The pattern of methylation in human genes for 18S and 28S ribosomal RNA has been investigated using methylation-sensitive restriction enzymes. We find that the transcribed region of the repeat unit is predominantly unmethylated, in agreement with previous studies. In contrast the non-transcribed spacer, which makes up the majority of the 43 kb repeat unit, is highly methylated in blood cell DNA. The boundaries between methylated and non-methylated domains appear to be relatively sharp, and occur ~1.5 kb upstream of the 5' edge of the proximal promoter and ~1.0 kb downstream of the 3' end of the transcribed region. A small proportion of all repeat units are methylated throughout the transcribed region, and may represent silent genes. The coincidence between the methylation pattern, the transcription pattern and other features of the repeat unit has implications for our understanding of the mechanism by which patterns of DNA methylation are generated.

INTRODUCTION

The human 18S and 28S ribosomal RNA (rRNA) genes are present at ~400 copies per human haploid genome, clustered on the short arms of the five acrocentric chromosomes (1). Each gene is part of a 43 kb repeat unit that can be divided into two regions: a 13.3 kb transcribed region which contains the highly conserved genes for 18S, 5.8S and 28S rRNA subunits of the ribosome, and a 30 kb non-transcribed spacer (NTS) (2). Repeat unit clusters consist of head-to-tail arrays of ~80 repeats (3).

Vertebrate DNA is frequently modified at the dinucleotide CpG by addition of a methyl group to give m5CpG (4). The modification can alter the ability of various restriction enzymes to cleave DNA, and this facilitates investigation of methylation at these sites (5,6). A previous study of methylation levels in the transcribed region of the rDNA repeat of various species demonstrated that the conserved region of the repeat unit exists in a predominantly unmethylated form in DNA from a variety of mammals, including man (7). In the mouse, a small fraction of all rDNA repeats are methylated throughout the transcribed region, and indirect evidence has suggested that these correspond to transcriptionally inactive repeats (8). The findings in mammals contrast with those in amphibia and fish, where chromosomal rDNA is heavily methylated in both transcribed and non-transcribed regions of the repeat unit (9,10).

This study began with isolation of fragments of the rDNA spacer from a library of densely methylated DNA fragments. The library was constructed by fractionating human blood DNA over a column that specifically binds to densely methylated DNA (11). The frequent occurrence of rDNA fragments in the library was unexpected given previous evidence that the transcribed region of mammalian rDNA is predominantly unmethylated. Further experiments showed that, in fact, human rDNA is mosaic with respect to CpG methylation. The transcribed region is indeed largely methylation-free at testable sites, but the spacer is densely methylated.

RESULTS

A library of densely methylated sequences was constructed after fractionating human blood DNA on a methyl-CpG binding column (G.J.R. Brock, unpublished). Fragments that repeatedly bound to the column were cloned in plasmids. The library was tested by examination of the DNA sequence and methylation status of random inserts. Clones containing p1A7 and p1B12 showed DNA sequence identity with the NTS of human rDNA. Each clone was represented many times in the library. Methylation was analysed by using inserts from the library as probes against Southern blots of genomic DNA that had been digested with the methylation-sensitive enzyme Smal (CCCGGG) or its methylation-insensitive isoschizomer Xmal. Probes p1A7 and p1B12 both hybridised to very high molecular weight Smal fragments, around 28 kb, whereas Xmal generated smaller fragments as predicted by the sequence of the spacer (Fig. 1). Weaker Smal bands <27 kb indicated that sites in this part of the repeat unit are occasionally non-methylated. The large size of the major band in the Smal lanes showed that p1A7 and p1B12 are part of long tracts of DNA in which multiple (at least 14) Smal/Xmal sites are usually in a methylated state. The equivalent experiment with sperm DNA again showed that the sequence environment of p1A7 and p1B12 is methylated, but the level of methylation is significantly lower than in blood DNA. Most of the rDNA repeats in sperm DNA were digested by Smal to give fragments <20 kb, and some fragments were of the same size as in the Xmal digested lane.

Hybridisation of a similar blot with probes from the transcribed region (pHsrDNA5.1/7.9) gave a contrasting result. Smal and Xmal patterns of both blood and sperm DNA were predominantly the same, indicating that the majority of sites in this part of the
In order to map the boundary between methylated and non-methylated regions of the repeat, genomic DNA was digested with EcoRI or HindIII together with a selection of restriction endonucleases which contain CpG in their recognition sequence and are sensitive to methylation. The CpG enzymes were AcII (CCGC), BstUI (CGCG), Hhal (GCGG) and HpaII (CCGG). Mspl, a methylation-insensitive isoschizomer of HpaII, was used as a control. To map the 5' boundary of the methylated domain, blots were probed with pIA7 (Fig. 2). The 15 kb HindIII fragment that hybridised to the probe spans the 3' half of the transcription unit and 6.9 kb of downstream spacer. All four methylation-sensitive enzymes in combination with HindIII produced bands in the region of 5.5-6.0 kb. This locates the furthest downstream site for each enzyme that is consistently non-methylated at 1.0-1.5 kb downstream of the 3' end of the transcription unit (Fig. 2, broken vertical line close to nucleotide 14 500). Digestion with EcoRI plus methylation-sensitive enzymes confirmed this location for the boundary between methylated and non-methylated domains of the repeat unit. EcoRI alone gave the expected band of 18.1 kb when probed with pIA7. Further digestion with methylation-sensitive enzymes gave bands clustered at 16 kb due to cleavage at position 14 500. These enzymes also generated bands at ~6 kb, indicating a discrete hypomethylated region ~200 bp downstream of the HindIII site (Fig. 2B, open arrow at nucleotide 20 500). The hybridisation signal is distributed roughly evenly between the 16 and 6 kb bands in these lanes. Survival of a high proportion of the 16 kb fragments, in spite of the presence of multiple sites for the methylation-sensitive enzymes, indicates that the 186 CpGs in the NTS that were tested in these experiments are methylated in a high proportion of repeat units.

Similar experiments were used to map the methylation boundary at the 5' end of the spacer. Figure 3 shows that the 13 kb HindIII band is reduced to ~7 kb by all four methylation-sensitive enzymes. This places the boundary between methylated spacer and the non-methylated transcription unit at ~1 kb upstream of the transcription start site. In addition there is a collection of weaker bands just below 3 kb, suggesting that this discrete region of the repeat unit is somewhat undermethylated (Fig. 3B, see open arrow at position 38 000).

**DISCUSSION**

We have demonstrated a mosaic pattern of CpG methylation in the majority of human rDNA repeat units. The 13.3 kb transcribed region is apparently free of methylation, whereas the NTS which lies between consecutive transcribed regions is highly methylated at ~300 tested CpG sites (Fig. 4A). The transition between methylated and non-methylated domains appears to be sharp and occurs near the boundaries of the transcribed region. Previous studies have only determined the methylation status of the transcribed region of the repeat unit. In this respect there is a great contrast within the vertebrates. Fish and amphibia have heavily methylated transcribed regions, whereas the mammals (and perhaps birds and reptiles) have transcribed regions that are predominantly non-methylated (7). The present results reduce the contrast somewhat, as it now seems likely that a major part of the chromosomal rDNA repeat unit may be methylated in most or all vertebrates.

The sequence of the human rDNA repeat unit shows that the dramatic variation in methylation is matched by less dramatic, but
Figure 2. Investigation of the methylation status of sites downstream of the termination of transcription at the 5′ end of the NTS. (A) Southern blot analysis of normal male lymphocyte DNA digested first with HindIII or EcoRI followed by the methylation-sensitive enzymes shown and then hybridised with p1A7. (B) Restriction map of part of the rDNA repeat unit showing the sites for EcoRI (E) and HindIII (H) and site maps for each of the methylation-sensitive restriction enzymes that were used. The boundary between methylated DNA (to the right) and non-methylated DNA (to the left) is indicated by a dotted vertical line. A hypomethylated region is marked by an open arrow. The position of probe p1A7 is shown. The origin of prominent bands seen on the autoradiographs is diagrammed below.

easily detectable, differences in sequence composition. Firstly, base composition fluctuates through the repeat unit from a high average GC content over the transcribed region, to a lower GC content over the NTS. This is apparent from the uneven frequency of GC-rich restriction enzyme sites (Fig. 4A), and also from a plot of average GC-content across the repeat unit (Fig. 4B). In some respects the structure of the repeat resembles the large scale isochore structure of the genome in which regions of differing average base composition are juxtaposed (20). One possible reason for the GC-richness of the transcribed region is selective pressure to conserve the rRNA precursor sequence. This seems an inadequate explanation, however, as the 5′ transcribed spacer evolves quite rapidly (there is little overall sequence similarity between human and rat or mouse in this region) and is therefore not apparently under strong sequence constraint (21–25). In spite of the sequence variation, the GC-rich character of the external transcribed spacer is maintained in mammals (23,24). Indeed, the human 18S sequence is less GC-rich than the transcribed spacers that flank it (see Fig. 4B). It is also apparent that the GC-rich domain, like the non-methylated domain, extends outside the region that is transcribed, to include the proximal promoter sequences (see also ref. 26) and regions downstream of the transcription termination site. This would not be expected if selection only for GC-rich RNA was operating. It seems more likely that the phenomena that relate lack of methylation to transcription also lead to GC-richness in this domain.

Another prominent feature of the repeat unit sequence is the fluctuation in CpG frequency (observed/expected). In the transcribed region, CpG occurs at near the expected level, but in the NTS there is a significant CpG deficiency (Fig. 4C). This finding fits well with the observation that methylation occurs in the NTS, as the primary cause of CpG deficiency is the mutability of m5C (26). It is possible to see analogies between the non-methylated GC-rich domain of rDNA and CpG islands, which have similar properties (26). Both are also associated with transcription, although the details differ. CpG islands usually
extend for -1 kb downstream from the promoter of a gene and rarely encompass the entire transcription unit. Only in cases where the gene is unusually small does it lie entirely within the island (for example, \( \alpha \) globin; see ref. 27). In rDNA the non-methylated domain is much larger than a typical CpG island (13 kb) and includes all of the transcription unit. Although the mosaic character of the rDNA repeat unit cannot yet be satisfactorily explained, it is notable that the region that is methylation-free and GC-rich corresponds with parts of the repeat unit that are probably associated with proteins. These are bound to the promoter and termination sites, and pass repeatedly through the transcribed region during transcription. It is conceivable that protein–DNA complexes block access to the DNA methyltransferase. A more radical possibility is that the act of transcription actively demethylates this part of the repeat unit. Preliminary evidence for RNA-mediated demethylation has been reported (28). Whatever the mechanism, it is important to keep in mind that the transcription unit of amphibian rDNA is maintained in a heavily methylated state in spite of protein associations and transcription (10).

Analysis of amplified ribosomal RNA genes in a human lymphoblastoid cell line has suggested that the methylated genes are transcribed at low levels (29). Active and relatively inactive populations of rDNA were also detected in HeLa cells (30). In the mouse, it has been shown that a minority of rDNA repeats are methylated in the transcribed region and are probably transcriptionally inactive (8). The present work shows a parallel phenomenon in normal human cells, as a small proportion of rDNA from blood is resistant to methylation-sensitive restriction enzymes in the transcribed region of the repeat unit. By analogy with the mouse, the methylated repeats may represent inactive human rRNA genes. There is evidence that lack of methylation at CpG islands depends on the binding of factors that are also required for transcription to occur (31,32). A similar requirement may apply to rDNA. Inactivity of a gene may invite de novo methylation. Once methylated, the silent repeat unit may be unable to return to transcriptional activity due to the repressive effect of methylation (33). In this way CpG methylation may stabilise the repressed state.

MATERIALS AND METHODS

DNA

DNA derived from blood was extracted using standard protocols (12), a gift from L. Strain (MRC. Human Genetics Unit,
Figure 4. (A) Schematic representation of two consecutive rDNA repeat units showing the frequency of GC-rich sites for three restriction enzymes Bsp1201, Eagl and BstUI. The transcription unit is represented by a box with an arrow indicating the start and direction of transcription. The 18S, 5.8S and 28S genes are shown in black. Approximate methylation levels across the repeat are plotted below. (B) A plot of %GC content across the repeat unit based on a window size of 1000 bp and a step of 30 bp. (C) A plot of the frequency of CpG across the repeat based on the observed frequency in a 1000 bp window divided by the expected frequency. The step is 30 bp.

Edinburgh). Restriction enzyme digests were performed according to the manufacturer’s instructions (New England Biolabs). As a control for complete digestion, an aliquot was removed and incubated with plasmid DNA (200 ng) that contained sites for the relevant enzyme. The plasmid and an undigested control were then visualised on an agarose gel stained with ethidium bromide to check for complete digestion (data not shown). DNA derived from sperm was extracted by addition of 50 mM EDTA pH 8.0, 1% SDS and 500 μg/ml Proteinase K with overnight incubation at room temperature. This was followed by addition of DTT to 50 mM and proteinase K to 100 μg/ml and further incubation overnight at 50°C (13). After phenol/chloroform extraction, the solution was ethanol-precipitated and resuspended in 10 mM Tris pH 7.5, 1 mM EDTA.

Southern blot analysis

In some cases DNA from human blood and sperm was first digested with EcoRV, which has no recognition site within the rDNA repeat, in order to reduce the molecular weight of bulk DNA. The DNA was then redigested with Xmal or its methylation-sensitive isoschizomer Smal (14). Where DNA was first digested with EcoRI or HindIII followed by digestion with methylation-sensitive restriction enzymes, EcoRV was not used. Digested DNA (5 μg/lane) was separated by electrophoresis in 0.8% or 0.5% agarose gels and blotted onto Hybond N+ filters (Amersham) following the manufacturer’s protocol. Probes were labelled using the random priming method to incorporate [α-32P]dCTP (15). Hybridisation was performed overnight at
68°C in 0.5 M Na2HPO4 (pH 7.2), 7% SDS with 5% powdered milk added to reduce background signal. After probe p1A7 had been denatured, 30 μg Cot-1 DNA (Gibco BRL) was added, competition was necessary due to the presence of a 100 bp Alu consensus sequence within p1A7 (16). Filters were washed three times for 20 min in 0.1x SSC, 0.2% SDS at 68°C (12).

Probes

Base pair co-ordinates used in this study correspond to those published in the GenBank database for the human ribosomal DNA complete repeating unit (accession number U13369). The coordinates of restriction sites shown in the figures were determined using either the GeneJockey III (Biosoft Cambridge) or DNAsStar (DNAsStar Incorporated) programs. Locations of restriction sites agree with previously published studies (17,18). Probes pHSrDNA5.9 and pHsrrDNA7.1 comprise EcoRI fragments from coordinates 1 to 5900 bp and from 5900 to 13 000 bp, respectively, cloned into pUC9 (a gift from Rakesh Anand). Probes p1A7 and p1B12 contain MseI fragments from coordinates 17 062 to 17 369 bp and from 35 695 to 36 231 bp cloned into pGem T-vectors (Promega). Probes p1B12 and p1A7 were both obtained during the construction of a library of densely methylated sequences from human blood DNA (G.J.R.Brock, unpublished). Both sequences were used as queries in BLAST (19) searches of the databases maintained by the NCBI Bethesda.

ACKNOWLEDGEMENTS

We thank Vicky Clark for DNA sequencing, Martin Simmen for computing assistance, Joan Davidson and Aileen Greig for technical help and Christine Struthers for secretarial support. We also thank Donald MacLeod, Susan Tweedie, Xinsheng Nan and Sally Cross for comments on the manuscript. GJRB was supported by a studentship from the Wellcome Trust. The work was supported by the Welcome Trust and the Howard Hughes Medical Institute.

REFERENCES