Volume 334, number 2, 237–240 FEBS 13249 November 1993

Length polymorphisms in new human collagen-like loci

Jeffrey L. Doering[a,*], Anne E. Burket[a], Lawrence C. Vogel[b]

[a]*Department of Biology, Loyola University of Chicago, 6525 North Sheridan Road, Chicago, IL 60626, USA*
[b]*Shriners Hospital for Crippled Children, Chicago, IL 60635, USA*

(Received 20 September 1993)

Two new collagen-like loci have been identified in the human genome which have sequence similarity to the triple-helical coding region of the pro-α2(I) gene. Both loci exhibit length polymorphism due to alleles that contain deletions. The deletion at one locus is 400 bp while the deletion at the second locus is 200 bp. The second locus is on chromosome 17 and its two alleles are not in Hardy-Weinberg equilibrium. These loci are candidates for involvement in connective tissue disorders.

Pro-α2(I) collagen gene; Deletions; Alleles; Restriction fragment length polymorphisms

1. INTRODUCTION

The collagens are a large family of proteins which are essential components of extracellular matrices. At least 14 genetically distinct collagen types have so far been identified, many of which are heterotrimers [1]. Thus, more than 20 genes in the human genome are involved in encoding these proteins [1]. While the functions for many of the collagens are not yet well-understood, the major fibrillar collagens Types I, II and III have been extensively studied [1].

Defects in the structural genes for the fibrillar collagens are now known to be involved in several of the heritable disorders of connective tissue [2,3]. A number of restriction fragment length polymorphisms (RFLPs) have been found to be associated with each of these collagen genes. In the case of the genes for Type I collagen, RFLPs have already proven useful in following the inheritance of some forms of diseases like osteogenesis imperfecta and Ehlers-Danlos syndrome [4]. The majority of fibrillar collagen gene RFLPs identified to date are the result of base substitutions that change a restriction enzyme site [4]. There have been only a few reports of length polymorphisms due to insertion/deletions at human fibrillar collagen gene loci [5,6]. Here we describe two previously-undetected collagen-like loci which both exhibit length polymorphism and which could be involved in connective tissue disorders.

2. MATERIALS AND METHODS

Twenty ml peripheral blood samples were obtained from individuals having no personal or family history of connective tissue disorders. DNA was isolated from the whole blood by a standard procedure [7], which we have modified. The modifications involve using spermine precipitation of the DNA [8], rather than ethanol precipitation, and omitting RNase treatment of the sample. We find that these changes improve the yield of DNA and reproducibly give samples lacking restriction enzyme inhibitors. The yield from each 20 ml of blood is at least 200 μg. Normal placental DNA was purified as previously described [9].

Our described methods were used for restriction digests and gel electrophoresis [10,11]. The DNA was transferred from the gel by the alkaline method [12] onto Gene Screen Plus membranes (NEN). cDNA probes were cut by restriction enzymes to release the inserts from the vector and then ^{32}P-labelled by nick translation as described [10,13]. Membranes were pre-hybridized at 37°C for 5 h in hybridization solution (50% formamide, 1 M NaCl, 50 mM Tris pH 7.5, 1% SDS, 10 μg/ml denatured *E. coli* DNA). Denatured labelled probe was added and hybridization took place at 37°C for 16 h. Membranes were then given two 10 min washes at room temperature in 2 × SSC (1 × SSC is 150 mM NaCl, 15 mM sodium citrate, 0.1 mM EDTA), followed by two 30 min washes at 60°C in 2 × SSC plus 1% SDS and two 30 min washes at room temperature in 0.5 × SSC. For the CosHcol1 probe, which contains repetitive sequences, the hybridization temperature was 55°C with denatured human placental DNA added to the hybridization mixture at 30 μg/ml. The second two sets of washes were in 0.5 × SSC plus 1% SDS at 65°C and in 0.1 × SSC at room temperature. Autoradiography of the membranes took place as described [10].

3. RESULTS AND DISCUSSION

In the course of other studies, the pro-α2(I) cDNA probe Hf32 [14] was used to examine the collagen gene structure in a number of normal individuals. Two previously-undescribed length polymorphisms were found. An example of one of the polymorphisms (Site 1) is shown in Fig. 1, where an individual has, instead of the usual 7.4 kb *Hin*dIII fragment, both a 7.0 kb and 7.4 kb fragment, both of which are present in about half the amount as the 7.4 kb *Hin*dIII fragment found in most individuals. Similarly, this same individual has both a 2.6 kb and a 3.0 kb *Pst*I fragment instead of just the

*Corresponding author. Fax: (1) (312) 508-3646.

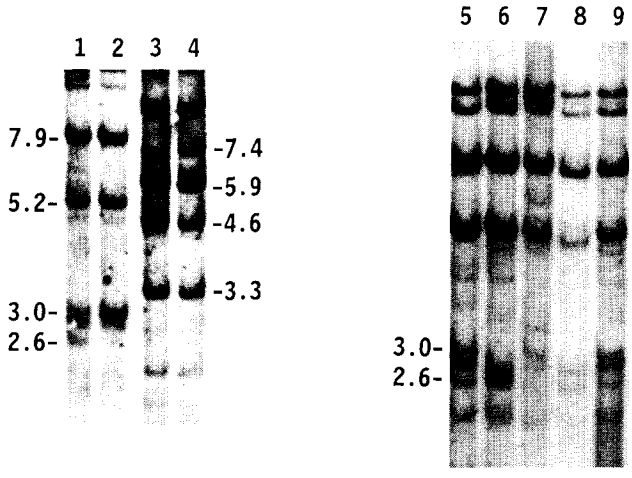

Fig. 1. Length polymorphism at Site 1. DNAs from two individuals were digested with *Pst*I (lanes 1 and 2) or *Hin*dIII (lanes 3 and 4). DNAs from five other individuals were digested with *Pst*I (lanes 5–9). All were probed with the pro-α2(I) cDNA clone Hf32 [14]. The fragment sizes in kb are indicated. Note the heterozygous polymorphism of the 3.0 kb *Pst*I fragment and the 7.4 kb *Hin*dIII fragment (lanes 1 and 3) and the homozygous polymorphism of the 3.0 kb *Pst*I fragment (lane 6).

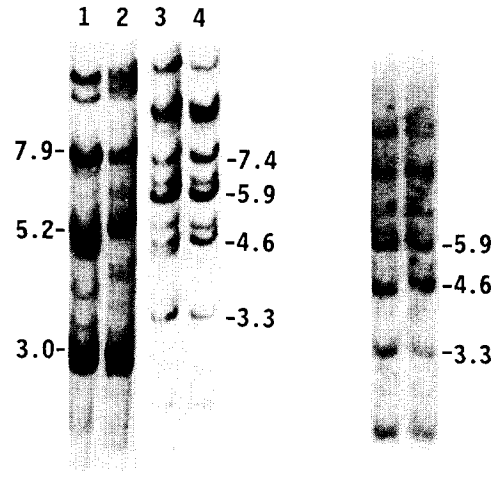

Fig. 2. Length polymorphism at Site 2. DNAs from two individuals were digested with *Pst*I (lanes 1 and 2) or *Hin*dIII (lanes 3 and 4). DNAs from two other individuals were digested with *Hin*dIII (lanes 5 and 6). All were probed with Hf32. The fragment sizes in kb are indicated. Note that one individual is heterozygous for the length polymorphism (lanes 1 and 3) and another individual is homozygous (lane 5).

usual 3.0 kb fragment. Again both of these fragments are present in about half the amount as the 3.0 kb fragment present in most individuals. Such complementary results with more than one restriction enzyme demonstrate that the polymorphism at Site 1 is due to a 400 bp deletion. The individual in question is thus heterozygous for this deletion, and examples of homozygotes for the deletion have also been observed (Fig. 1).

Examples of the second polymorphism (Site 2) are shown in Fig. 2. Here, instead of the usual 4.6 kb *Hin*dIII fragment, both a 4.4 and a 4.6 kb fragment are seen. Similarly, both a 5.0 and a 5.2 kb *Pst*I fragment are present, instead of just the usual 5.2 kb *Pst*I fragment. Again, when the two fragments are found, they are each present in about half the amount as the usual full-sized fragment. These results indicate the presence of 200 bp deletions at Site 2, and both homozygotes and heterozygotes for the deletion have been observed (Fig. 2).

Twenty individuals were examined at Site 1, and 15 were found homozygous for the large (non-deleted) allele, 4 were heterozygous, and 1 was homozygous for the small allele. Thus, the frequency for the small allele is 0.15 ± 0.06, and the two alleles are in Hardy-Weinberg equilibrium ($P > 0.7$). At Site 2 there were 14 individuals homozygous for the large allele, 2 heterozygotes, and 4 homozygous for the small allele. The frequency for the small allele at Site 2 is thus 0.25 ± 0.07. Interestingly, the alleles at Site 2 are *not* in Hardy-Weinberg equilibrium ($P < 0.006$) due to an unexpectedly low frequency of heterozygotes. Given the sample sizes, the Yates correction was used in determining Hardy-Weinberg equilibria.

The fragments at Sites 1 and 2 are detected with a pro-α2(I) cDNA probe, but the restriction map of the COL1A2 gene does not contain the 7.4 kb *Hin*dIII and 3.0 kb *Pst*I fragments characteristic of Site I nor the 4.6 kb *Hin*dIII and 5.2 kb *Pst*I fragments characteristic of Site 2 [15,16,17]. COL1A2 genomic clones that span the same region of the gene as does the cDNA were used to probe samples where the cDNA probe had detected length polymorphisms at Sites 1 or 2 (Fig. 3). In no case did the genomic clones detect the usual-sized or deletion-containing restriction fragments characteristic of either Sites 1 or 2. The fragments containing Site 2 are similar in size to ones found in COL1A2, but by sequentially hybridizing the same blot with the cDNA probe and then the genomic clones we clearly demonstrated that the restriction fragments characteristic of Site 2 are distinct from those in COL1A2 (Fig. 3). Thus the pro-

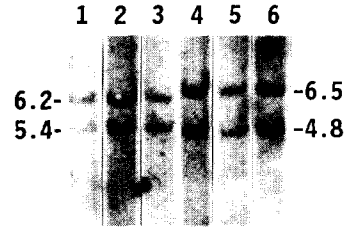

Fig. 3. Absence of length polymorphism in COL1A2. DNAs from three individuals were digested with *Pst*I (lanes 1–3) or *Hin*dIII (lanes 4–6) and hybridized with the COL1A2 genomic clone NJ-3 [15]. Lanes 1 and 5 contain a sample in which Hf32 detected a homozygous deletion at Site 2. This same filter had previously been hybridized with Hf32. Note the lack of any length polymorphism in this sample when probed with NJ-3. Similarly, no polymorphism was detected when the NJ-1 genomic clone of COL1A2 [15] was used as probe. The fragment sizes in kb are indicated.

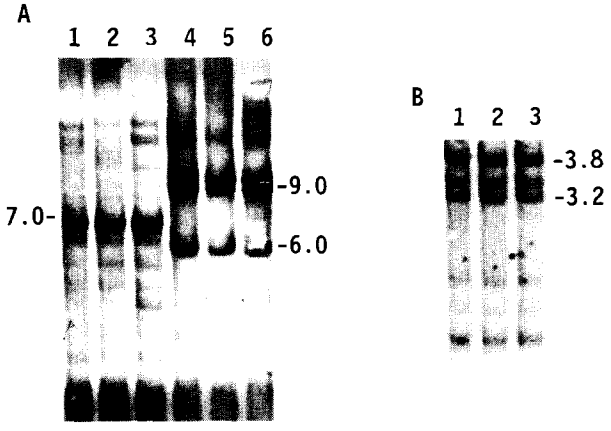

Fig. 4. Absence of the deletion sites in other collagen genes. (A) DNAs from three individuals where length polymorphisms were found with Hf32 were digested with *Pst*I (lanes 1–3) or *Hin*dIII (lanes 4–6) and probed with the COL1A1 cDNA Hf404 [24]. Note the absence of any polymorphisms with this probe. Similarly, no polymorphisms were detected with the Hf677 cDNA probe for COL1A1 [24]. (B) DNAs from three individuals where length polymorphisms were detected with Hf32 were digested with *Pst*I and probed with the CosHcol1 cosmid clone for the full COL2A1 gene [25]. No length polymorphisms are detected. Sizes in kb of selected fragments are indicated.

α2(I) cDNA must be able to hybridize under our conditions to sequences besides those in COL1A2, and Sites 1 and 2 must be in a pro-α2(I)-like genetic region(s) other than COL1A2. Since the COL1A2 gene is primarily made up of very short exons (usually less than 100 bp) [15,17] one would not expect the highly intron-interrupted genomic probes to detect the cross-hybridizing new collagen-like loci. The restriction fragments of both sites are detected only by the triple helical coding domain of the cDNA probe (data not shown), so these regions do not contain any sequences related to the pro-α2(I) carboxy-propeptide.

We have attempted to determine whether the deletion sites are in some other known human fibrillar collagen genes. cDNA probes corresponding to the full COL1A1 gene were used to probe samples where the pro-α2(I) cDNA probe had detected deletions, and again none of the deletion-containing restriction fragments characteristic of Sites 1 or 2 were detected (Fig. 4A). A genomic cosmid clone for the full COL2A1 gene (Fig. 4B) and cDNA probes for the full COL5A2 [18] and COL3A1 [19] genes similarly detected none of the deletion-containing fragments (not shown). Thus, Sites 1 and 2 are not in these other collagen genes, but are likely in one or more still-undescribed genes closely related to COL1A2. At least for Site 2, its chromosomal localization also rules out its being in any of the other known fibrillar collagen genes (see below).

A human-hamster somatic cell hybrid panel was used to identify the chromosome(s) carrying Sites 1 and 2. Site 2 was localized to chromosome 17 (Fig. 5). Site 1 has not yet been localized due to the fact that the hamster genome contains restriction fragments of similar size to those containing Site 1 that also cross-hybridize with the pro-α2(I) cDNA (Fig. 5). The only other known collagen gene on chromosome 17 is COL1A1 [2], and, as shown above, Site 2 is not in that gene. All other known human collagen genes have been localized to chromosomes other than 17 [2]. This provides additional evidence that Site 2 is not in an already-characterized collagen gene.

The majority of fibrillar collagen gene RFLPs identified to date are the result of base substitutions [4]. There have been only a few reports of length polymorphisms at human fibrillar collagen gene loci. A 38 bp insertion/deletion is a common polymorphism within an intron of the COL1A2 gene [5], and a region adjacent to the 3′ end of the COL2A1 gene exhibits a high degree of length polymorphism due to variation in the number of copies of a simple tandemly repeating sequence [6,20]. The COL3A1 gene contains a polymorphic block of dinucleotide repeats in one intron [21]. Recently a chromosome 11 locus containing some uncharacterized collagen-like sequences has been shown to have substantial length polymorphism arising from variable numbers of a tandem repeat [22]. However, the properties of this locus clearly distinguish it from the two polymorphic sites we have described here.

Thus, the two collagen-like sites identified are not in any previously-described loci, and, at this point, it is not known whether both sites are in the same gene. Based on the length polymorphisms described above for other human collagen loci, it is likely that these new deletion sites are in introns or other non-coding regions. Of particular interest is the fact that at least at one of the sites the alleles are not in Hardy-Weinberg equilibrium, indicating that the locus produces a product and that some genotypes can lead to a detrimental phenotype. Since the short allele is found in phenotypically unaffected

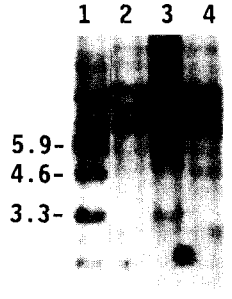

Fig. 5. Identifying the chromosome carrying deletion Site 2. The indicated DNAs from a human-hamster somatic cell hybrid panel (Bios Corp., New Haven) were digested with *Hin*dIII and hybridized with the Hf32 probe. Representative hybrid DNAs from the chromosome panel are shown. Lanes: 1, total human DNA; 2, hybrid cell DNA containing neither chromosomes 7 or 17; 3, hybrid cell DNA containing chromosome 7; 4, hybrid cell DNA containing chromosome 17. Sizes in kb of selected fragments are indicated. The 4.6 kb fragment characteristic of Site 2 is only found in hybrid cells containing chromosome 17. The 5.9 and 3.3 kb fragments from the COL1A2 gene map to chromosome 7 as previously shown [26].

individuals, it suggests that the deletion serves as a marker for an allele that is not itself harmful, but that predisposes to detrimental phenotypes. Such alleles in the unaffected population have been found to be involved in fragile X syndrome and myotonic dystrophy [23]. These two new collagen-like sites should therefore be examined for their possible involvement in various connective tissue disorders. The polymorphic nature of the sites may prove useful in this analysis.

Acknowledgements: This work was supported by grants from the Shriners Hosptitals for Crippled Children and the Osteogenesis Imperfecta Foundation, Inc. We thank F. Ramirez and E. Weiss for generously providing probes.

REFERENCES

[1] Fleischmajer, R., Olsen, B.R. and Kuhn, K. (eds.) (1990) Structure, Molecular Biology, and Pathology of Collagen, Ann. NY Acad. Sci., vol. 580.
[2] Byers, P.H. (1990) Trends Genet. 6, 293–300.
[3] Kuivaniemi, H., Tromp, G. and Prockop, D.J. (1991) FASEB J. 5, 2052–2060.
[4] Pope, F.M., Daw, S.C.M., Narcisi, P., Richards, A.R. and Nicholls, A.C. (1989) J. Inherit. Metab. Dis. 12, Suppl. 1, 135–173.
[5] Dalgleish, R., Williams, G. and Hawkins, J.R. (1986) Hum. Genet. 73, 91–92.
[6] Sykes, B.C., Ogilvie, D.J. and Wordsworth, B.P. (1985) Hum. Genet. 70, 35–37.
[7] Wilson, J.T., Milner, P.F., Summer, M.E., Nallaseth, F.S., Fadel, H.E., Reindollar, R.H., McDonough, P.G. and Wilson, L.B. (1982) Proc. Natl. Acad. Sci. USA 79, 3628–3631.
[8] Hoopes, B.C. and McClure, W.R. (1981) Nucleic Acids Res. 9, 5493–5504.
[9] Blin, N. and Stafford, D.W. (1976) Nucleic Acids Res. 3, 2303–2308.
[10] Doering, J.L., Jelachich, M.L. and Hanlon, K.M. (1982) FEBS Lett. 146, 47–51.
[11] Sims, M.A., Doering, J.L. and Hoyle, H.D. (1983) Nucleic Acids Res. 11, 277–290.
[12] Reed, K.C. and Mann, D.A. (1985) Nucleic Acids Res. 13, 7207–7221.
[13] Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.
[14] 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.
[15] 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.
[16] Tsipouras, P., Myers, J.C., Ramirez, F. and Prockop, D.J. (1983) J. Clin. Invest. 72, 1262–1267.
[17] de Wet, W., Bernard, M., Benson-Chanda, V., Chu, M.-L., Dickson, L., Weil, D. and Ramirez, F. (1987) J. Biol. Chem. 262, 16032–16036.
[18] Weil, D., Bernard, M., Gargano, S. and Ramirez, F. (1987) Nucleic Acids Res. 15, (1987) 181–198.
[19] Chu, M.-L., Weil, D., de Wet, W., Bernard, M., Sippola, M. and Ramirez, F. (1985) J. Biol. Chem. 260, 4357–4363.
[20] Stoker, N.G., Cheah, K.S.E., Griffin, J.R., Pope, F.M. and Solomon, E. (1985) Nucleic Acids Res. 13, 4613–4622.
[21] Lee, B., D'Alessio, M.D., Vissing, H., Ramirez, F., Steinmann, B. and Superti-Furga, A. (1991) Am. J. Hum. Genet. 48, 511–517.
[22] Brookes, A.J., Hedge, P.H. and Solomon, E. (1989) Nucleic Acids Res. 17, 1792.
[23] Caskey, C.T., Pizzuti, A., Fu, Y-H., Fenwick, R.G. and Nelson, D.L. (1992) Science 256, 784–789.
[24] Chu, M.-L., Myers, J.C., Bernard, M.P., Ding, J.-F. and Ramirez, F. (1982) Nucleic Acids Res. 10, 5925–5934.
[25] Cheah, K.S.E., Stoker, N.G., Griffin, J.R., Grosveld, F.G. and Solomon, E. (1985) Proc. Natl. Acad. Sci. USA 82, 2555–2559.
[26] Junien, C., Weil, D., Myers, J.C., Cong, N.V., Chu, M.-L., Foubert, C., Gross, M.-S., Prockop, D.J., Kaplan, J.-C. and Ramirez, F. (1982) Am. J. Hum. Genet. 34, 381–387.