The EMBO Journal vol.4 no.7 pp.1745-1748, 1985

Detection of a high frequency *Rsa*I polymorphism in the human proα2(I) collagen gene which is linked to an autosomal dominant form of osteogenesis imperfecta

A.F. Grobler-Rabie, G. Wallis[1], D.K. Brebner, P. Beighton[1], A.J. Bester and C.G. Mathew

MRC Molecular and Cellular Cardiology Research Unit, University of Stellenbosch Medical School, Tygerberg 7505, and [1]MRC Research Unit for Inherited Skeletal Disorders, Department of Human Genetics, University of Cape Town, Observatory 7925, South Africa

Communicated by R. Williamson

Screening of the proα2(I) collagen genes of Southern African populations for restriction fragment length polymorphisms (RFLPs) has revealed a locus polymorphic for the restriction enzyme *Rsa*I. The frequency of the RFLP was 0.38 in Afrikaners, but much lower in indigenous Southern African populations, which suggests that it is of European origin. The polymorphism was used to study 19 affected and non-affected individuals in a four generation family with the autosomal dominant disorder, osteogenesis inperfecta (OI) type I. Co-inheritance of the loss of the *Rsa*I site and the OI phenotype was observed with a lod score of 3.91 at a recombination fraction (θ) of zero, indicating strong linkage. This suggests that the defect in this family is caused by a structural mutation within or close to the proα2(I) collagen gene. The use of this high frequency RFLP together with other recently described polymorphisms at this locus will facilitate the analysis of the role of this gene in OI and other inherited disorders of connective tissue.

Key words: linkage/osteogenesis imperfecta/polymorphism/proα2(I) collagen

Introduction

Osteogenesis imperfecta (OI) is a heterogeneous group of inherited disorders characterised chiefly by brittle bones, and associated to a variable degree with abnormalities of other connective tissues (reviewed in Prockop and Kivirikko, 1984). At least four clinical types have been recognised, which vary in the severity of fracturing, mode of inheritance and the degree of involvement of extraskeletal tissues (Sillence *et al.*, 1979). Although it has been suspected for some time that abnormal collagen synthesis is responsible for the OI phenotype, this has only recently been confirmed by the demonstration of shortened proα1 or proα2 chains of type I collagen in several variants of OI (Prockop and Kivirikko, 1984).

A major advance in the study of this condition has been the molecular cloning of the human proα2(I) and proα1(I) collagen genes (Myers *et al.*, 1983; Dalgleish *et al.*, 1982; Chu *et al.*, 1984). Both genes are represented only once in the human genome and are not syntenic, with the proα2(I) and proα1(I) genes having been mapped to chromosomes 7 and 17, respectively (Huerre *et al.*, 1982; Henderson *et al.*, 1983). The structural characterisation of the normal genes has opened up the possibility of direct analysis of the molecular defect in OI. This has recently been accomplished for a variant of OI type II, where the defect was shown to be a deletion of ~500 bases from the proα1(I) gene (Chu *et al.*, 1983). However, the genes are very large, and their products subject to an elaborate pathway of post-transcriptional and post-translational processing and modification (Prockop and Kivirikko, 1984), so that localisation of the defect can be very difficult.

A related approach is to find restriction fragment length polymorphisms (RFLPs) in the type I collagen genes, and to test these markers for linkage to OI in affected families. This will establish whether the disorder is caused by a structural defect in a particular gene. It will also allow diagnosis of the disorder in future offspring of the family, and could be used for pre-natal diagnosis of the severe variants of OI. This approach has been particularly successful in the pre-natal diagnosis of sickle cell anaemia and β thalassaemia. (Boehm *et al.*, 1983).

We have screened the proα2(I) collagen genes of three South African populations for RFLPs as part of a programme aimed at the characterisation and diagnosis of inherited disorders of connective tissue, and have previously reported the presence of an *Msp*I RFLP in these populations (Grobler-Rabie *et al.*, 1985) and a *Bgl*II and *Eco*RI RFLP in South African Blacks (Brebner *et al.*, 1985). The *Msp*I RFLP has also been described by Tsipouras *et al.* (1984), and this group has detected an *Eco*RI RFLP in the 5′ half of the proα2(I) gene which was linked to the phenotype in a family with OI (Tsipouras *et al.*, 1983). Here we report the presence of a high frequency *Rsa*I RFLP in the 3′ half of the proα2(I) gene, and demonstrate linkage between this marker and an autosomal dominant form of OI.

Results

Screening of RFLPs in the human proα2(I) collagen gene

DNA from nine unrelated Afrikaners (white South Africans of predominantly Dutch ancestry) was digested with the restriction enzymes *Bgl*II, *Sca*I, *Hinc*II, *Hinf*I, *Eco*RV or *Rsa*I, electrophoresed, blotted and hybridised to a recombinant plasmid Hf32. This probe contains sequences complementary to the coding regions of the 3′ half of the human proα2(I) collagen gene, including 1443 nucleotides coding for the proα2(I) helical domain and 597 nucleotides coding for the C-propeptide region (Myers *et al.*, 1981). Identical restriction fragment patterns were obtained in all individuals with five of the enzymes (data not shown), whereas variant patterns were observed with *Rsa*I (Figure 1).

Molecular basis of the RsaI RFLP

When *Rsa*I digests of human DNA are probed with Hf 32 at least eight fragments are detected, the largest of which is 2.1 kb in length (Figure 1, lanes a and b). In some individuals an additional fragment of 2.9 kb was detected (Figure 1, c and d), whereas in others the 2.9-kb fragment was present in the absence of the 2.1-kb fragment (Figure 1e), suggesting homozygosity for the polymorphism.

To localise the polymorphism in the gene and to establish its molecular basis, blots were probed with four *Eco*RI genomic subclones from the 3′ half of the human proα2(I) gene (Dalgleish

et al., 1982; Tajima *et al.*, 1984). The variant 2.9-kb *Rsa*I fragment mapped to two adjacent subclones, 2.9 and 1.2 kb in length (see Figure 2), which are located ~10 kb from the 3′ end of the gene. The polymorphism was analysed further by carrying out double digests on DNA from individuals homozygous for the 2.1-kb or 2.9-kb fragment, using the restriction enzyme *Rsa*I with *Xba*I, *Hind*III or *Eco*RI (data not shown). The data summarized in Figure 2, show that the variant 2.9-kb fragment results from the loss of an *Rsa*I site between a 2.1-kb and 0.8-kb *Rsa*I fragment at the 3′ end of the 1.2-kb *Eco*RI subclone. The three genotypes of the polymorphism are therefore: 2.1 kb (+/+), 2.1/2.9 kb (+/−) and 2.9 kb (−/−).

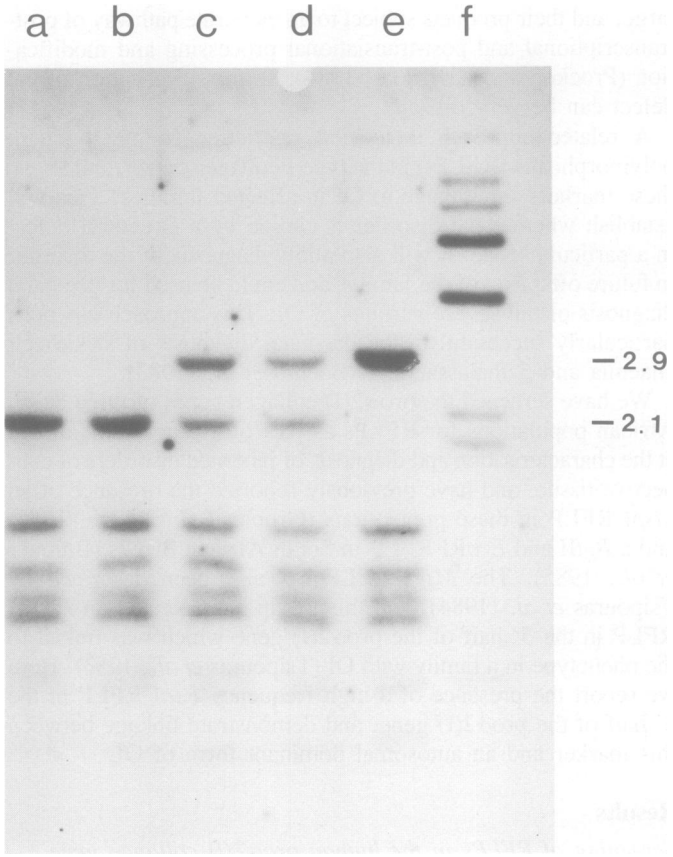

Fig. 1. DNA samples were digested with *Rsa*I, blotted and hybridised to a human proα2(I) collagen probe, Hf 32. **Lanes a** and **b** are homozygotes for the presence of the *Rsa*I site (2.1 kb), **lane c** and **d** are heterozygotes (2.1 + 2.9 kb) and **lane e** is a homozygote for the absence of the site (2.9 kb). **Lane f** is λ DNA digested with *Hind*III.

Frequency of the RsaI polymorphism

The usefulness of a genetic marker in a population is determined by its frequency. We therefore screened three South African populations for the presence of the *Rsa*I RFLP. The results are presented in Table I. The frequency was highest in the Afrikaners, who originated from Western Europe, but is apparently substantially lower in the population of mixed ancestry (the so-called 'Cape Coloured' – Botha and Pritchard, 1972), and absent in South African Blacks. There is no significant difference between the observed and expected distribution of the genotypes for both the Afrikaner and 'Cape Coloured' populations ($p > 0.90$, Table I). These figures indicate that the heterozygote frequency in the Afrikaner population will be 47%.

Linkage analysis of a family with OI

The inheritance of the *Rsa*I RFLP was studied in a family with an autosomal dominant form of OI. Affected individuals presented with blueing of the sclera, a moderate fracturing tendency, dentinogenesis imperfecta and a normal stature. According to the current clinical and genetic classification of OI (Sillence *et al.*, 1979), the condition in this family can be categorised as OI type I.

The *Rsa*I RFLP was analysed in 19 individuals from four generations, and the pedigree and results are shown in Figure 3. Co-inheritance of the loss of the *Rsa*I site and the OI phenotype was observed. The lod score was 3.91 at a recombination fraction (θ) of zero, which indicates strong linkage (Ott, 1974).

Discussion

The data presented in this report establish the existence of a new RFLP which results from the loss of an *Rsa*I site ~10 kb from the 3′ end of the human proα2(I) collagen gene. Since genes in which this site was absent produced normal restriction fragment patterns with 14 other enzymes (Grobler-Rabie *et al.*, 1985 and this report), it is likely that the mutation is a single base change or a minor structural alteration. It occurs in the region of the gene which codes for the α-chain domain of proα2(I) collagen, but since the gene is highly interrupted by intervening sequences (Myers *et al.*, 1983) and the structure of this region has not yet been published, it is not clear whether coding or non-coding sequences are affected. The fact that three apparently normal individuals are homozygous for the mutation (Table I) suggests

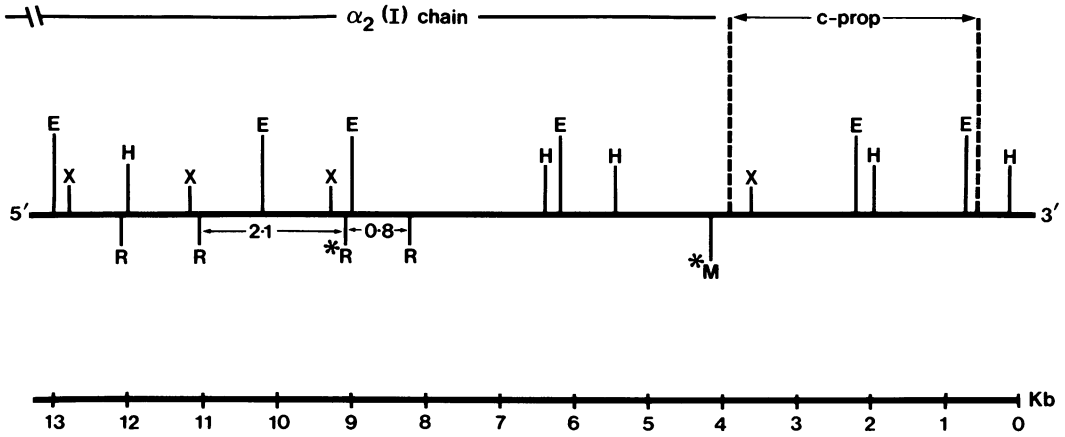

Fig. 2. Restriction map of the 3′ half of the human proα2(I) collagen gene (modified from Tajima *et al.*, 1984 and Myers *et al.*, 1983). Polymorphic sites are indicated by an asterisk. The scale in kilobases is numbered from the 3′ end of the gene. Restriction sites: E = *Eco*RI, H = *Hind*III, X = *Xba*I, R = *Rsa*I, M = *Msp*I, (only the relevant *Rsa*I and *Msp*I sites have been indicated).

Table I. Frequency of *Rsa*I polymorphism in three South African populations

Genotypes	Afrikaners		'Cape Coloureds'		South African Blacks	
	Observed	Expected[a]	Observed	Expected	Observed	Expected
+/+	9	9.375	7	6.96	9	9
+/−	12	11.24	2	1.73	0	0
−/−	3	3.375	0	0.10	0	0
n	24		9		9	
Frequency of minor (−) allele	0.38		0.11		0	

[a]The expected occurrence of the genotypes has been calculated from the frequency of the polymorphism by assuming that it is distributed randomly in the populations.
n = number of individuals tested.

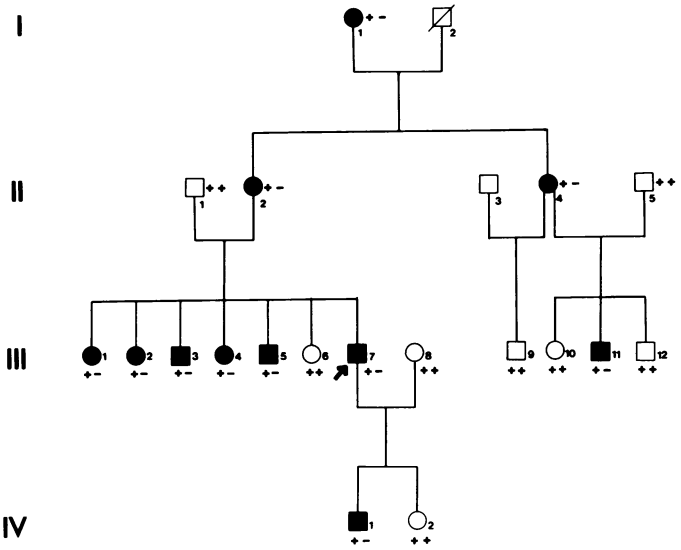

Fig. 3. Pedigree of a family with an autosomal dominant form of OI. The presence and absence of the *Rsa*I polymorphic site is indicated by + and −, respectively. The symbols ■ and ● indicate affected individuals. The proband is arrowed.

that it does not lead to a significantly altered phenotype.

The frequency of the RFLP was very different in the three populations studied (Table I). It was high in the Afrikaners, who are of Western European origin, moderate in the 'Cape Coloured' population and zero in South African Blacks. This implies that the polymorphism originated in Europe and its presence in the 'Cape Coloured' population may result from admixture of settlers from Western Europe with this group. The frequency of 0.38 in the Afrikaner makes it, together with the 5′ *Eco*RI RFLP described by Tsipouras *et al.* (1983), the most informative RFLP detected in the gene to date, with an expected heterozygote frequency of 47%. *Msp*I, *Bgl*II and *Eco*RI RFLPs of lower frequency have been detected in the 3′ half of the proα2(I) gene (Grobler-Rabie *et al.*, 1985; Brebner *et al.*, 1985; Tsipouras *et al.*, 1984). The detection of an increasing number of markers at this locus should significantly improve the chances of finding an informative polymorphism in a given family. It also raises the possibility of establishing polymorphic haplotypes at this locus. This approach has been very useful in the systematic characterisation of globin gene disorders, where the association of a particular haplotype with a specific molecular defect is high (Orkin *et al.*, 1982).

The co-inheritance of the *Rsa*I RFLP and the OI phenotype observed in this study (Figure 3) and the evidence of type I collagen involvement in OI provide a strong indication that a mutation within the proα2(I) collagen gene is responsible for the disorder in this family. Therefore a detailed analysis of this gene and its product should reveal the molecular basis of the disorder. The finding of a proα2(I) defect in OI type I is surprising since all well-defined cases of this type have involved mutations in the proα1(I) gene (Prockop and Kivirikko, 1984), although a variant form of OI type I with a structurally altered proα2(I) chain has been described (Byers *et al.*, 1983). It is also interesting that the phenotype in this family is dominantly inherited. The complete absence of α2(I) chain synthesis is compatible with life, since α1(I) chains can form a functional trimer. Thus a reduction in α2(I) synthesis from one allele might not be expected to have a significant effect. This suggests that the mutation may result in the production of a structurally altered proα2(I) chain that prevents procollagen trimers into which it is incorporated from folding into a triple-helical conformation. Mutations in the proα2(I) gene have now been implicated in the mild autosomal dominant OI type I (this report), the perinatal lethal OI type II (DeWet *et al.*, 1983), the recessively inherited OI type III (Pihlajaniemi *et al.*, 1984) and an autosomal dominant form of OI type IV (Tsipouras *et al.*, 1983). Molecular characterisation of the defect in all of these variants will be required to explain the differences in phenotype and mode of inheritance observed in these four conditions.

The establishment of linkage in this family also implies that future affected offspring can be diagnosed at birth, rather than after the development of symptoms of the disorder. Furthermore the *Rsa*I RFLP could be used for the pre-natal diagnosis of severe variants of OI which involve the proα2(I) gene. Finally, the marker could be applied in linkage studies of other inherited disorders of connective tissue, such as the Ehlers Danlos and Marfan syndromes (Prockop and Kivirikko, 1984) where defects of type I collagen synthesis are suspected.

Materials and methods

High mol. wt. human DNA was prepared from whole blood or from white blood cells of randomly selected individuals as described by Vandenplas *et al.* (1984).

DNA samples were digested to completion with restriction enzymes, which were obtained from Bethesda Research Laboratories, Boehringer Mannheim and New England Biolabs, and used as directed by the suppliers. The resulting DNA fragments were separated by electrophoresis on 1% agarose gels and transferred to nitrocellulose as described (Vandenplas *et al.*, 1984). DNA from λ phage digested with *Hind*III and labelled with [α-^{32}P]dCTP using T4 polymerase (Maniatis *et al.*, 1982) was used as a mol. wt. marker. The nick-translation kit and protocol supplied by Bethesda Research Laboratories was used to label human proα2(I) collagen cDNA (Myers *et al.*, 1981) and genomic subclones (Dalgleish *et al.*, 1982; Tajima *et al.*, 1984) to a specific activity of 1−2 x 10^8 c.p.m./μg with [α-^{32}P]dCTP. The restricted DNA on the filters was hybridised to the probes at 65°C as described (Vandenplas *et al.*, 1984), with final stringency washes of 0.1 x SSC for genomic and 0.5 x SSC for cDNA probes (1 x SSC buffer contains 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0). Autoradiography was for 1−7 days at −80°C.

Acknowledgements

We should like to thank Drs. J. Myers, F. Ramirez and D. Prockop for the human proα2(I) collagen cDNA clone, Hf 32, and Drs. R. Dalgleish and R. Kaufman for the human proα2(I) collagen genomic subclones. This work was supported by grants from the South African Medical Research Council.

References

Boehm,C.D., Antonarakis,S.E., Phillips,J.A., Stetten,G. and Kazazian,H.H. (1983) *N. Engl. J. Med.*, **308**, 1054-1058.

Botha,M.C. and Pritchard,J. (1972) *S. Afr. Med. J. Suppl.*, **1**, 1-27.

Brebner,D.K., Grobler-Rabie,A.F., Bester,A.J., Mathew,C.G. and Boyd,C.D. (1985) *Hum. Genet.*, in press.

Byers,P.H., Shapiro,J.R., Rowe,D.W., David,K.E. and Holbrook,K.A. (1983) *J. Clin. Invest.*, **71**, 689-697.

Chu,M.-L., Williams,C.J., Pepe,G., Hirsch,J.L., Prockop,D.J. and Ramirez,F. (1983) *Nature*, **304**, 78-80.

Chu,M.-L., de Wet,W., Bernard,M., Ding,J.-F., Morabito,M., Myers,J., Williams,C. and Ramirez,F. (1984) *Nature*, **310**, 337-340.

Dalgleish,R., Trapnell,B.C., Crystal,R.G. and Tolstoshev,P. (1982) *J. Biol. Chem.*, **257**, 12816-12822.

De Wet,W.J., Pihlajaniemi,T., Myers,J., Kelly,T.E. and Prockop,D.J. (1983) *J. Biol. Chem.*, **258**, 7721-7728.

Grobler-Rabie,A.F., Brebner,D.K., Vandenplas,S., Wallis,G., Dalgleish,R., Kaufman,R.E., Bester,A.J., Mathew,C.G.P. and Boyd,C.D. (1985) *J. Med. Genet.*, **22**, in press.

Henderson,A.S., Myers,J.C. and Ramirez,F. (1983) *Cytogenet Cell. Genet.*, **36**, 586-587.

Huerre,C., Junien,C., Weil,D., Chu,M.-L., Morabito,M., Van Cong,N., Myers,J.C., Foubert,C., Gross,M.-S., Prockop,D.J., Bone,A., Kaplan,J.-C., de la Chapelle,A. and Ramirez,F. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6627-6630.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.

Myers,J.C., Chu,M.-L., Faro,S.H., Clark,W.J., Prockop,D.J. and Ramirez,F. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3516-3520.

Myers,J.C., Dickson,L.A., de Wet,W.J, Bernard,M.P., Chu,M.-L., Di Liberto,M., Pepe,G., Sangiorgi,F.O. and Ramirez,F. (1983) *J. Biol. Chem.*, **258**, 10128-10135.

Orkin,S.H., Kazazian,H.H., Antonarakis,S.E., Goff,S.C., Boehm,C.C., Sexton,J.P., Waber,P.G. and Giardina,P.J.V. (1982) *Nature*, **296**, 627-631.

Ott,J. (1974) *Am. J. Human Genet.*, **26**, 588-597.

Pihlajaniemi,T., Dickson,L.A., Pope,F.M., Korhonen,V.R., Nicholls,A., Prockop,D.J. and Myers,J.C. (1984) *J. Biol. Chem.*, **259**, 12941-12944.

Prockop,D.J. and Kivirikko,K.I. (1984) *N. Engl. J. Med.*, **311**, 376-386.

Sillence,P.O., Senn,A. and Darks,D.M. (1979) *J. Med. Genet.*, **16**, 101-116.

Tajima,S., Ting,J.P.V., Pinnell,S.R. and Kaufman,R.E. (1984) *J. Invest. Dermatol.*, **82**, 265-269.

Tsipouras,P., Myers,J.C., Ramirez,F. and Prockop,D.J. (1983) *J. Clin. Invest.*, **72**, 1262-1267.

Tsipouras,P., Borresen,A., Dickson,L.A., Berg,K., Prockop,D.J. and Ramirez,F. (1984) *Am. J. Hum. Genet.*, **36**, 1172-1179.

Vandenplas,S., Wiid,I., Grobler-Rabie,A., Brebner,K., Ricketts,M., Wallis,G., Bester,A., Boyd,C. and Mathew,C. (1984) *J. Med. Genet.*, **21**, 164-172.

Received on 25 March 1985; revised on 13 May 1985