Volume 8 Number 20 1980 Nucleic Acids Research

A novel type of secondary modification of two CCGG residues in the human γδβ-globin gene locus

L.H.T.van der Ploeg*, J.Groffen+ and R.A.Flavell+*

*Section for Medical Enzymology and Molecular Biology, Lab. Biochemistry, University of Amsterdam, Jan Swammerdam Inst., 1005 GA Amsterdam, Netherlands, and +Lab. Gene Structure and Expression, National Inst. Medical Research, Mill Hill, London NW7 1AA, UK

Received 8 September 1980

ABSTRACT

The majority of the CCGG residues in the human γδβ-globin gene locus are cleaved by Msp I, irrespective of the tissue of origin of the DNA, although these sites show differential sensitivity to Hap II as a result of methylation of the internal C residue of the cleavage site (ref. 6). Two CCGG sites, at homologous positions 54 nucleotides in front of the Gγ- and Aγ-globin genes respectively, are not cleaved by Msp I in DNA from several human tissues, although DNA from placenta, foetal liver and from some established cell lines is cut at these sites. We have cloned the Aγ-globin gene from foetal blood DNA where the relevant CCGG site is not cut by Msp I. After cloning, the CCGG site can be cut by Msp I. The failure to cleave at this CCGG site in foetal blood DNA therefore, is not the result of a change in the DNA sequence of the cleavage site. Most likely the external C residue and perhaps both C residues are blocked by methylation at these two specific sites.

INTRODUCTION

Since the introduction of restriction endonucleases to the study of DNA methylation in eukaryotes[1], considerable attention has been focused on specific modification in the genomes of higher organisms and the role of DNA modification in gene expression[2-6]. Until now, attention has been paid to modifications of the cytosine residue of the CpG sequence present in the recognition sequence of several restriction endonucleases. Early work has shown, however, that 5 methyl cytosine (^{m}C) is also found at positions other than next to G residues: in particular, ^{m}CC residues were found in both sea urchin and mouse DNAs (see e.g. refs. 7 and 8).

The finding that the restriction endonuclease Msp I cleaves CCGG and $C^{m}CGG$ residues, whereas Hpa II or Hap II can only cleave the unmethylated CCGG, has greatly facilitated the study of DNA methylation in eukaryotes[9-11]. We now show that two specific CCGG residues in the human γ-globin gene locus are not cleaved by Msp I in DNA from certain human cell-types, whereas the CCGG residues flanking these sites are efficiently cleaved in DNA from all

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

tissues that we have examined. The failure of Msp I to cleave at these sites must be due to secondary modification of the DNA since cloned DNA can be cut at this site.

MATERIALS AND METHODS

Procedures for the isolation of DNA and cleavage with restriction endonucleases are as described[12, 13]. DNA digests were electrophoresed on 0.7% agarose gels for 24-48 h at 1v/cm and the DNA was then blotted onto nitrocellulose filters according to Southern[14]. The filter strips were hybridized to ^{32}P-labelled pHβG1 or pHγG1 DNAs as described previously[13] with the following modifications:

(i) The Hha I fragments which contain the γ or β-globin cDNA were prepared from pHβG1[13] and pHγG1 DNA[15] by Hha I-cleavage and electrophoresis on a 1% agarose gel. The fragments were recovered by the method of Tabak and Flavell[16]. The γ or β-globin cDNA Hha I fragment was labelled *in vitro* by nick-translation to a specific activity of 0.5-4 x 10^8 cpm/μg and used in filter-hybridizations at a final concentration of 5ng/ml.

(ii) The dextran sulphate procedure of Wahl *et al.* was used[17]. 1g of dextran sulphate (Sigma, 500,000 molecular weight) was dissolved in 9 ml. of 3 x SSC, 10 x Denhardts solution containing salmon sperm DNA and poly(A) as described in Jeffreys and Flavell[12]. Hybridization was for 10-12 h at 65° and post hybridization washes were as described[13]. It should be noted that while the incorporation of dextran sulphate in the hybridization increases the signal obtained, heterologous hybrids (such as β-γ globin gene DNA-DNA hybrids) are detected with this method and cannot be entirely eliminated by stringent washes.

RESULTS AND DISCUSSION

Mapping Msp I sites in the human γδβ-globin gene locus

We have mapped several Msp I sites around the human γ, δ and β globin genes in DNA from human placenta, from KB cells, or from peripheral blood (i.e. lymphocytes). DNA was cleaved with Msp I, either alone, or in combination with other restriction endonucleases, and the γ, δ and β-globin gene fragments detected by blot-hybridizations[12, 13]. The map described is shown in Fig. 4 and the Msp I sites are numbered Ms_1 etc. according to their position in the locus. This nomenclature will be used in the rest of this paper.

Mapping Msp I sites around the δ and β-globin genes

Cleavage of human DNA with Msp I generates a single 10.5 kb band

which contains both the δ- and β-globin genes (Fig. 1). In an Msp I + Bgl II double digest, Msp I cleaves the 8.4 kb Bgl II-δ fragment at Ms_6 to give the 7.0 kb double digest fragment, and cuts the 5.0 kb Bgl II-β fragment at Ms_7 to give a 3.4 kb fragment. Of the possible orientations for the Msp I sites in the respective Bgl II fragments, only one possible combination can generate the 10.5 kb Msp I-δβ fragment. This is confirmed by the following (Fig. 1):

- the 3.6 kb 3' Eco RI-β-globin gene fragment is cut by Msp I at Ms_7

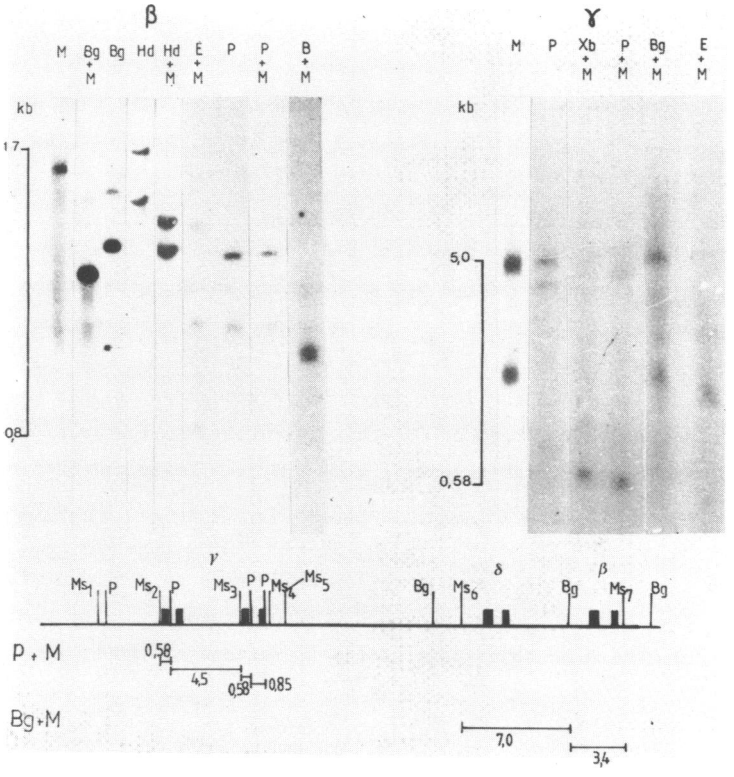

Fig. 1. Mapping Msp I sites around the δ, β and γ-globin genes in human placenta.

Human placental DNA was cleaved with the combination of restriction endonucleases indicated, and the samples electrophoresed and blotted onto nitrocellulose filters. The δ + β globin gene fragments were detected by hybridizing the filters with ^{32}P-labelled pHβG1 DNA and the γ-globin gene fragments were detected with pHγG1 DNA. The enzymes used were Msp I (M), Bal II (Bg), Hind III (Hd), Eco RI (E), Pst I (P), Bam HI (B) and Xba I (Xb). The simplified map under the figure explains the bands obtained and the Msp I sites are labelled (Ms_1 etc.) according to Fig. 4.
*Denotes a blocked Msp I site and —— a restriction enzyme cut.

to yield a 0.8 kb double digest fragment.

- the 7.6 kb Hind III-β fragment is cut by Msp I at Ms_7 to give a 4.8 kb fragment and the 17.0 kb-δ fragment is cut to give a 6.1 kb double digest fragment.
- neither the 4.4 kb Pst-β globin-gene fragment, nor the 2.3 kb Pst-I-δ fragment are cut by Msp I.
- the 15 kb 5' Bam HI-δ and the 8.6 kb 3' Bam HI-β fragments are cut by Msp I to give poorly resolved bands of 2.0 kb (cleavage at Ms_6) and 1.7 kb (cleavage at Ms_7) respectively; the uncut 5' 1.8 kb Bam HI-β and 4.8 kb 3' Bam HI-δ fragments are also visible (Fig. 1).

The cleavage pattern for Msp I around the β- and δ-globin genes is the same in DNA isolated from all human tissues that we have examined, i.e. KB cells, blood, sperm, placenta (Fig. 2); and brain, bone marrow, foetal liver, adult liver and the cultured cells, K562 and hela (not shown).

Mapping Msp I sites around the γ-globin genes in KB or placental DNA

Msp I-cut KB cell or placental DNA contains 2 fragments that hybridize with γ-globin gene probes, a 5.0 kb fragment which contains the Gγ-globin gene and a 1.8 kb fragment which contains the Aγ-globin gene. The Msp I sites have been mapped in the following way (Fig. 1):

- Pst I cleaves in the large intervening sequences of both γ-globin genes to generate a 5 kb fragment[15]; this is cut by Msp I at a site (Ms_3) just in front of the Aγ-globin gene to give a 4.5 kb fragment containing the Gγ gene and a 0.58 kb fragment containing the 5' regions of the Aγ-globin gene. The 4.1 kb Pst I-fragment which contains the 5' regions of the Gγ-globin gene[15] is cut at Ms_2 (at a position homologous to that described above) in front of the Gγ gene to give a second 0.58 kb fragment. The 0.85 kb Pst I-A-γ fragment (which contains part of the 3' regions of the Aγ gene) is uncut by Msp I.
- Xba I also cuts in the large intervening sequence of both Gγ and Aγ-genes and it gives 3 γ-globin gene fragments[15]: a 3.7 kb fragment containing the 5' segment of the Gγ-globin gene which is cut by Msp I at Ms_2 to give a 0.6 kb fragment; and a 5.0 kb fragment, containing the intergenic region and part of the Gγ and part of the Aγ globin genes, which is cut in front of the Aγ gene at Ms_3 by Msp I to give a 4.5 kb fragment and a 0.6 kb fragment. The third Xba I fragment contains the 3' regions of the Aγ-globin gene[15] and is cut by Msp I to give a 1.2 kb fragment; this indicates the presence of an Msp I site (Ms_4) to the 3' side of the Aγ globin gene.

The above data are explained by the positions of the Msp I sites

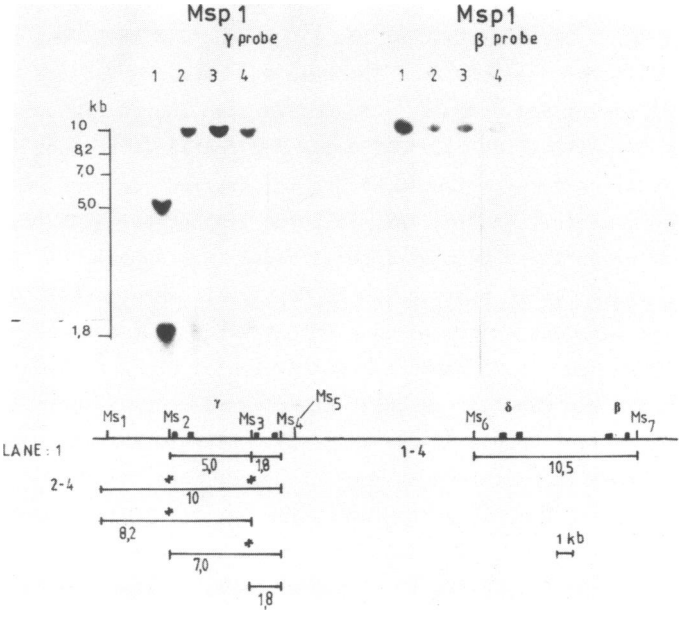

Fig. 2. Tissue-specific differences in Msp I digestion patterns around the γ-globin but not the β-globin genes.

DNA from various human tissues was cleaved with Msp I and then electrophoresed on a 0.7% agarose gel. The DNA was then blotted onto a nitrocellulose filter and hybridized to either ^{32}P-labelled pHγG1 DNA (γ-probe) or ^{32}P-labelled pHβG1 DNA (β probe): post hybridization washes and autoradiography were as described in experimental procedures. The tissues used were: 1. KB cells; 2. sperm; 3. and 4. foetal umbilical cord blood.

The simplified map under the figure explains the bands obtained and the Msp I sites are labelled (Ms_1 etc.) according to Fig. 4.
*Denotes a blocked Map I site and —— denotes a restriction enzyme cut.

Ms_2 and Ms_3, which give the 5 kb Msp I fragment, and Ms_3 and Ms_4, which give the 1.8 kb Msp I fragment (Fig. 4). This is confirmed by Eco RI + Msp I double digests (Fig. 1): the 5' Eco RI fragments of the Gγ and Aγ genes are both cleaved to give a 1.4 kb double band and the 3' regions of the Gγ gene are present as the 1.5 kb Eco RI fragment (the 0.6 kb 3' Aγ Eco RI fragment[15] is not visible in this picture). Recently the DNA sequence of the human Gγ and Aγ genes has been determined[18]. Sites Ms_2 and Ms_3 are localized on this sequence at positions -54 in both genes, where +1 is the first nucleotide of the Gγ and Aγ globin genes.

Sites Ms_2 and Ms_3 are not cut by Msp I in blood DNA

The Msp I pattern for blood DNA around the δ + β genes was the same

as shown above for placental DNA, but the pattern for the γ-genes was entirely different (Fig. 2). Instead of the 5.0 kb and 1.8 kb bands, a single major band of 10 kb was visible. The same was seen for 9 blood DNA preparations from different individuals. Sperm DNA also showed the same pattern, but DNA from KB cells (Fig. 2), K562 cells and foetal liver (not shown) showed the placental pattern. The 10 kb Msp I-γ fragment is also found in DNA from human bone marrow, adult liver, cultured lymphocytes (EBV transformed) and granulocytes purified from peripheral blood (not shown).

It seemed likely that the 10 kb Msp I-γ fragment resulted from the failure of Msp I to cut at one or more of the sites that are cleaved to generate the 5.0 kb and 1.8 kb bands seen in placental DNA and KB cell DNA. We therefore mapped the Msp I sites which are cleaved to give the 'blood'-type 10 kb band in various double digest (Fig. 3). The 10 kb band containing the two γ-globin genes is explained by the failure of Msp I to cleave at the sites Ms_2 and Ms_3 of Fig. 1 as diagrammed in Fig. 2; instead, the 10 kb band is generated by cleavage at Ms_1 and Ms_4. This is shown by double digestion of sperm DNA with Msp I and Eco RI or Bam HI respectively (Fig. 3). The 5' Eco RI-Gγ fragment of 6.5 kb is cleaved to give a 5 kb band, not the 1.4 kb band seen for placental or KB cell DNA (cf. Fig. 1); similarly, both the 2.7 kb Bam HI fragment containing the 5' regions of the Gγ gene and the 5 kb fragment Bam HI, which contains segments of both γ-genes, are uncleaved by Msp I (Fig. 3). Hence, sites Ms_2 and Ms_3 are both blocked in sperm DNA. In contrast, site Ms_4 is cleaved by Msp I in sperm and blood DNA (Fig. 3), as is found for KB cell or placental DNA (Fig. 1). Thus, the 15 kb Bam HI fragment which contains the 3' regions of the Aγ globin gene is cut by Msp I to give the 1.3 kb double digest fragment characteristic for cleavage at site Ms_4 (see Fig. 2); and in the same way, Bgl II + Msp I double digests of 'blood' DNA show cleavage of the 14 kb Bgl II-γ fragment at site Ms_4 to give a characteristic 8.5 kb double digest band (Fig. 3).

The modification at Ms_2 and Ms_3 is not 100% in these DNAs. In the blood and sperm DNA preparations additional minor bands of 8.5 kb, 7.0 kb and 1.8 kb are visible (Figs. 2 and 5). These bands are the result of partial cleavage by Msp I at sites Ms_2 and Ms_3, as judged by their sizes and the products seen in double digests. The 1.8 kb fragment contains the Aγ-globin gene and is the same as the major component seen in KB and placental DNA; it derives from chromosomes where Ms_3 is unmodified. The Gγ-globin gene fragment from the same chromosome is 8.2 kb and it is the result of cleavage at Ms_1 and Ms_3. The 7 kb fragment derives from chromosomes where Ms_3, but not

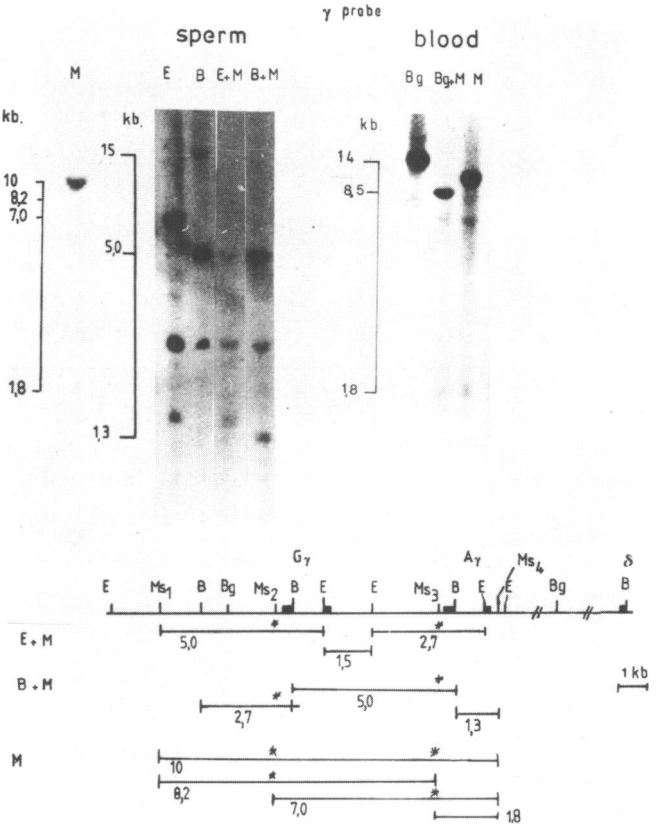

Fig. 3. Mapping the Msp I sites in human sperm and blood DNA.

Human sperm or blood DNA was digested twice to completion with Msp I and then digested with Eco RI, Bam HI or Bgl II. The DNA samples were then electrophoresed, blotted and the filters hybridized with pHγG1 DNA (γ-probe) as described in Methods. The maps shown under the figure gives the interpretation of the results.
*Denotes a blocked Msp I site.

Ms_2, is modified; here, the 5.0 kb and 1.8 kb fragments are linked to form the 7 kb segment. The partial cleavage at sites Ms_2 and Ms_3 shows that these sites are indeed present in DNA from tissues like blood and sperm.

Site Ms_3 is cleaved by Msp I in the Aγ-globin gene cloned from blood DNA

The failure of Msp I to cleave at site Ms_2 and Ms_3 in blood (lymphocyte) DNA can be explained by secondary modification as considered above, or by somatic mutation at these sites which causes the sites to be lost. If the latter were the case, the γ-globin genes cloned from blood DNA should

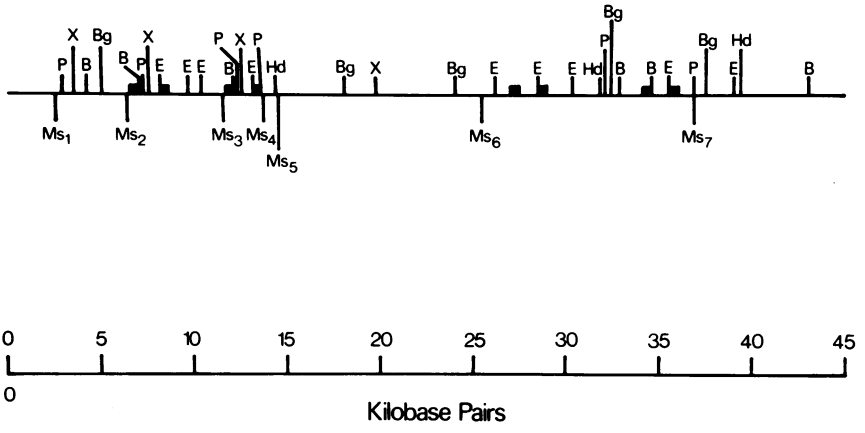

Fig. 4. A physical map of Msp I sites around the γ, δ and β-globin genes.

The Msp I sites have been mapped relative to the known restriction sites around these genes. The genes are shown as filled boxes and are (from left to right) Gγ, Aγ, δ and β. The restriction enzymes used were Bam HI (B), Bgl II (Bg), Eco RI (E), Hind III (Hd), Msp I (Ms), Pst I (P), Xba I (X). The Msp I sites are shown below the line and have been mapped in this study. The numbers (Ms_1 etc.) refer to the position within the locus.

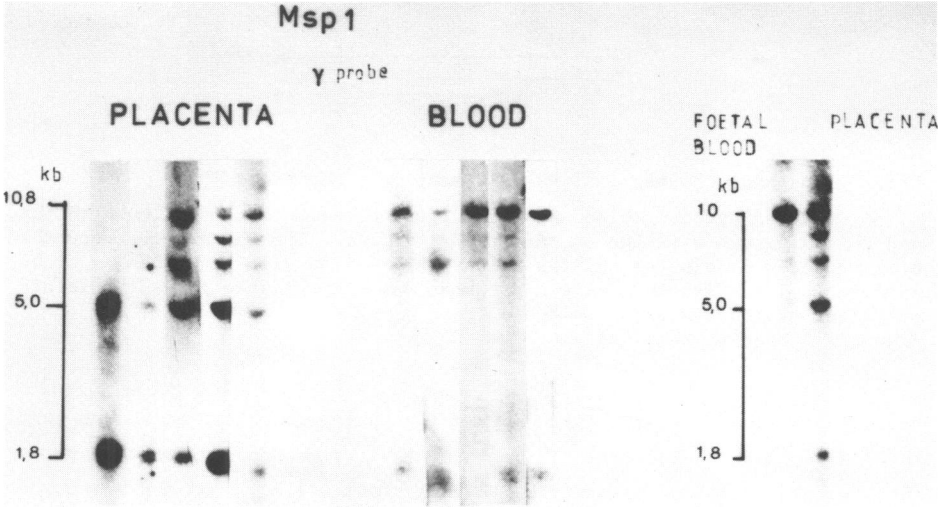

Fig. 5. Msp I patterns of blood and placenta DNAs from different individuals and the Msp I pattern of blood and placenta from the same foetus.

Several placental or foetal umbilical-cord blood DNA preparations were digested with Msp I and analysed for the γ-globin DNA bands as described in Methods.

lack these two Msp I sites. Alternatively, if the sites were simply blocked by secondary modification, these modifications would be lost during propagation in bacteria and hence the sites would be detectable after cloning.

To test this possibility, we isolated the γ-globin gene region from a phage library made from human foetal blood DNA (the isolation of these recombinants will be described elsewhere). Several independent recombinant clones were isolated which contain the region from the Aγ gene to the δ-globin gene. The break point on the 5' side of these clones was between the Gγ and Aγ genes and the clones should therefore contain site Ms_3.

To test this we digested the phage DNA with Msp I or Eco RI. As a control a cosmid containing the Gγ Aγ δ and β globin genes isolated from a human placental DNA library (F. G. Grosveld, H. H. Dahl, E. de Boer and RAF, unpublished), was also analysed in the same way. These DNAs were electrophoresed in a 1% agarose gel, transferred to a nitrocellulose filter and hybridized with a probe for the γ-globin genes (Fig. 6). In both the phage and cosmid DNAs, the 1.8 kb Msp I fragment containing the Aγ-globin gene and which is produced by cleavage at Ms_3 and Ms_4, is present. The same result was obtained for another two γ-globin gene clones. We conclude that site Ms_3 is present, but in a modified form in blood (lymphocyte) DNA.

Several trivial artefacts might formally explain the failure of Msp I to cut at Ms_2 and Ms_3 and we have, therefore, included control experiments to eliminate these uncertainties.

1. The result could be caused by partial digestion with the enzyme Msp I. This is excluded by the fact that the DNA was digested twice with a 50 fold excess of Msp I enzyme. Between digests, the DNA was deproteinized and recovered by ethanol precipitation. Moreover, two additional controls show that Msp I digestion was complete. In incubations with Msp I, internal phage λ DNA is added as a control and is fully digested in both cases. More significantly, when the same Msp I digested DNA is hybridized with β-gene probes, the 10.5 kb Msp I limit-digest product is seen in all cases.

2. The result could be caused by extensive polymorphism in the human population such that some people show the 'blood' pattern, others the placental pattern and others a heterozygous pattern. This is excluded:

a) by the analysis of a large number of DNA samples from different individuals. Blood DNA always shows the blood pattern and placental DNA shows mainly the placental pattern with variable amounts of the blood pattern (which may be caused in part by contaminating blood superposed on this (Fig. 5).

b) By the fact that umbilical cord blood and placental DNA from the

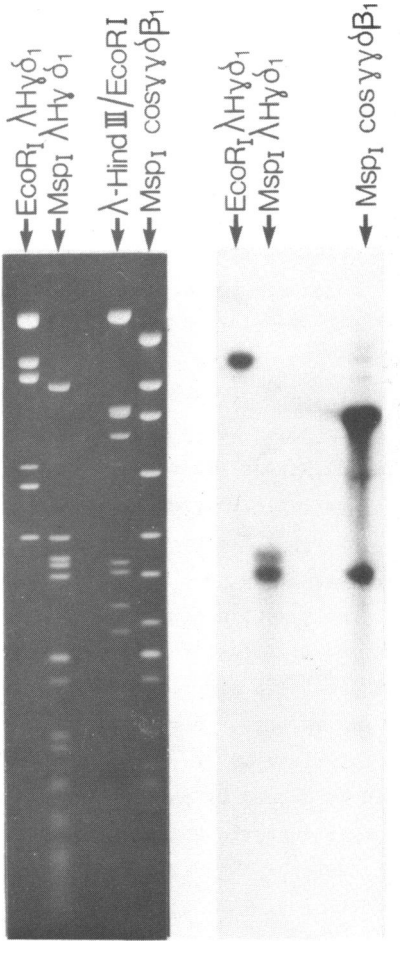

Fig. 6. Cleavage at site Ms_3 in an A-γ globin gene cloned in phage λ from blood DNA.

The phage DNA (λ Hγδ1) or cosmid DNA (cos γγδβ1 cloned from placental DNA was cleaved with Eco RI or Msp I and electrophoresed on a 1% agarose gel. The gel was blotted onto a nitrocellulose filter and hybridized with a 5 kb Xho I fragment (isolated from the cosmid) which runs from the large intron of the Gγ globin gene to the corresponding site in the Aγ globin gene. The cosmid shows up the 5 kb and 1.8 kb Msp I fragments running from Ms_2 to Ms_3 and Ms_3 to Ms_4. The phage shows the 1.8 kb fragment from Ms_3 to Ms_4, together with a junction fragment of about 2 kb which contains the phage vector DNA attached to the 5' breakpoint of the human DNA.

same infant show the respective tissue-specific patterns (Fig. 5), not the same pattern as would be expected if the differences are caused by polymorphism.

c) By the fact that DNA cloned from a blood DNA sample blocked at site Ms_3 can be cleaved by Msp I at Ms_3.

Sites Ms_2 and Ms_3 are modified in DNA from several human tissues. Since Msp I is not blocked by methylation of the internal C residues (refs. 9, 10, 11), it seems likely that the modification responsible is methylation of the external C, perhaps in conjunction with modification of the internal C residue. This is supported by the fact that ^{m}CC pyrimidine tracts have been found in mammalian DNA [7, 8]. This type of modification is described here for the first time and it clearly occurs at specific locations in the genome. We have recently observed similar modifications in the vicinity of the rabbit β-globin gene (unpublished results).

It is obviously of interest that this type of modification is found at specific sites and that tissue-specific differences occur. Elsewhere, we have shown that these sites are unmodified in DNA from K562 cells and that they exhibit a low level of modification in foetal liver DNA (ref. 6); in both these cell types the γ-globin genes are expressed. The fact that sites Ms_2 and Ms_3 are located in the putative γ-globin gene promotor region is of interest with regard to the function of this type of modification, although it should be noted that they are also unmodified in cell types where γ-globin gene expression would not be expedited. This subject is discussed in depth elsewhere[6].

ACKNOWLEDGEMENTS

We are grateful to Adrian Bird for providing Moraxella species, Piet Borst for discussions, Christine Shewmaker for comments on the manuscript and Cora O'Carroll for preparing the manuscript. This work was supported in part by a grant to R.A.F. from The Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization for the Advancement of Pure Research (ZWO) and by the British Medical Research Council.

REFERENCES

1. *Bird, A. and Southern, E. (1978). J. Mol. Biol. 118, 27-48.*
2. *Waalwijk, C. and Flavell, R. A. (1978). Nucl. Acids Res. 5, 4631-4641.*
3. *McGhee, J. D. and Ginder, G. D. (1979). Nature 280, 419-420.*

4. Desrosiers, R. C., Mulder, C. and Fleckenstein, B. (1979). Proc. Natl. Acad. Sci. U.S.A. 76, 2125-2132.
5. Mandel, J. L. and Chambon, P. (1979). Nucl. Acids. Res. 7, 2081-2103.
6. van der Ploeg, L. H. T. and Flavell, R. A. (1980). Cell, 19, 947-958.
7. Grippo, P., Iaccarino, M., Parisi, E. and Scarano, E. (1968). J. Mol. Biol. 36, 195-208.
8. Solomon, R. and Kaye, A. M. (1970). Biochem. Biophysica Acta. 204, 340-351.
9. Waalwijk, C. and Flavell, R. A. (1978). Nucl. Acids Res. 5, 3231-3236.
10. Singer, J., Roberts, E. M. S. and Riggs, A. D. (1979). Science 203, 1019-1021.
11. Cedar, H., Solage, A., Glaser, G. and Razin, A. (1979). Nucl. Acids Res. 6, 2125-2132.
12. Jeffreys, A. J. and Flavell, R. A. (1977). Cell 12, 429-439.
13. Flavell, R. A., Kooter, J. M., de Boer, E., Little, P. F. R. and Williamson, R. (1978). Cell 15, 25-41.
14. Southern, E. M. (1975). J. Mol. Biol. 98, 503-517.
15. Little, P. F. R., Flavell, R. A., Kooter, J. M., Annison, G. and Williamson, R. (1979). Nature 278, 227-231.
16. Tabak, H. F. and Flavell, R. A. (1978). Nucl. Acids Res. 5, 2321-2332.
17. Wahl, G. M., Stern, M. and Stark, G. (1979). Proc. Natl. Acad. Sci. U.S.A. 76, 3683-3687.
18. Slighton, J., Blechl, A. and Smithies, O. (1980). Cell, in press.