Absence of methylation at HpaII sites in three human genomic tRNA sequences

Daniel F.Schorderet and Stanley M.Gartler
Departments of Medicine and Genetics, SK-50, University of Washington, Seattle, WA 98195, USA

Received August 13, 1990; Revised and Accepted November 6, 1990

ABSTRACT

It has been known since the development of nearest neighbor analysis that the frequency of the dinucleotide CpG is markedly suppressed in vertebrate DNA (i.e. less than %C x %G). This suppression appears to be heterogeneous since it was shown some years ago that three vertebrate tRNA genes did not exhibit CpG suppression. We have analyzed 13 different human tRNA genes and found that they also do not exhibit CpG suppression. Because CpG suppression has been linked, to some extent at least, to the methylation-deamination process by which a methylated CpG is mutated to TpG, we investigated whether the lack of suppression of CpG in tRNAs could originate from an absence of methylation.

Three human tRNA genes were selected from Genbank (Lysine, Proline, and Phenylalanine) and examined for methylation at HpaII sites by polymerase chain reaction (PCR) and Southern blot analysis. The observed patterns were consistent with the absence of methylation at the seven HpaII sites analyzed in and around the tRNA genes, and we predict that the remaining CpGs in these genes will be unmethylated. Since GC-rich promoter regions also escape CpG suppression and since they are generally unmethylated, avoidance of methylation may be a general explanation for the absence of CpG suppression in selected regions of vertebrate genomes.

INTRODUCTION

It has been known since the development of nearest neighbor analysis in the early 1960's that the frequency of the dinucleotide CpG is much lower in vertebrate DNA than its expected value calculated from the frequency of cytosine (C) and guanine (G) bases (1−3). Some 15 years later, Russell et al. showed that CpG suppression was not present in three vertebrate tRNA genes (4) which represent GC-rich regions such as the more recently discovered HTF islands 5′ to housekeeping genes (5). On the basis of their observation that tRNA genes, in contrast to coding regions, did not exhibit CpG suppression, Russell et al. (4) proposed that CpG avoidance was a consequence of selection acting on the translation apparatus.

The fact that some vertebrate mRNAs are not CpG-suppressed prompted Salser (6) to point out that the translational basis for CpG suppression could not be universal. He suggested instead that CpG might represent a mutational hotspot, especially since it was known to be methylated and methylation might lead to incorrect base pairing. A year later, Coulondre et al. (7) showed that methylated cytosine represented a mutational hotspot in E. coli not because of incorrect pairing but by its deamination to thymine. If a cytosine is deaminated, a uracil is produced which is removed from the DNA by the repair enzyme, uracil glycosylase. However, in the case of a G-T mismatch resulting from deamination of a methylated cytosine, it was assumed that the resulting thymine would not be recognized as abnormal and a high mutation rate would result. Methylated cytosines are deaminated spontaneously at a relatively high rate and, if uncorrected, could result in mutation rates orders of magnitude higher than at other bases. In recent years, specific G-T mismatch repair systems have been described in both microbial and animal cells which preferentially correct a G-T mismatch to G-C (8−11). However, these repair systems are not perfect, and it is clear that methylated cytosines represent mutational hotspots in humans as well as bacteria (12). These mutations could, at least partially, explain the CpG suppression found in these species.

We analyzed the CpG composition of 13 human tRNA genes and found in all cases that the observed frequencies of CpGs were not significantly different from expected values. How does one account then for the lack of CpG suppression in certain DNA regions? The simplest explanation would be that they remain methylation-free. This, in fact, is the case for the GC-rich mammalian promoters which, like the tRNA genes, are not suppressed in CpGs. We suggest that other non-suppressed CpG regions are also methylation-free and, in this paper, report the absence of methylation at seven HpaII sites in and around three different human tRNA genes.

MATERIALS AND METHODS

Source of tRNAs

Three human tRNA genes were selected from Genbank with the condition that they were accompanied by flanking regions so that, after restriction enzyme digestion and amplification by the polymerase chain reaction (PCR), their DNA could easily be distinguished by size. Proline (Pro) (13), phenylalanine (Phe) (14), and lysine (Lys) (14), available under the respective accession numbers M15274, M17622, and M17620, fulfilled these conditions, and a pair of primers, flanking each tRNA gene (Table 1), were synthesized and used to amplify the corresponding DNA region by PCR technology.

PCR amplification of the tRNA genes and methylation analysis

Three hundred nanograms of human male genomic DNA, extracted from sperm or blood leukocytes as described (15), were digested either with HpaII or MspI according to the manufacturer's conditions and PCR-amplified using their respective pair of primers described above. The PCR reactions were carried out in a total volume of 100 ul with the final concentration of primers being 1 uM in 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin (w/v), 1.25 mM of dATP, dCTP, dGTP, and dTTP and 2.5 units of Taq DNA polymerase (Amplitaq, Perkin Elmer Cetus). The cycler (Ericomp) conditions were as follows: 94°C, 1 min; 60°C, 1 min; 72°C, 2 min; 23 cycles. The same amount of uncut DNA derived from the same individual was amplified as a control. All the PCR products were resolved on a 0.7% agarose gel and stained with ethidium bromide. Since the absence of a product in the target region could be due either to an unmethylated HpaII site or to failure of PCR amplification, we simultaneously amplified in the same reaction a region containing exon 9 of the phosphoglycerate kinase 1 (PGK1) gene which contains a HpaII site. This site is methylated on the active X chromosome (data not shown) and would give rise to a 700 base pair (bp) band upon PCR amplification. Because we amplified male DNA, there is only one site present and that site will only be cut by MspI.

Table 1. Primers used for PCR analysis and probe generation

Gene	Primers
Proline	5′ TGT AAT ACC ATG ACA AAG AGC 3′
	5′ AGG AGA GAA AAC CTG CAG TGG 3′
Phenylalanine	5′ TGC CTA CCT CTA GCT GAA TCC 3′
	5′ GGT AAG ACA GTA TGT CCC CAG 3′
Lysine	5′ GCA AGA ACC CAC TGA GCA AGT T 3′
	5′ ATG CGG CCA CGT GGC CTA TTT 3′

Southern blot analysis

DNA derived from blood leukocytes was double-digested with TaqI and HpaII or MspI for the proline tRNA gene and with DraI and HpaII or MspI for the phenylalanine and lysine tRNA genes. The fragments were separated by polyacrylamide gel electrophoresis (7%), electrotransferred onto a nylon membrane (Zeta-Probe, Bio-Rad). The resultant membrane was prehybridized for 5 minutes in 50% formamide, 4×SSPE (1×SSPE = 180 mM NaCl, 10 mM $Na_2HPO_4 \cdot 7H_2O$, 1 mM EDTA), 0.5% 'Blotto' (w/v) (Carnation), 1% SDS, 0.5 mg/ml denatured salmon testes DNA, and 10% dextran sulfate (w/v) at 43°C. The different tRNA probes were generated as follows: 300 ng of male DNA were amplified during 25 cycles in a total volume of 100 ul using the same PCR conditions as described above. After amplification, the PCR products were gel-purified in low-melting agarose (Sea Plaque, FMC BioProducts) and ^{32}P-labeled using a random primed DNA labeling kit (Boehringer-Mannheim). Hybridization was performed for 16 hrs using the labeled tRNA probe which was denatured and added to the prehybridization solution. The final wash buffer contained 0.25×SSC (1×SSC = 150 mM NaCl, 15 mM sodium citrate) at 65°C, and the blot was exposed to a Kodak XAR film for 2 to 7 days.

RESULTS

The CpG composition of 21 tRNA genes is shown in Table 2. The data were derived from sequences found in Genbank. There is no significant difference in the number of CpG dinucleotides per gene between species with methylated DNA (human) and species without methylation (Caenorhabditis elegans, Drosophila melanogaster, and Saccharomyces cerevisiae), and the actual CpG composition is not significantly different from an expected value calculated from the base composition, indicating that there is no CpG suppression in these tRNA genes (chi-square, $p > 0.5$).

To examine the methylation status of the tRNA genes, we selected three genes (proline, phenylalanine, and lysine) for

Table 2. CpG and GpC content of tRNA genes

Species	Name	Length [bases]	Composition G	C	Observed CG	GC	Expected CG or GC
H. sapiens	ala	76	24	10	6	9	6
	phe	76	23	18	5	4	5
	gly[a]	74	23	23	8	8	7
	gly[b]	74	24	22	6	7	7
	asn	77	26	21	8	7	7
	val[a]	76	22	21	9	3	6
	val[b]	76	21	22	9	3	6
	pro	72	26	18	4	6	6
	thr	73	24	17	6	3	6
	met[a]	72	23	19	5	8	6
	lys[a]	73	23	18	4	4	6
	gln	72	19	22	3	2	4
	leu	82	28	23	9	8	8
C. elegans	leu	82	28	22	6	6	8
	trp	72	24	18	8	7	6
D. melanogaster	phe[2]	76	22	19	5	4	6
	glu[4]	75	21	20	3	3	6
	his	76	24	22	6	3	7
S. cerevisiae	ala[1]	75	28	22	10	6	8
	phe	76	23	18	3	5	5
	gly	73	22	22	7	6	7

further analysis using PCR amplification. In each gene, one HpaII site can be analyzed by this PCR technique (Fig. 1). Figure 2 shows the PCR products obtained when human male white blood cell (Fig. 2 top) or human sperm (Fig. 2 bottom) DNA was amplified after HpaII or MspI digestion, using each set of primers for the three different tRNA genes. Only uncut DNA is amplified and shows the expected tRNA products (Fig. 2, lanes 3,6,9). As expected, when the DNA is cut with MspI, no product is obtained (Fig. 2, lanes 2,5,8). The same pattern is found when the DNA is cut with HpaII (Fig. 2, lanes 1,4,7), indicating the

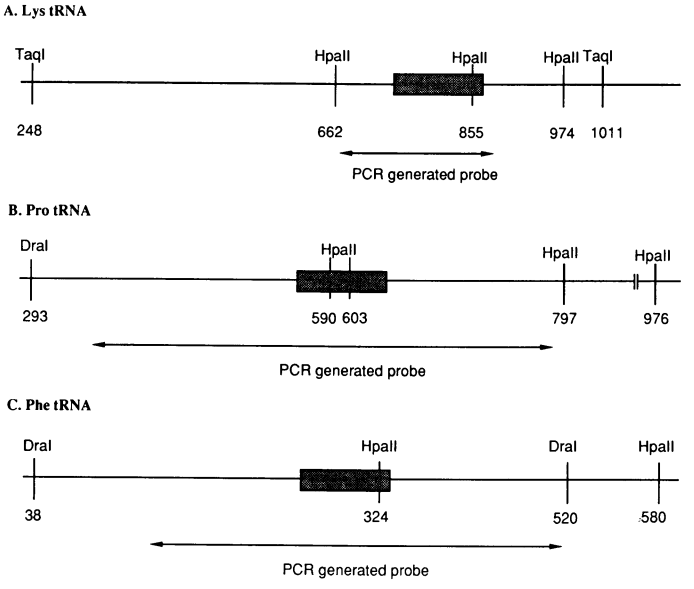

Figure 1. Regional location of the analyzed HpaII sites for the human Lys (A), Pro (B), and Phe (C) tRNA genes. The base numbers are from Genbank, and the tRNA genes are represented by a shaded box. Using the primers described in Table 1, the region marked by the arrow was PCR-amplified and used for the direct PCR analysis or was used as a probe for Southern blot analysis.

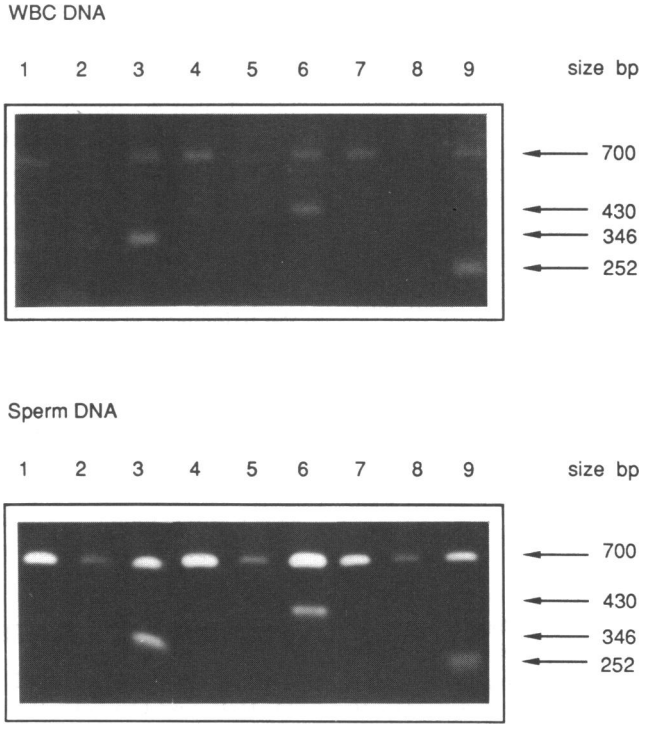

Figure 2. PCR analysis of the methylation status of the human Lys, Pro, and Phe tRNA genes. Male human DNA (top: white blood cell; bottom: sperm) was amplified with the primers for the following tRNAs: Phe lanes 1–3; Pro lanes 4–6; Lys lanes 7–9. Prior to PCR amplification, the DNA was digested as follows: lanes 1,4,7: HpaII; lanes 2,5,8: MspI; lanes 3,6,9: no digestion. The control band at 700 bp is from the amplification of a methylated HpaII site of the PGK1 gene.

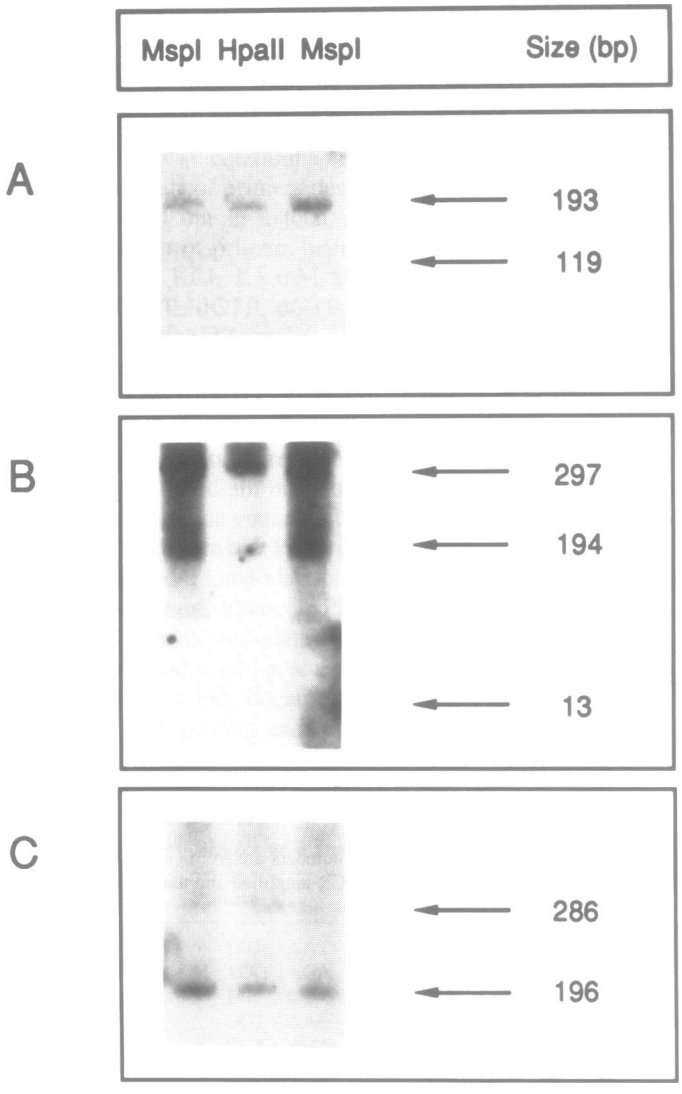

Figure 3. Southern blot analysis of the methylation status of the human Lys, Pro, and Phe tRNA genes. DNA was double-digested with TaqI and MspI or HpaII for the Lys tRNA gene (A); with DraI and MspI or HpaII for the Pro (B) and Phe (C) tRNA genes. The locations of the probes are indicated in Figure 1. The expected band size is indicated at the right.

absence of methylation of the three (or possibly four) HpaII sites found in the tRNA genes. Because the primers flank the two HpaII sites of the proline tRNA gene, hypomethylation of either or both sites would allow HpaII to cut the DNA strand and inhibit PCR amplification. Therefore, at least one of the two sites is unmethylated. The control band at 700 bp results from the coamplification of part of the PGK1 gene in each reaction. As expected, this band was present when the DNA was digested with HpaII, absent when digested with MspI, and present in undigested DNA (Fig. 2).

These PCR-based results were confirmed and extended with Southern blot analysis. For the lysine tRNA gene, DNA was double-digested with TaqI and HpaII or MspI which results in the analysis of 3 different HpaII sites, one being inside the lysine tRNA gene (Fig. 1A). Methylation of all or any of these HpaII sites will lead to distinctive patterns. The Southern blot analysis shows a 193 bp band, indicating that both the lysine tRNA internal HpaII site and HpaII site #662 are not methylated (Fig. 3A). A very faint band of 119 bp size is also present, suggesting hypomethylation of HpaII site #974. For the proline tRNA gene, three HpaII sites are in the region covered by the probe, with two of them being intragenic (Fig. 1B). The Southern blot analysis using DraI with MspI or HpaII digestion revealed three bands: one at 297 bp, a second band at 194 bp, and, after a longer exposure, a third band at 13 bp (Fig. 3B). This is the unique pattern expected when all three HpaII sites are unmethylated. We also observed a faint band around the 140 bp marker. We do not know whether this represents the HpaII(797)-HpaII(976) fragment (which is not covered by the probe), or if this represents nonspecific hybridization to some proline isoacceptor tRNA genes. Sequence analysis of the region around the phenylalanine tRNA gene shows only one HpaII site covered by the PCR-generated probe. The Southern blot analysis, using DraI with MspI or HpaII digestion, showed two bands of 286 bp and 196 bp (Fig. 3C), which represents the expected pattern if the HpaII site inside the tRNA gene is unmethylated.

DISCUSSION

As indicated in the INTRODUCTION, the well-known CpG suppression found in vertebrate DNA is heterogeneous within the genome. CpG-rich regions, such as promoters of housekeeping genes (5,6) and tRNA genes (4), do not exhibit CpG suppression. This was confirmed by the sequence analysis of 21 tRNA genes. The number of CpG dinucleotides was not statistically different from the number of GpCs, which is known to occur at the expected frequency in all the DNA studied so far and is roughly equal to the predicted number.

Considerable evidence supports the hypothesis that at least part of the general CpG suppression found in vertebrate DNA is due to the mutational instability of 5-methylcytosine (5), and several ideas have been suggested over the years to explain the lack of CpG suppression in high GC-regions. These include resistance of GC-rich regions to deamination (16) and absence of methylation of these regions in the germ lines (6,13). We have investigated the possibility that CpG is not suppressed in these regions because it is not methylated in somatic tissues and in the male germ line.

Our analysis of the methylation status of three human tRNA genes using the methyl-sensitive enzyme HpaII showed that the intragenic and nearby HpaII sites in all three tRNA genes were unmethylated. These PCR and Southern blot analyses support our hypothesis of absence of methylation of tRNA genes in human leukocytes and the male germ line. However, our evidence is limited and further studies of all the CpGs in a particular tRNA and of more tRNAs will be necessary before one can generalize our observations to an entire tRNA molecule and to all tRNAs. The data presented here, as well as the more extensive observations of GC-rich promoters of mammalian housekeeping genes, are beginning to make a case for the general proposition of absence of methylation in regions without CpG suppression.

ACKNOWLEDGEMENTS

The authors wish to thank Drs. Christine M. Disteche, Benjamin D. Hall, and Carol H. Sibley for comments on the manuscript and Kenneth L. Roy for his help. This work was supported by National Institutes of Health Grant HD-16659. D.F.S. is supported by the Swiss National Science Foundation (grant # 32-9253.87) and S.M.G. is a recipient of a National Institutes of Health Career Award.

REFERENCES

1. Josse,J., Karser,A.D. and Kornberg, A. (1961) J. Biol. Chem. **236**:864–875.
2. Swartz,M.N., Trautner,T. and Kornberg,A. (1962) J. Biol. Chem. **237**:1961–1967.
3. Setlow,P. (1976) In Fasman,G.D. (ed.), Handbook of Biochemistry and Molecular Biology, CRC Press, Cleveland, pp. 312–318.
4. Russell,G.J., Walker,P.M.B., Elton,R.A. and Subak-Sharpe,J.H. (1976) J. Mol. Biol **108**:1–23.
5. Bird,A., Taggart,M., Frommer,M., Miller,O.J. and Macleod,D. (1985) Cell, **40**:91–99.
6. Salser,W. (1977) Cold Spring Harbor Symp. Quant. Biol. **42**:985–1002.
7. Coulondre,C., Miller,J.A., Farabaugh,P.J. and Gilbert,W. (1978) Nature **274**:775–780.
8. Lindhal,T. (1982) Ann. Rev. Biochem. **51**:61–87.
9. Brown,T.C. and Jiricny,J. (1987) Cell **50**:945–950.
10. Brown,T.C. and Jiricny,J. (1988) Cell **54**:705–711.
11. Rideout III,W.M., Coetzee,G.A., Olumi,A.F. and Jones,P.A. (1990) Science **249**:1288–1290.
12. Cooper,D.N. and Youssoufian,H. (1988) Hum. Genet. **78**:151–155.
13. Chang,Y.-N., Pirtle,I.L. and Pirtle,R.M. (1986) Gene **48**:165–174.
14. Doran,J.L., Wei,X. and Roy,K.L. (1987) Gene **56**: 231–243.
15. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular cloning. A laboratory manual. 2nd edition, Cold Spring Harbor Laboratory Press, 1989.
16. Adams,R.L.P. and Eason,R. (1984) Nucleic Acids Res. **12**:5869–5877.
17. Tykocinski,M.L. and Max,E.E. (1984) Nucleic Acids Res. **12**:4385–4396.