Restriction Fragment Length Polymorphism Associated with the Proα2(I) Gene of Human Type I Procollagen

APPLICATION TO A FAMILY WITH AN AUTOSOMAL DOMINANT FORM OF OSTEOGENESIS IMPERFECTA

PETROS TSIPOURAS, JEANNE C. MYERS, FRANCESCO RAMIREZ, and
DARWIN J. PROCKOP, *Departments of Biochemistry, Pediatrics, and Obstetrics
and Gynecology, University of Medicine and Dentistry of New Jersey-
Rutgers Medical School, Piscataway, New Jersey 08854*

ABSTRACT One cloned complementary DNA and one genomic subclone were used to detect restriction fragment length polymorphism associated with the proα2(I) gene for human type I procollagen. The restriction fragments obtained from examination of 30–122 chromosomes confirmed previous indications that the proα2(I) gene is found in a single copy in the human haploid genome. One highly polymorphic site was detected with EcoRI in the 5′-half of the gene. The restriction site polymorphism at the site had an allelic frequency of 0.38, and it generated two fragments of 10.5 and 3.5 kilobase in homozygous individuals. The restriction fragment length polymorphism generated at the EcoRI site was used to study affected and nonaffected individuals in four generations of a family with an autosomal dominant form of osteogenesis imperfecta. The data demonstrated a linkage of the phenotype to a proα2(I) allele with a lod score of 2.41 at a recombination fraction (θ) of 0. The data therefore provided presumptive evidence that osteogenesis imperfecta in this family is caused by a mutation in the proα2(I) gene or some contiguous region of the genome. The relatively high frequency of polymorphism at the EcoRI site makes it useful for studying a broad range of genetic disorders in which mutations in type I procollagen are suspected. In addition, the polymorphic site should provide useful markers for linkage studies with other loci located on human chromosome 7.

INTRODUCTION

Variations in base sequences in and around genes generate DNA fragments of varying length when genomic DNA from different individuals is cleaved with restriction endonucleases. The detection of such fragments of varying lengths, known as RFLP[1], was first used to examine the inheritance of specific alleles of mitochondrial and yeast genes (1–3). Within the past several years, it has proven to be a powerful tool for the study of human genetic disorders such as sickle cell disease (4–5), thalassemias (6), and human growth hormone deficiency (7, 8).

Recently, clones of complementary (c)DNA and genomic DNA for type I procollagen have become available (9–19). Studies with these DNA have revealed some unique characteristics of the collagen genes such as a complex intron-exon structure (9–11), a high degree of nucleotide conservation in an area coding for the carbohydrate attachment site in the C-propeptide of the proα2(I) chain (16), and transcription of each gene into several different messenger RNA (19). Furthermore, it has been established that the two human genes for proα1(I) and proα2(I) chains are not syntenic

---

but they are located on chromosomes 17 (17) and 7 (18), respectively. Here we have used one cloned cDNA and a genomic subclone to detect RFLP associated with the proα2(I) gene of type I procollagen.

METHODS

Materials. Restriction endonucleases and other enzymes were purchased from New England Nuclear, Boston, MA, and Boehringer-Mannheim Biochemicals, Indianapolis, IN. ^{32}P-Labeled deoxynucleotides were obtained from New England Nuclear and nitrocellulose paper was obtained from Schleicher & Schuell, Inc., Keene, NH.

Nuclear DNA preparation. High-molecular-weight DNA was prepared (20) from peripheral blood obtained from randomly selected normal individuals and from affected and nonaffected members of one family with autosomal dominant forms of osteogenesis imperfecta (OI).

Preparation of DNA probes. The cloned cDNA Hf–32 was used here as an insert in pBR322 (14). The probe specific for the 5'-end of the gene (19) consisted of two contiguous EcoRI genomic subclones (NJ-3' and NJ-3" in Fig. 1). The EcoRI site at the 5'-terminus of NJ-3' is an artificial site, generated by the linker in the course of preparing a lambda phage library with a AluI/HaeIII digest of genomic DNA (22). The probes were labeled to a specific activity of 2 to 5 × 10⁸ cpm/μg by nick-translation.

Analysis of RFLP. Samples of nuclear DNA prepared from leukocytes were digested to completion under conditions recommended by the commercial supplier. Digested DNA and appropriate DNA size markers were separated by electrophoresis in 0.6–1.0% (wt/vol) agarose gels. The DNA fragments were transferred to nitrocellulose filters and hybridized with the human proα2(I) probes for 24–48 h at 40°C (20). The filters were then washed for 10 min at 68°C with each of the following solutions: 2 × SSC, 1 × SSC, 0.5 × SSC, and 0.1 × SSC (SSC, buffer containing 0.15 M NaCl in 0.015 M sodium citrate, pH 6.8) (20).

RESULTS

DNA probes for the human proα2(I) gene. The cDNA probe (Hf–32) used in these experiments contained 1,443 nucleotides coding for the α2(I) domain and 597 nucleotides coding for the C-propeptide, and it included about half of the coding sequences of the gene (16). Because of the large number and large size of the intervening sequences in the gene, the Hf–32 probe hybridized to fragments spanning over 12 kilobase (kb) of the gene (Fig. 1). The probe for the 5'-end of the gene consisted of two DNA fragments (NJ-3' and NJ-3"), which contained 6.75 kb of genomic DNA extending downstream from the codon for the amino acid residue 19 of the α2(I) chain (Fig. 1).

RFLP associated with the human proα2(I) gene. To search for RFLP associated with the 3'-half of the proα2(I) gene, leukocyte DNA from 61 individuals was cleaved with 11 restriction endonucleases and hybridized with Hf–32. The 61 individuals were randomly selected among members of our laboratory staff and other volunteers. The number of chromosomes examined with each restriction endonuclease varied from 82 to 122. As indicated in Table I, no RFLP was detected with 10 of the restriction endonucleases. One individual had a polymorphic site for PstI. In individuals lacking the site, three PstI fragments of 5.7, 8.2, and 15.5 kb were generated. Restriction mapping of genomic clones for proα2(I) indicated that the PstI fragments hybridizing with Hf–32 are arranged in the 5'- to 3'-order of 5.7, 8.2, and 15.5 kb. Hf–32 itself contains a single PstI site (16), which corresponds to
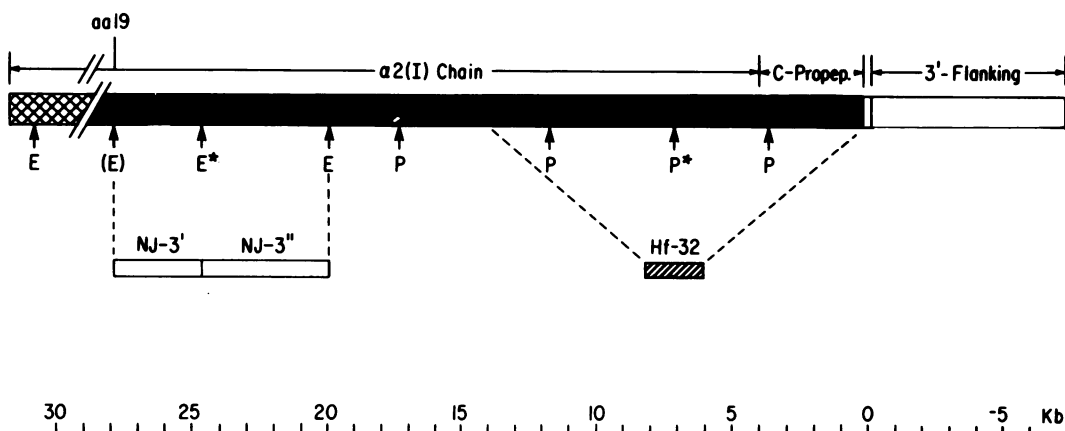

FIGURE 1 Schematic map of the proα2(I) gene of human type I procollagen and the DNA probes used here. The 5'-boundary of the gene has not been precisely defined but is ~38 kb from the 3'-end (10, 19). E and P indicate EcoRI and PstI sites, respectively. The locations of the two polymorphic sites studied here are indicated by the asterisks. The polymorphic site for PstI may or may not be in an exon. (E) indicates an artificial EcoRI site in the genomic probe.

the site generating the 8.2- and 15.5-kb fragments. With DNA from the individual displaying the polymorphism, fragments of 4.7 and 3.5 kb were detected in addition to the fragments of 5.7, 8.2, and 15.5 kb. Therefore, it appeared that the RFLP in this individual consisted of an additional PstI site within the 8.2-kb fragment. Based on the data provided in Table I, the frequency of the polymorphic site for PstI is ∼0.01.

To search for RFLP associated with the 5′-end of the gene, similar experiments were carried out with the combined probe NJ-3′ and NJ-3″. A highly polymorphic site for EcoRI was detected at the 5′-end of the gene. DNA of individuals homozygous for the absence of the polymorphic site (−/−) generated one fragment of 14 kb (Fig. 2). DNA from individuals homozygous for the presence of the polymorphic site (+/+) generated two fragments of 10.5 and 3.5 kb. DNA from individuals heterozygous for the polymorphic site (+/−) generated three fragments of 14, 10.5, and 3.5 kb. Thus, we identified three different genotypes. Analysis of DNA from a normal family demonstrated that the polymorphic site segregated as an autosomal dominant trait (Fig. 2). A total of 100 chromosomes were examined from randomly selected individuals, and the allelic frequency for the presence of the polymorphic site was 0.38. The three genotypes were in Hardy-Weinberg equilibrium with observed values close to the theoretical ($P < 0.10$).

To confirm the polymorphic nature of the EcoRI site, we digested DNA from individuals homozygous for the presence (+/+) or absence (−/−) of the polymorphic site with other enzymes mapped in the same
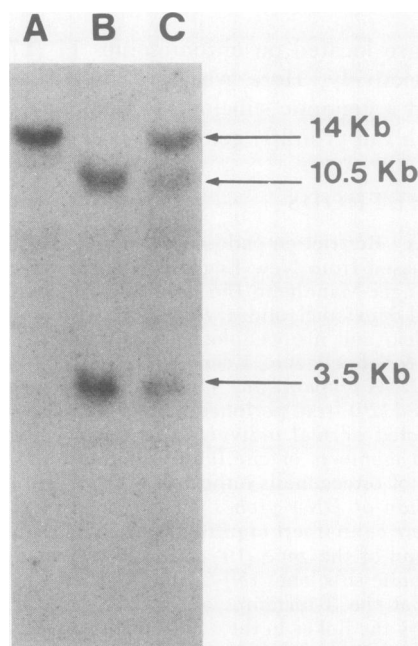

FIGURE 2 RFLP detected with the probe for the 5′-half of the gene (NJ-3′ and NJ-3″) after digestion of genomic DNA with EcoRI. (A) Parent who is homozygous for absence of the polymorphic site (−/−); (B) parent who is homozygous for the site (+/+); (C) offspring who is heterozygous for the site (+/−).

region. Single digestions with the restriction endonucleases HindIII and XbaI, and double digestions with the enzymes EcoRI and HindIII produced fragments with sizes consistent with the map distances of the recognition sites in the same region of the gene (19).

Analysis of a family with OI. The RFLP generated at the EcoRI in the 5′-region of the gene was used to study a family with an autosomal dominant form of OI. Leukocyte DNA was obtained from 18 individuals in four generations (Fig. 3). To a variable degree, affected individuals presented with fractures, dentinogenesis imperfecta, hearing loss, small joint laxity, and white sclerae. In generation I, the affected individual (I-2) was heterozygous (−/+) for the presence of the EcoRI site and the nonaffected (I-1) was homozygous for absence of the site (−/−). In the subsequent three generations the phenotype cosegregated with a proα2(I) allele bearing the EcoRI recognition site. The lod score was 2.41 at a recombination fraction (θ) of 0, indicating strong linkage (21).

DISCUSSION

Several human genes have now been surveyed for DNA polymorphism, readily detectable as RFLP. In

TABLE I
*RFLP Associated with Coding Sequences of 3′-Half of the Proα2(1) Gene**

Number of chromosomes screened	Restriction endonuclease	Cleavage sites screened
88	BamHI	2
94	BgII	3
86	BgIII	5
122	EcoRI	4
106	HindIII	5
82	HpaI	3
94	KpnI	3
100	PstI	3–4‡
88	PvuII	4
88	SacI	6
84	XbaI	5

* Hybridization with Hf-32, a 2.2-kb cDNA (see Fig. 1).
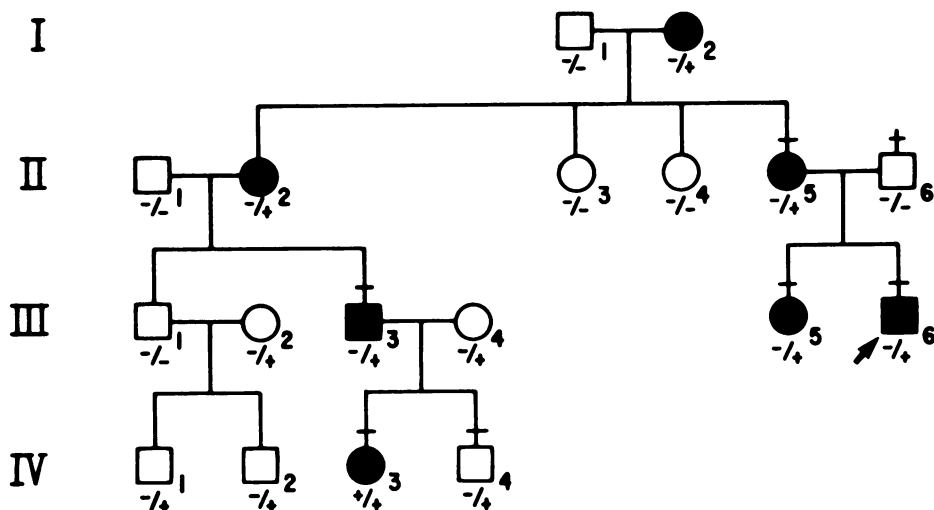‡ Polymorphic site found in one chromosome examined.

FIGURE 3 Family with an autosomal dominant form of OI. Symbols: Individuals heterozygous (−/+) and homozygous for the absence (−/−) and presence (+/+) of the polymorphic EcoRI site; horizontal bars, individuals examined by one of us (Dr. P. Tsipouras). (■ ●): affected by OI; (+): examined clinically; arrow: proband.

the cluster of β-globin genes, seven sites of relatively high frequency have been identified in ~60 kb of DNA (23). In addition, RFLP generated by either of two restriction endonucleases, DdeI or MstII, has been used to identify directly the mutation in the β^s-globin gene (4, 5). In the case of the gene for human growth hormone, five polymorphic restriction sites with allelic frequencies ranging from 0.30 to 0.44 have been detected within ~40 kb of genomic DNA (7, 8). With the human insulin gene, extensive RFLP is found in the 5'-flanking region as a result of either insertions or deletions at a site ~800 bases upstream from the gene (24). The most common alteration is a 1.6-kb insertion, which has an allelic frequency of ~0.20 in normal individuals, and a significantly higher frequency in diabetics who are noninsulin dependent.

The probes used here made it possible to survey ~25 kb of the proα2(I) gene for RFLP (Fig. 1). Only one polymorphic site in one individual was found with the cDNA Hf−32, even though this probe hybridized to fragments spanning over 12 kb of the gene and a total of 210 bases were surveyed with 11 restriction endonucleases. It is possible, however, that additional polymorphic sites will be found as the same region is explored more extensively with additional restriction endonucleases or with genomic probes. In contrast, the polymorphic EcoRI site in the 5'-half of the gene had the very high allelic frequency of 0.38.

The data developed here rigorously confirm previous indications that the proα2(I) gene is found in a

single copy in the human haploid genome (11, 25, 26). All the fragments detected with the genomic probe were accounted for by the linear map developed from overlapping phage clones of the human proα2(I) gene (19). Since the intervening and flanking sequences of highly homologous genes, such as the two α-globin genes (27) are dissimilar, a different restriction pattern would have been expected if more than one copy of the proα2(I) procollagen gene existed per haploid genome. The three genotypes for the polymorphic EcoRI site are in equilibrium. Therefore it is likely that because of a founder effect, the mutation creating or abolishing the recognition site for the EcoRI was fixed in the human genome before any mutations producing genetic disorders of collagen appeared.

The EcoRI site in the 5'-end of the proα2(I) gene made it possible to demonstrate linkage of an OI phenotype with the presence of a proα2(I) allele in four generations of a family with an autosomal dominant form of OI. Since a lod score of 2.41 is generally regarded as significant if derived from a single family (21), the data obtained here indicate linkage. The results therefore provide presumptive evidence that the OI phenotype in this family is produced by a mutation in the proα2(I) gene or in some contiguous region of the genome.

It should be noted that use of RFLP to establish linkage is probably more important for examining genetic disorders of connective tissue than for most other genetic diseases. A variety of observations suggest that

mutations in genes for type I procollagen are the cause of many forms of OI, of Ehlers-Danlos syndrome, and of Marfan syndrome (28–30). However, it has been difficult to generate definitive data about the molecular defects, because of the large size of both the genes and the proteins. Also, the assembly of the procollagen molecule involves at least 11 posttranslational enzymes, and defects in one or more of these can produce disorders of connective tissue. In addition, defects of genes for other components of connective tissue may well produce similar phenotypes. Therefore, data demonstrating linkage of clinical phenotypes to specific alleles should provide invaluable information for systematically defining the molecular basis of a variety of disorders of the extracellular matrix. It should also be noted that the relatively high frequency of the EcoRI RFLP associated with the proα2(I) gene can be used as molecular marker for linkage studies with other foci assigned on chromosome 7, a chromosome for which only a few markers are currently available. A partially characterized α1-like collagen gene with a HindIII RFLP has recently been located on chromosome 7 (31).

ACKNOWLEDGMENTS

REFERENCES

1. Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet. 32:314–331.
2. Grodzicker, T., J. Williams, P. Sharp, and J. Sambrook. 1974. Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harbor Symp. Quant. Biol. 39:439–446.
3. Hutchinson, C., J. Newbold, S. Potter, and M. Edgell. 1974. Maternal inheritance of mammalian mitochondrial DNA. Nature (Lond.). 251:536.
4. Chang, J. C., and Y. W. Kan. 1982. A sensitive new prenatal test for sickle-cell anemia. N. Engl. J. Med. 307:30–31.
5. Orkin, S. H., P. F. R. Little, H. H. Kazazian, and C. D. Boehm. 1982. Improved detection of the sickle mutation by DNA analysis. N. Engl. J. Med. 307:32–36.
6. Kazazian, H. H., J. A. Phillips, C. D. Boehm, T. A. Vik, M. J. Mahoney, and A. K. Ritchey. 1980. Prenatal diagnosis of thalassemias by amniocentesis: linkage analysis using multiple polymorphic restriction endonuclease sites. Blood. 56:926–930.
7. Phillips, J. A., B. L. Hjelle, P. H. Seeburg, and M. Zachmann. 1981. Molecular basis for familial isolated growth

hormone deficiency. Proc. Natl. Acad. Sci. USA. 78:6372–6375.
8. Phillips, J. A., J. S. Parks, B. L. Hjelle, J. E. Herd, L. P. Plotnick, C. J. Migeon, and P. H. Seeburg. 1982. Genetic analysis of familial isolated growth hormone deficiency type I. J. Clin. Invest. 70:489–495.
9. Vogeli, G., E. V. Avvedimento, M. Sullivan, J. V. Maizel, G. Lozano, S. L. Adams, I. Pastan, and B. De Crombrugghe. 1980. Isolation and characterization of genomic DNA coding for α2 type I collagen. Nucleic Acids Res. 8:1823–1837.
10. Wozney, J., D. Hanahan, V. Tate, H. Boedtker, and P. Doty. 1981. Structure of the proα2(I) collagen gene. Nature (Lond.). 294:129–135.
11. Dagleish, R., B. C. Trapnell, R. G. Crystal, and P. Tolstoshev. 1982. Copy number of a human type I α2 collagen gene. J. Biol. Chem. 257:13816–13822.
12. Weiss, E. H., K. S. E. Cheah, F. G. Grosveld, H. H. M. Dahl, E. Solomon, and R. A. Flavell. 1982. Isolation and characterization of a human collagen α1(I)-like gene from a cosmid library. Nucleic Acids Res. 10:1981–1994.
13. Monson, J. M., J. Friedman, and B. J. McCarthy. 1982. DNA sequence analysis of a mouse proα1(I) procollagen gene: evidence for a mouse B1 element within the gene. Mol. Cell Biol. 2:1362–1371.
14. Myers, J. C., M.-L. Chu, S. H. Faro, W. J. Clark, D. J. Prockop, and F. Ramirez. 1981. Cloning a cDNA for the proα2 chain of human type I collagen. Proc. Natl. Acad. Sci. USA. 78:3516–3520.
15. Chu, M.-L., J. C. Myers, M. P. Bernard, J.-F. Ding, and F. Ramirez. 1982. Cloning and characterization of five overlapping cDNAs specific for the human proα1(I) collagen chain. Nucleic Acids Res. 10:5925–5934.
16. Bernard, M. P., J. C. Myers, M.-L. Chu, F. Ramirez, E. F. Eikenberry, and D. J. Prockop. 1983. Structure of a cDNA for the proα2 chain of human type I procollagen. Comparison with chick cDNA for proα2(I) identifies structurally conserved features of the protein and the gene. Biochemistry. 22:1139–1145.
17. Huerre, C., C. Junien, D. Weil, M.-L. Chu, M. Morabito, N. V. Cong, J. C. Myers, C. Foubert, M.-S. Gross, D. J. Prockop, A. Boue, J.-C. Kaplan, A. de la Chapelle, and F. Ramirez. 1982. Human type I procollagen genes are located on different chromosomes. Proc. Natl. Acad. Sci. USA. 79:6627–6630.
18. Junien, C., D. Weil, J. C. Myers, N. V. Cong, M.-L. Chu, C. Foubert, M.-S. Gross, D. J. Prockop, J. C. Kaplan, and F. Ramirez. 1982. Assignment of the human proα2(I) collagen structural gene (COLIA 2) to chromosome 7 by molecular hybridization. Am. J. Hum. Genet. 34:381–387.
19. Myers, J. C., L. A. Dickson, W. de Wet, M. P. Bernard, M.-L. Chu, M. DiLiberto, G. Pepe, F. O. Sangiorgi, and F. Ramirez. 1983. Analysis of the 3' end of the human proα2(I) collagen gene: utilization of multiple polyadylation sites in cultured fibroblasts. J. Biol. Chem. In press.
20. Mears, J. G., F. Ramirez, D. Leibowitz, and A. Bank. 1978. Organization of human δ- and β-globin genes in cellular DNA and the presence of intragenic inserts. Cell. 15:15–23.
21. Ott, J. 1974. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am. J. Hum. Genet. 26:588–597.

22. Lawn, R. M., E. F. Fritsch, R. Parker, G. Blake, and T. Maniatis. 1978. The isolation and characterization of linked δ- and β-globin genes from a cloned library of human DNA. *Cell.* 15:1157–1174.

23. Orkin, S. H., H. H. Kazazian, S. E. Antonarakis, S. C. Goff, C. D. Boehm, J. P. Sexton, P. G. Weber, and P. J. V. Giardina. 1982. Linkage of β-thalassemia mutations and β-globin gene polymorphisms with DNA polymorphisms in human β-globin gene cluster. *Nature (Lond.).* 296:627–631.

24. Rotwein, P. S., J. Chirgwin, M. Province, W. C. Knowler, D. J. Pettitt, B. Cordell, H. M. Goodman, and M. A. Permutt. 1983. Polymorphism in the 5′ flanking region of the human insulin gene: a genetic marker for non-insulin-dependent diabetes. *N. Engl. J. Med.* 308:65–71.

25. Steinmann, B., L. Tuderman, L. Peltonen, G. R. Martin, V. A. McKusick, and D. J. Prockop. 1980. Evidence for a structural mutation of procollagen type I in a patient with the Ehlers-Danlos syndrome type VII. *J. Biol. Chem.* 255:8887–8893.

26. Byers, P. H., R. C. Siegel, K. E. Peterson, D. W. Rowe, K. A. Holbrook, L. T. Smith, Y. H. Chang, and J. G. Fu. 1981. Marfan syndrome: abnormal α2 chain in type I collagen. *Proc. Natl. Acad. Sci. USA.* 78:7745–7749.

27. Nishioka, Y., and P. Leder. 1979. The complete sequence of a chromosomal mouse α-globin gene reveals elements conserved throughout vertebrate evolution. *Cell.* 16:875–882.

28. McKusick, V. A. 1972. Heritable Disorders of Connective Tissue. Fourth ed. The C.V. Mosby Company.

29. Hollister, D. W., P. H. Byers, and K. A. Holbrook. 1982. Genetic disorders of collagen metabolism. *Adv. Hum. Genet.* 12:1–87.

30. Prockop, D. J., K. I. Kivirikko, L. Tuderman, and N. A. Guzman. 1979. The biosynthesis of collagen and its disorders. *N. Engl. J. Med.* 301:13–23; 77–85.

31. Driesel, A. J., A. M. Schumacher, and R. A. Flavell. 1982. A *Hind*III restriction site polymorphism in the human collagen α1(I)-like gene on chromosome No. 7. *Hum. Genet.* 62:175–176.