]	LMW[d]	Dimer	Tetramer
A) Brain	0.98 ± 0.03(6)	1.52 ± 0.05	1.90	1.54 ± 0.02	1.94 ± 0.01	1.97 ± 0.03	1.36	1.04
B) Placenta	0.76 ± 0.03(7)	1.31 ± 0.06	1.63	1.61 ± 0.06	1.51 ± 0.01	1.61 ± 0.00	0.82	0.61
C) Sperm	0.84 ± 0.01(6)	1.20 ± 0.03	-	-	-	-	0.67	-
A ÷ B[e]	1.29	1.16	1.16	0.96	1.28	1.22	1.66	1.70

[a]The mean ± standard error or standard deviation is given for data from more than one sample. Unless the number of samples is given in parenthesis, such analyses were performed on two or three independently prepared samples.

[b]The HAP fraction refers to the double-stranded DNA eluted from a hydroxyapatite column after reassociation of DNA sonicated to ~0.3 kb. The S1 digest fraction is double-stranded DNA isolated as nuclease S1-resistant material after reassociation of DNA fragments of ~2 kb.

[c]The *Alu* family sequences were obtained from two different placentas or brains. One placental *Alu* family fraction was isolated from a 0.3 kb band after a single electrophoresis of nuclease S1-treated, reassociated DNA. The other *Alu* family fractions were further purified by digestion with *Alu* I and electrophoresis. Three determinations were performed on placental samples and two on brain.

[d]HMW and LMW refer respectively to high and low molecular weight DNA fractions from the first electrophoresis gel for isolating *Alu* family sequences. The former are >∿320 bp and the latter are <∿280 bp.

[e]The ratio of mol% m^5C in the brain DNA fraction divided by the mol% m^5C in the same fraction from placenta is given.

sequences as was done in the procedure of Houck et al. (19). This was expected since almost all the Alu family repeats reanneal at Cot 0.05. The HMW and LMW subfractions from the Cot ≤68 DNA fraction of brain had 47 and 48%, respectively, more m^5C than the analogous placental subfractions although they were much lower in m^5C content than the corresponding Cot ≤0.05 fractions. This result on S1-treated DNA is consistent with our previous finding that the Cot 0.05-50 hydroxyapatite fraction of brain DNA has 38% more m^5C than the same type of fraction from placenta (2) and indicates that moderately repetitive DNA sequences can show striking tissue-specific differences in DNA methylation.

Levels of m^5C in the Eco RI repeat family

Like the above HMW and LMW nuclease S1-treated, repeated DNA sequences (Table 1), the Eco repeat family sequences from brain and placental DNAs differed considerably in their m^5C content. The Eco family of DNA repeats from brain was even more hypermethylated compared to that from placenta than were the analogous Cot ≤0.05 sequences of which the Eco family comprises ∿5% (Table 1). This family of human DNA repeats was isolated as discrete 0.34 kb and 0.68 kb bands after exhaustive digestion of total DNA with EcoRI (21). The former band is a dimer with 73% homology between the two halves and the latter band is a tetramer (21). The Eco family dimer and tetramer fractions from brain and placenta had equivalent C contents, 19 mol%; dimer sequences from both brain and placental DNAs were ~31-34% more highly methylated than the analogous tetrameric sequences. The significance of the differences between the dimer fraction and the analogous tetramer fraction is uncertain because the band of dimer sequences isolated by gel electrophoresis is ~75-80% pure while the tetrameric sequence band is somewhat less pure (17). Analysis of the Eco repeat family dimer from sperm showed it to be yet lower in its m^5C content than the same sequence family from placenta (Table 1).

Hypomethylation of a HhaI site in the Eco repeat family of sperm DNA

Wu and Manuelidis (21) showed that a minor fraction of Eco family repeats in human placental DNA contain an unmethylated HhaI site which gives a ladder of multimeric repeats upon HhaI digestion, gel electrophoresis, and blot hybridization to an Eco dimer repeat probe. We have confirmed their result and obtained similar findings with human brain, lymphocyte, lung, heart, and liver DNAs (Fig. 1 and data not shown). However, sperm DNA gave a five to ten fold more intense ladder pattern of HhaI-produced multimeric bands of moderate or low molecular weight in an identical experiment (Fig. 1). This indicates that in sperm a much higher percentage of HhaI sites in

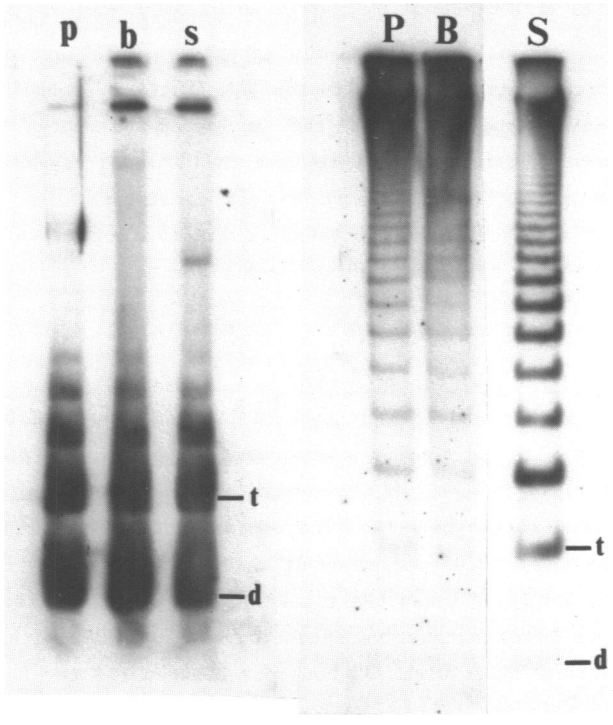

Fig. 1. Restriction and blot hybridization analysis of the human Eco repeat family. Samples of human DNA (4 µg each) were digested with EcoRI or HhaI, electrophoresed in a 1.4% agarose gel and then blotted and hybridized to the cloned Eco dimeric repeat probe as described in Material and Methods. The three samples on the left are EcoRI digests of placental (p), brain (b), and sperm (s) DNAs which were blot hybridized and processed for autoradiography simultaneously. The three samples on the right (P, B, and S) are the analogous HhaI digests except that the placental and brain DNAs were cut out and then exposed to X-ray film twice as long as was the sperm DNA to enhance their ladder of hybridizing bands. The dimeric (d) and tetrameric (t) bands of Eco repeats are indicated. The bands at the top of lanes b and s are probably artifactual. The extra minor band seen in the middle of lane s was observed in several experiments but only with sperm DNA and is of unknown significance. The ladder of bands in lane B is less distinct than that in lane P because this brain DNA preparation was slightly degraded.

the Eco repeat family are unmethylated than in DNA from various somatic cell populations. Quantitatively similar results were obtained with sperm DNA isolated by our standard procedure (2) or by a method involving 7 M urea (29) instead of trypsin treatment. That we obtained a multimeric ladder in the HhaI digest with a cloned Eco dimeric repeat probe just as Wu and Manuelidis (21) did with a genomic Eco dimeric probe, indicates that this ladder is due

to the Eco repeat sequence itself and not to some minor contaminating sequence in the EcoRI digest, which co-electrophoreses with the Eco dimer repeat. In EcoRI digests, as expected, the Eco family dimeric and tetrameric repeats from equivalent sperm and somatic cell DNA preparations gave hybridizing bands of the same intensity (Fig. 1).

DISCUSSION

Just as the unique sequence fraction of human brain DNA has a considerably (36%) higher level of m^5C than that of placental DNA (2) so do several repetitive subfractions of brain DNA compared to the analogous subfractions of placental DNA (Table 1). Most notably, the Eco repeat family, which consists of ∿0.34 kb, tandemly repeated sequences comprising ~1% of the human genome and localized in centromeres in vivo (17,18), showed a large difference in m^5C content between brain and placental samples. The former had 66% more m^5C than the latter (Table 1). Brain Eco family dimeric (17) sequences had 1.36 mol% m^5C, which might represent complete methylation of all CG dinucleotide sequences, the predominant sites of vertebrate DNA methylation (1), because in the human Eco family consensus sequence 1.5% of the residues are C residues bordered on the 3'-side by a G residue (2). Furthermore, the dimeric Eco family fraction purified by the methods used is ~75-80% pure (17). The tissue specificity of methylation of the Eco family sequences probably cannot be correlated with transcriptional activity because these low complexity sequences are presumed to be transcriptionally inactive in vivo and transcripts from these sequences could not be detected in HeLa cells (J. Maio, pers. commun.).

Eco family dimer sequences from human sperm DNA showed even greater hypomethylation compared to those of brain or placental DNA (Table 1). This is in agreement with our previous finding that the Cot ≤0.05 fraction of sperm DNA contained less of the total m^5C of the genome than did the Cot ≤0.05 fraction from a number of other human cell populations (2) and with the finding of extreme hypomethylation of one tandemly repeated DNA sequence family in bovine sperm DNA (3,30). The hypomethylation of sperm and placental Eco genomic repeats compared to analogous brain repeats was not similarly distributed over the CG sites of these repeats since only the sperm repeats were extensively hypomethylated at a HhaI recognition site (Fig. 1). All the HhaI sites in the sperm sequences might have been unmethylated. The multimeric series of Eco family bands in the HhaI digests of sperm DNA could have been a result of most of these repeats not having the GCGC (HhaI) recognition

sequence (21) rather than a consequence of partial methylation of HhaI sites.

Unlike the other studied repeated DNA sequence subfractions (Table 1) and the single copy sequences with which it is interspersed (2), the AluI family of repeats, which constitutes ~5% of the human genome, exhibited a remarkable conservation of the extent of methylation. These sequences isolated from brain and placenta differed by <5% in their m^5C content (Table 1). Although with ~1.6 mol% m^5C, these sequences were considerably hypermethylated compared to the total human genome, they are very rich in CG dinucleotides (27,28). According to the consensus sequence deduced from 15 independently cloned copies of the Alu family repeat there could be ~4.7 mol% m^5C if all the CG dinucleotides were methylated and if CG were the only methylated dinucleotide as has been found for several other repeated DNA sequences in vertebrates (31-34). Therefore, in contrast to the unfractionated genome and the single copy sequences (2) with which it is interspersed, the Alu family contained only a minor portion of its CG dinucleotide sequences in the methylated form. The Alu family is, therefore, an example of a group of sequences enriched in both unmethylated CG dinucleotides and m^5C residues, which, in contrast to other major sequence classes of the human genome, shows little overall variation in methylation levels.

ACKNOWLEDGEMENTS

We are grateful to Drs. Monroe Samuels and Richard Harrison for providing us with various human organ and cell samples and to Dr. Joseph Maio for giving us the cloned Eco repeat sequence. We thank Mrs. Agnes Boyd for help with the manuscript. This study was supported in part by National Institutes of Health grant CA-19942.

REFERENCES

1. Ehrlich, M. and Wang, R.Y.-H. (1981) Science 212, 1350-1357.
2. Ehrlich, M., Gama-Sosa, M.A., Huang, L.-H., Midgett, R.M., Kuo, C.C., McCune, R.A., and Gehrke, C., (1982) Nucleic Acids Res. 10, 2709-2721.
3. Sturm, K.S. and Taylor, J.H. (1981) Nucleic Acids Res. 9, 4536-4546.
4. Shmookler Reis, R.J. and Goldstein, S. (1982) Proc. Natl. Acad. Sci. USA 79, 3949-3953.
5. Ott, M.-O., Sperling, L., Cassio, D., Levilliers, J., Sala-Trepat, J., and Weiss, M. C., (1982) Cell 30, 825-833.
6. DeSimone. J., Heller, P., Hall, L., and Zwiers, D., (1982) Proc. Natl. Acad. Sci. USA 79,4428-4431.
7. Stein, R., Razin, A., and Cedar, H. (1982) Proc. Natl. Acad. Sci. USA 79, 3418-3422.
8. Harris, M. (1982) Cell 29, 483-492.
9. Vardimon, L., Kressmann, A., Cedar, H., Maechler, M. and Doerfler, W. (1982) Proc. Natl. Acad. Sci. USA 79, 1073-1077.

10. Jahner, D., Stuhlmann, H., Stewart, C.L., Harbers, K., Lohler, J., Simon, I., and Jaenisch, R. (1982) Nature 298, 623-628.
11. Hoffmann, J.W., Steffen, D., Gusella, J., Tabin, C., Bird, S., Cowing, D., and Weinberg, R.A., (1982) Virology, 44, 144-157.
12. Clough, D.W., Kunkel, L.M., and Davidson, R.L. (1982) Science 216, 70-73.
13. Kappler, J.W. (1971) J. Cell Physiol. 78, 33-36.
14. Vanyushin, B.F., Mazin, A.L., Vasilyev, V.K., and Belozersky, A.N. (1973) Biochim. Biophys. Acta 299, 397-403.
15. Jelinek, W.R. and Schmid, C.W. (1982) Ann. Rev. Biochem. 51, 813-844.
16. Maio, J. J., Brown, F. L., and Musich, P. R. (1977) J. Mol. Biol. 117, 637-655.
17. Manuelidis, L. (1978) Chromosoma 66, 1-21.
18. Manuelidis, L. (1976) Nucleic Acids Res. 3, 3063-3076.
19. Houck, C.M., Rinehart, F.P., and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306.
20. Yang, R.C.-A., Lis, J., and Wu, R. (1979) in Methods in Enzymology, R. Wu, Ed., Vol. 68, pp. 176-182. Academic Press, New York.
21. Wu, J. C. and Manuelidis, L. (1980) J. Mol. Biol. 142, 363-386.
22. Kuo, K.C., McCune, R.A., Gehrke, C.W., Midgett, R., and Ehrlich, M. (1980) Nucleic Acids Res. 8, 4763-4776.
23. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
24. Wahl, G.M., Stern, M., and Stark, G.R. (1979) Proc. Natl. Acad. Sci. 76, 3683-3687.
25. Shafit-Zagardo, B., Maio. J.J., and Brown, F.L. (1982) Nucleic Acids Res. 10, 3175-3193.
26. Rigby, P.W., Dieckmann, M., Rhoades, C., and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
27. Deininger, P.L., Jolly, D.L., Rubin, C.M., Friedmann, T., and Schmid, C.W. (1981) J. Mol. Biol. 151, 17-33.
28. Rubin, C.M., Houck, C.M., Deininger, P.L., Friedman, T., and Schmid, C.W. (1980) Nature 284, 372-374.
29. Bird, A.P. and Taggert, M.H. (1980) Nucleic Acids Res. 8, 1485-1497.
30. Pages, M. and Roizes, G. (1982) Nucleic Acids Res. 10, 565-576.
31. Sano, H. and Sager, R. (1982) Proc. Natl. Acad. Sci. USA 79, 3584-3588.
32. Manuelidis, L. (1981) FEBS Let. 129, 25-28.
33. Miller, J. R., Cartwright, E. M., Brownlee, G. G., Fedoroff, N. V., and Brown, D. D. (1978) Cell 13, 717-725.
34. Pech, M., Igo-Kemens, T., and Zachau, H. G. (1979) Nucleic Acids Res. 7, 417-432.