Polymorphism of DNA sequence in the human pro α2(I) collagen gene

A F GROBLER-RABIE, D K BREBNER, S VANDENPLAS, G WALLIS*, R DALGLEISH†, R E KAUFMAN‡, A J BESTER, C G P MATHEW, AND C D BOYD

*From the MRC Molecular and Cellular Cardiology Research Unit, University of Stellenbosch Medical School, PO Box 63, Tygerberg 7505, South Africa; *the MRC Research Unit for Inherited Skeletal Disorders, Department of Human Genetics, University of Cape Town, Observatory 7925, South Africa; †the Pulmonary Branch, National Heart, Lung and Blood Institute, Bethesda, Maryland, 20205, USA; and ‡the Department of Medicine, Duke University Medical Center, Durham, North Carolina 27710, USA.*

SUMMARY The human pro α2(I) collagen gene was analysed for the presence of restriction fragment length polymorphisms. DNA from randomly selected unrelated persons of three Southern African populations was cleaved with one of eight different restriction enzymes, electrophoresed, blotted, and hybridised with cDNA and genomic probes specific for the pro α2(I) gene. An MspI polymorphism was detected which results from the loss of a cleavage site within the 3' half of the gene. In two of the populations studied, the polymorphism occurred at significant frequencies, and should therefore prove useful as a genetic marker for the study of inherited disorders of connective tissue involving collagen structure or biosynthesis.

The collagens are a family of closely related proteins which constitute a major component of the extracellular matrix of nearly every organ and tissue in vertebrates. At least eight structurally distinct collagen types have been characterised to date, some of which are composed of more than one type of chain. Thus, a multigene family of a minimum of 13 genes is necessary to code for their constituent chains.[1-4]

The modulated expression of these different gene loci appears to be critically important for normal embryonic development, as the correctly programmed spatial and temporal production of collagen is required for differentiation and tissue repair processes. Aberrations in collagen gene expression are the cause of at least three groups of inherited disorders in man: osteogenesis imperfecta (OI), Marfan's syndrome, and Ehlers-Danlos syndrome.[5-7] Some of these defects result from structural alterations in the collagen genes, since shortened pro α1(I) or pro α2(I) chains have been found in a number of cases of osteogenesis imperfecta.[8-10] Recently this has been verified by direct analysis of the gene, since a 300 bp

deletion was detected in the pro α1(I) gene of a case of lethal OI.[11] A deletion within an α1(I)-like collagen sequence may also be associated with OI and Ehlers-Danlos syndrome type II.[12]

The isolation and characterisation of human type I collagen genes[13-15] has led to a detailed understanding of their structure. The pro α2(I) collagen gene is represented only once in the human genome[16] and has been mapped to chromosome 7.[17 18] It is about 38 kb in length and contains at least 50 intervening sequences.[14] The detailed structural information available opens up the possibility of studying variations in the structure of the pro α2(I) gene. Sequence specific probes can be used to detect variations either related to inherited disorders of collagen gene expression, or which produce restriction fragment length polymorphisms (RFLP) within and around the pro α2(I) collagen gene. The RFLPs could then serve as genetically linked markers to trace the inheritance of structural alterations of the pro α2(I) gene.[19 20] The relevance of this approach has been amply demonstrated by the use of RFLPs for the antenatal diagnosis of sickle cell anaemia and thalassaemia.[21]

In this paper, we report the existence of a restriction fragment length polymorphism of relatively high frequency, detectable by the restriction

enzyme MspI, in apparently normal subjects of South African populations.

Materials and methods

High molecular weight human DNA was prepared from whole blood or from white blood cells of randomly selected subjects of three South African populations, Afrikaners, so-called 'Cape Coloureds', and Blacks,[22] as described by Vandenplas *et al.*[23]

Restriction enzymes were obtained from Bethesda Research Laboratories, Boehringer-Mannheim, and New England Biolabs, and used as directed by the suppliers. Typically, 10 μg of human DNA was digested with 30 units of restriction enzyme overnight and the DNA fragments separated by electrophoresis as described.[23] DNA from λ phage digested with HindIII and labelled with [α–^{32}P]dCTP using T4 DNA polymerase[24] was used as a molecular weight marker.

DNA was transferred from the gel to nitrocellulose as described by Southern.[25] Human pro α2(I) collagen cDNA[13] and genomic probes[16] were nick translated to a specific activity of at least 1×10^8 cpm/μg using the nick translation kit and protocol supplied by Bethesda Research Laboratories. The procedure described by Jeffreys and Flavell[26] was followed for hybridisation of the DNA on the filters to the labelled probes, with final stringency washes of 0·1×SSC for genomic and 0·5×SSC for cDNA probes, at 65°C (1×SSC buffer contains 0·15 mol/l NaCl, 0·015 mol/l sodium citrate, pH 7·0). Autoradiography was for one to seven days at −80°C, using Cronex Safety or Kodak X-Omat AR films, backed by Fuji Mach 2 calcium tungstate intensifying screens.

Results and discussion

POLYMORPHISM IN THE HUMAN PRO α2(I) COLLAGEN GENE

A study of polymorphism in the human pro α2(I) collagen gene of a South African population, the Afrikaners,[22] was undertaken. DNA was isolated from randomly selected, apparently normal subjects and restricted with eight different enzymes (EcoRI, HindIII, PstI, XbaI, TaqI, BamHI, SmaI, and MspI). The probe used to detect pro α2(I) sequences was a recombinant plasmid, Hf-32, which contains DNA sequences complementary to the 3′ half of the human pro α2(I) collagen gene,[13] including 1443 nucleotides coding for the pro α2(I) triple helical domain and 597 nucleotides coding for the C propeptides.[27] Only with the restriction enzyme MspI could a polymorphic site be detected
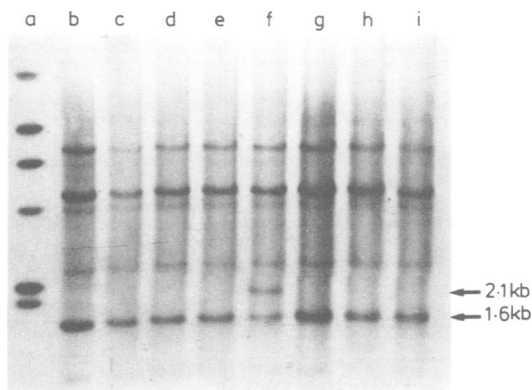

FIG 1 *DNA samples were digested with MspI, blotted, and hybridised to Hf-32,[13] a human pro α2(I) procollagen cDNA probe. Lane a is λ DNA, digested with HindIII. Lane f shows the variant 2·1 kb fragment.*

at a frequency high enough to warrant further investigation.

MOLECULAR BASIS OF THE MspI POLYMORPHISM

Restriction with MspI normally gives rise to six fragments when probed with Hf-32 (fig 1). However, in some subjects, an additional fragment of 2·1 kb was detected (fig 1, lane f). This fragment did not result from contamination of human DNA samples with pBR322-like sequences, because it was found in the DNA prepared from repeated blood samples and was not detected when blots were probed with nick translated pBR322.

An RFLP can result from the creation or elimination of a restriction site by base pair substitution or modification, or from rearrangements of DNA segments by insertions, deletions, or inversions.[19] The MspI polymorphism does not appear to be caused by gross rearrangement since restriction analysis of DNA from persons with the variant using other enzymes did not reveal any abnormal fragments (data not shown).

The presence of the 2·1 kb fragment could not be correlated with the loss of any restriction fragments normally found after digestion with MspI. However, if a person was heterozygous for the polymorphism, we would expect to see a difference in the intensity of hybridisation of at least one of the other fragments. Scans of the autoradiograph of normal and variant MspI patterns are shown in fig 2. The appearance of the 2·1 kb variant fragment was always accompanied by a decrease in the intensity of the 1·6 kb band. However, the reduction in intensity
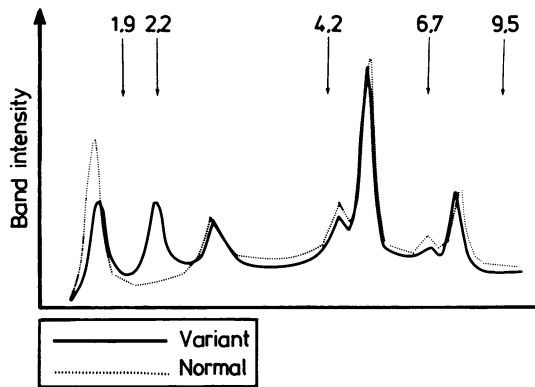
FIG 2 *Scan of normal and variant MspI patterns from the autoradiograph in fig 1. The position of the λ HindIII size markers are indicated by vertical arrows.*

MspI fragment when screened with the 4·0 kb subclone. Three genotypes were detected (table) using the 4·0 kb subclone as a probe: 1·6 kb (+/+), 1·6/2·1 kb (+/−), and 2·1 kb (−/−).

A series of double digests was carried out with MspI and EcoRI, HindIII, or XbaI, in order to map the polymorphism more exactly. The results are summarised in fig 3b, which shows that the 2·1 kb variant is caused by the loss of an MspI site at the 5' end of the 1·6 kb fragment. This site is about 4 kb from the 3' end of the gene, and is located near the 3' end of intron 6.[15] The loss of an MspI site may be caused by a base substitution within the recognition sequence of the enzyme (5'–CCGG–3') or by selective modification at the external cytosine of this site. However, the pro α2(I)-MspI pattern of DNA isolated from fibroblasts (which produce type I

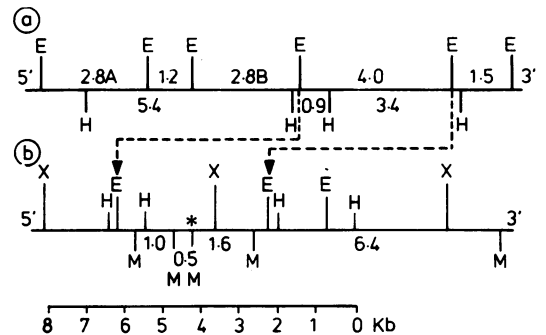was only about 25%, instead of the 50% expected for heterozygotes. This result could be explained by the presence of two 1·6 kb MspI fragments originating from different regions of the pro α2(I) gene. In order to test this hypothesis, and to localise further the variant fragment, we hybridised blots with a series of genomic subclones[16 28] corresponding to areas of the gene covered by Hf-32. The location of these subclones is shown in fig 3a. The four subclones used (1·2 kb, 1·5 kb, 2·8A kb, and 4·0 kb) are derived from EcoRI fragments of these sizes in the 3' half of the pro α2(I) gene. The 1·6 kb MspI band was detected with the 1·2 kb, 2·8A kb, and 4·0 kb subclones, whereas the 2·1 kb variant band hybridised to the 4·0 kb subclone only (data not shown). The results indicate that a 1·6 kb MspI fragment originates from sequences spanning the adjacent 1·2 and 2·8A kb subclones, and another 1·6 kb MspI fragment is located within the region of the gene encoded by the 4·0 kb subclone. The homozygote for the polymorphism lacks the 1·6 kb

FIG 3 *(a) Restriction map of the genomic subclones of the 3' half of the human pro α2(I) procollagen gene. The 5' end of the 2·8A kb EcoRI subclone is about 14 kb from the 3' end of the gene. (b) Location of the polymorphic MspI site in the 4·0 kb subclone. The 4 kb EcoRI subclone corresponds to 3 kb of sequence coding for the α chain domain and 1 kb coding for the C propeptide.[14 27] Restriction sites: E=EcoRI; H=HindIII; X=XbaI; M=MspI.*

TABLE *Frequency study of the MspI RFLP in South African populations.*

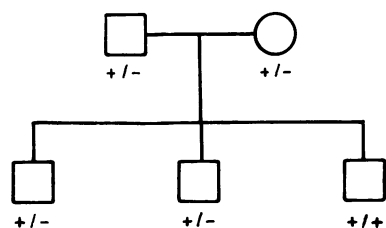| Genotypes	Afrikaners		'Cape Coloureds'		South African Blacks	
	Observed	Expected	Observed	Expected	Observed	Expected
+/+	22	22·9	22	22·5	28	27
+/−	11	9·2	8	6·9	1	2·85
−/−	0	0·9	0	0·5	1	0·08
n	33		30		30	
Frequency of minor allele (−)	0·17		0·13		0·05	

FIG 4 *Family study showing the inheritance of the 2·1 kb MspI variant. + indicates the presence and − the absence of the polymorphic MspI site.*

collagen) and white blood cells (which do not) is the same. It therefore seems unlikely that there should be differences in methylation at the polymorphic site in different persons.

INHERITANCE OF THE MspI RFLP
In order to establish the heritability of the RFLP, we examined DNA isolated from first degree relatives of a subject who showed the 2·1 kb fragment (fig 4). The family was screened with the 4·0 kb subclone, since homozygotes for the polymorphism would not be detected using Hf-32 as a probe. Both parents are heterozygous for the 2·1 kb MspI variant, as are two of the three sibs. The third child is homozygous for the normal restriction fragment pattern. These results are compatible with a simple Mendelian mode of inheritance.

FREQUENCY OF THE MspI POLYMORPHISM
The usefulness of a polymorphism as a genetic marker depends on its frequency in a population. An RFLP is considered 'useful' if the rarer allele has a frequency of 0·10 or greater.[19] The frequencies of the MspI polymorphism were determined in three local populations, the Afrikaners who originated from several regions of Western Europe, the so-called 'Cape Coloureds', and the South African Blacks,[22] using the 4·0 kb EcoRI subclone as a probe. The frequencies detected are presented in the table. In the Afrikaner and 'Coloured' samples, the frequency of the rare allele was 0·17 and 0·13 respectively, which suggests that it should prove a useful marker in these populations. However, the frequency of this polymorphism in the Blacks is substantially lower (0·05). The polymorphism may therefore be of European origin, and its relatively high frequency in the 'Cape Coloured' population may reflect the significant contribution made by settlers from Western Europe to the genetic composition of this group.[22]

The table also shows the observed and expected distribution of the RFLP in the three populations.

There was no significant difference between these values for any of the three groups (p=0·73, 0·61, and 0·47 for Afrikaners, 'Coloureds', and Blacks respectively).

Conclusion

The data presented here establish the existence of a restriction fragment length polymorphism of relatively high frequency in the 3′ half of the human pro α2(I) collagen gene of two South African populations. The MspI RFLP may provide a useful genetic marker for studying inherited disorders of connective tissue that are caused by abnormal collagen structure or biosynthesis.

This approach has been used by Tsipouras *et al*[29] who detected an EcoRI RFLP in the 5′ half of the pro α2(I) gene, and demonstrated linkage between the polymorphism and a dominant form of OI. The MspI and EcoRI RFLPs should complement each other in such studies, since one polymorphism may be informative in a family or parent in whom the other is not. The use of multiple polymorphisms in linkage studies has been clearly demonstrated by Boehm *et al*,[21] who were able to diagnose 86% of cases of sickle cell disease or β thalassaemia with RFLPs in the β globin gene region.

The human pro α2(I) collagen gene has recently been mapped to chromosome 7q22,[18] which is a region of the genome where few polymorphic markers have been observed. The MspI RFLP described in this report may therefore provide a useful additional marker for inherited disorders other than those involving collagen, should the defective gene be linked to this locus.

References

[1] Bornstein P, Sage H. Structural distinct collagen types. *Ann Rev Biochem* 1980;**49**:957–1003.
[2] Miller EJ, Gay S. Collagen: an overview. *Methods Enzymol* 1982;**82**:3–32.
[3] Furthmayr H, Wiedemann H, Timpl R, *et al*. Electron-microscopical approach to a structural model of intima collagen. *Biochem J* 1983;**211**:303–11.
[4] Sage H, Trüeb B, Bornstein P. Biosynthetic and structural properties of endothelial cell type VIII collagen. *J Biol Chem* 1983;**258**:13391–401.

186

5 Lapiere CM, Nusgens B. In: Ramachandran GN, Reddi AH, eds. *Biochemistry of collagen*. New York: Plenum Press, 1976:377–447.

6 Breitkreutz D, Diaz de Leon L, Daglia L *et al*. The synthesis of presumptive procollagen messenger ribonucleic acid in the calvaria of the developing chick. *Biochim Biophys Acta* 1978;**517**:349–59.

7 Prockop DJ, Kivirikko KI, Tuderman K, Guzman NA. The biosynthesis of collagen and its disorders. *N Engl J Med* 1979;**301**:77–85.

8 Williams CJ, Prockop DJ. Synthesis and processing of a type I procollagen containing shortened pro-α_1 (I) chains by fibroblasts from a patient with osteogenesis imperfecta. *J Biol Chem* 1983;**258**:5915–21.

9 De Wet WJ, Pihlajaniemi T, Myers J, *et al*. Synthesis of a shortened pro-α_2 (I) chain and decreased synthesis of pro-α_2 (I) chains in a proband with osteogenesis imperfecta. *J Biol Chem* 1983;**258**:7721–28.

10 Byers PH, Shapiro JR, Rowe DW, David KE. Abnormal α_2-chain in type I collagen from a patient with a form of osteogenesis imperfecta. *J Clin Invest* 1983;**71**:689–97.

11 Chu ML, Williams CJ, Gugliemina P, *et al*. Internal deletion in a collagen gene in a perinatal lethal form of osteogenesis imperfecta. *Nature* 1983;**304**:78–80.

12 Pope FM, Cheah KSE, Nicholls AC, *et al*. Lethal osteogenesis imperfecta congenita and a 300 base pair gene deletion for an α_1 (I)-like collagen. *Br Med J* 1984;**288**:431–4.

13 Myers JC, Chu ML, Faro SH, *et al*. Cloning a cDNA for the pro-α_2 chain of human type I collagen. *Proc Natl Acad Sci USA* 1981;**78**:3516–20.

14 Myers JC, Dickson LA, de Wet W, *et al*. Analysis of the 3' end of the human pro-α_2 (I) collagen gene. *J Biol Chem* 1983;**258**:10128–35.

15 Chu ML, de Wet WJ, Bernard MP, *et al*. Structure of the human pro-α_1 (I) collagen gene: evolutionary conservation of functional domains. *Nature* 1984;**310**:337–40.

16 Dalgleish R, Trapnell BC, Crystal RG, Tolstoshev P. Copy number of a human type I α_2 collagen gene. *J Biol Chem* 1982;**257**:13816–22.

17 Junien C, Weil D, Myers JC, *et al*. Assignment of the human pro-α_2 (I) collagen structural gene (COLIA 2) to chromosome 7 by molecular hybridization. *Am J Hum Genet* 1982;**34**:381–7.

18 Henderson AS, Myers JC, Ramirez F. Localization of the human α_2-(I) collagen gene (COLIA 2) to chromosome 7q22. *Cytogenet Cell Genet* 1983;**36**:586–7.

19 Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980;**32**:314–31.

20 Skolnick M, White R. Strategies for detecting and characterizing restriction fragment length polymorphism (RFLPs). *Cytogenet Cell Genet* 1982;**32**:58–67.

21 Boehm CD, Antonarakis SE, Phillips JA, *et al*. Prenatal diagnosis using DNA polymorphisms. Report on 95 pregnancies at risk for sickle-cell disease or β-thalassemia. *N Engl J Med* 1983;**308**:1054–8.

22 Botha MC, Pritchard J. Blood groups gene frequencies. An indication of the genetic constitution of population samples in Cape Town. *S Afr Med J* 1972;**suppl 1**:1–27.

23 Vandenplas S, Wiid I, Grobler-Rabie A, *et al*. Blot hybridisation analysis of genomic DNA. *J Med Genet* 1984;**21**:164–72.

24 Maniatis T, Fritsch EF, Sambrook J. *Molecular cloning. A laboratory manual*. New York: Cold Spring Harbor Laboratory, 1982:117.

25 Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 1975;**98**:503–18.

26 Jeffreys AJ, Flavell RA. A physical map of the DNA regions flanking the rabbit β-globin gene. *Cell* 1977;**12**:429–39.

27 Bernard MP, Myers JC, Chu ML, *et al*. Structure of a cDNA for the pro-α_2 chain of human type I procollagen. Comparison with chick cDNA for pro-α_2 (I) identifies structurally conserved features of the protein and the gene. *Biochemistry* 1983;**22**:1139–45.

28 Tajima S, Ting J, Pinnell S, Kaufman R. Isolation and characterisation of a human pro-α_2(I) collagen gene segment. *J Invest Dermatol* 1984;**82**:265–9.

29 Tsipouras P, Myers JC, Ramirez F, Prockop DJ. Restriction fragment length polymorphism associated with the pro-α_2 (I) gene of a family with an autosomal dominant form of osteogenesis imperfecta. *J Clin Invest* 1983;**72**:1262–7.

Correspondence and requests for reprints to Dr C Mathew, MRC Molecular and Cellular Cardiology Research Unit, University of Stellenbosch Medical School, PO Box 63, Tygerberg 7505, South Africa.