Journal of Cellular Biochemistry 27:169–174 (1985)
Membrane Receptors and Cellular Regulation 129–134

A 38 Base Pair Insertion in the Proα2(I) Collagen Gene of a Patient With Marfan Syndrome

Elizabeth Henke, Mark Leader, Shingo Tajima, Sheldon Pinnell, and Russel Kaufman

Divisions of Genetics (E.H., R.K.), Dermatology (E.H., S.T., S.P.), and Hematology-Oncology (M.L., R.K.), Departments of Medicine and Biochemistry (R.K.), Duke University Medical Center, Durham, North Carolina 27710

Abnormalities in type I collagen have been recognized in a number of connective tissue disorders. In the Marfan syndrome, an autosomal dominant condition producing a generalized abnormality in connective tissue, no consistent abnormality has been identified, although one individual has been found to have an elongated proα2(I) collagen chain [Byers et al, Proc Natl Acad Sci USA 78:7745, 1981]. To determine the nature of the alteration in the gene that produced this abnormality, we studied the proα2(I) gene from this individual by genomic blotting and gene cloning. Genomic mapping studies detected no abnormalities. However, analysis of the cloned segment of the proα2(I) collagen gene from the Marfan individual indicates that the gene contains a 38 base pair insertion in an intron near the collagenase cleavage site. Although the relationship of this insertion to the protein abnormality is unclear, it may be a useful marker for the diagnosis of the Marfan syndrome.

Key words: Marfan syndrome, connective tissue disorders, insertion DNA, collagen, proα2(I) gene

Defects in collagen synthesis and structure produce a wide variety of disorders of connective tissue. A number of defects of type I collagen have recently been reported to produce osteogenesis imperfecta, explaining in part the clinical heterogeneity within this disorder [1–3]. Likewise, it is possible that a number of different gene defects produce the clinical manifestations of the Marfan syndrome. Patients with this disorder characteristically have arachnodactyly, ocular lens dislocation, joint laxity, and cystic medial necrosis of the aorta, but many patients do not manifest all of these abnormalities. Such variation in the clinical manifestations of this disorder occurs even among affected members of the same family in whom the gene defect is presumably identical.

Shingo Tajima's present address is Department of Dermatology, Keio University School of Medicine, Shinjuku, Tokyo 160, Japan.

Received July 5, 1984; accepted September 21, 1984.

A defect in collagen metabolism has been suspected in the Marfan syndrome because abnormally soluble collagen has been found in some patients with this disorder [4–6]. In 1981 Byers et al [7] reported a 20–25 amino acid insertion in one of the proα2(I) collagen protein chains of a patient with arachnodactyly and cystic medial necrosis of the aorta. Protein studies indicated that the insertion was near the collagenase cleavage site, which is between amino acids 775 and 776 [8] (Fig. 1). We have attempted to determine the site and nature of this defect at the DNA level in order to develop methods for detecting affected individuals before the complications of the disease are manifested. Furthermore, we were interested in determining the mechanism by which an insertion in the protein could be produced. In previous studies we and others have isolated and identified specific landmarks on the proα2(I) collagen gene [8,9] (Fig. 1). These include the exon encoding amino acids 532 to 567 in the EcoRI fragment labeled 2.8A [9], and exons encoding the collagenase cleavage site in the EcoRI fragment labeled 2.8B [8]. Immediately 5′ to the exon encoding the collagenase cleavage site is an AluI repetitive element [8,9]. We first chose to examine the gross organization of the proα2(I) gene from this individual by Southern blotting [10]. Our study of the proα2(I) collagen gene was focused on that region of DNA encoding amino acids 532–776 because studies on the abnormal protein localized the protein abnormality to that region.

MATERIALS AND METHODS

Isolation of high-molecular-weight DNA from cultured fibroblasts was by a minor modification of the method of Mears et al [11]. Analysis of high-molecular-weight DNA by Southern blot-hybridization [10] was as reported except that hybridization was performed in 10% dextran sulfate. Hybridization probes were prepared by nick translation of fragments of the normal human proα2(I) collagen gene.

A human DNA bacteriophage library was prepared by ligating the HindIII cleaved genomic DNA fragments to HindIII cleaved Charon 28. Packaging was as previously reported [9].

Sequencing of the cloned collagen fragments was performed by subcloning appropriate fragments in M13 mp8 and mp9 followed by Sanger dideoxy termination sequencing [12].

RESULTS

This patient's proα2(I) collagen gene (20 kb) was studied by EcoRI digestion of genomic DNA, Southern blotting, and hybridization to nick translated probes derived from cloned fragments 3.3, 5.2, 1.4, 2.8A, 1.2, 2.8B, and 4.0 from the normal gene (Fig. 1). No difference was observed between control and Marfan restriction fragments, indicating the absence of a large deletion or insertion in this 20 kb region of DNA (data not shown).

Based on the analysis by Byers et al [7] of the abnormal collagen peptide produced by this patient, we felt that the defect in the gene would likely lie near the collagenase cleavage site. We therefore prepared a HindIII genomic DNA bacteriophage library from the DNA of the patient with Marfan syndrome and screened that library with the 1.2 and the 4.0 kb EcoRI fragments that encode the region flanking the collagenase cleavage site. Of several recombinants identified, one, λHM-I, con-

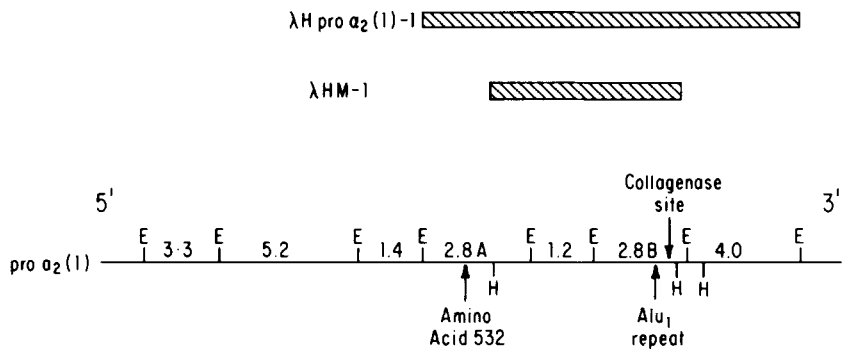

Fig. 1. Partial map of the human proα2(I) collagen gene. A 20 kb pair region of the proα2(I) collagen gene is represented in the figure and includes relevant landmarks [8,9]. The inserts from λ bacteriophage recombinants containing cloned segments of the normal collagen gene (λHproα2(I)1) [8] and the Marfan collagen gene (λHM-1) are displayed above the region of the gene to which they correspond. Restriction endonuclease sites EcoRI (E); HindIII (H).

tained a 5.2 kb HindIII fragment, which we compared by restriction endonuclease mapping to the recombinant λH proα2(I)-I representing a comparable region of the proα2(I) gene of a normal control (Fig. 1). Both recombinants were digested with EcoRI. Different mobility was observed on agarose gel electrophoresis of the corresponding 1.2 kb EcoRI fragments obtained from the digested Marfan and control recombinants. For detailed mapping studies, the 1.2 kb EcoRI fragments obtained from Marfan and control recombinants were subsequently subcloned into pUC 9 and pBR322, respectively. Further characterization of the Marfan and control 1.2 kb EcoRI fragments was performed by digestion with EcoRI and XbaI (Fig. 2A). This yielded two bands in addition to the plasmids on agarose electrophoresis; the 800 bp fragment was of equal size in both Marfan and control lanes. However, the smaller XbaI-EcoRI fragment of approximately 400 bp migrated more slowly in the Marfan DNA lane than in the control DNA lane. The slower migrating fragment was calculated to be almost 40 bp larger. The smaller (400 bp) EcoRI-XbaI fragments from Marfan and control were subcloned into M13-mp11 and M13-mp10 and the DNA sequence was determined by the Sanger dideoxy termination method [12] (Fig. 2B). A 38 bp insertion was found in the Marfan DNA 63 bp from the EcoRI site. The DNA sequence of the remainder of the 400 bp fragment was identical in Marfan and control.

DISCUSSION

We have analyzed these data to determine if the alteration in this gene is responsible for the alteration that has been described in the proα2(I) collagen chain from this patient. By necessity, the mutation would have to alter the proα2(I) mRNA to produce the protein alteration and could occur by two possible mechanisms. First, the insertion, if in an exon, could remain in the spliced mRNA and directly encode extra amino acids in the protein product. Our analysis of the entire base sequence of the EcoRI-XbaI fragments fails to identify any segment that would encode typical collagen α helical protein. This sequence should be easily identifiable because in collagen every third codon is for glycine, producing the Gly-X-Y repeating pattern.

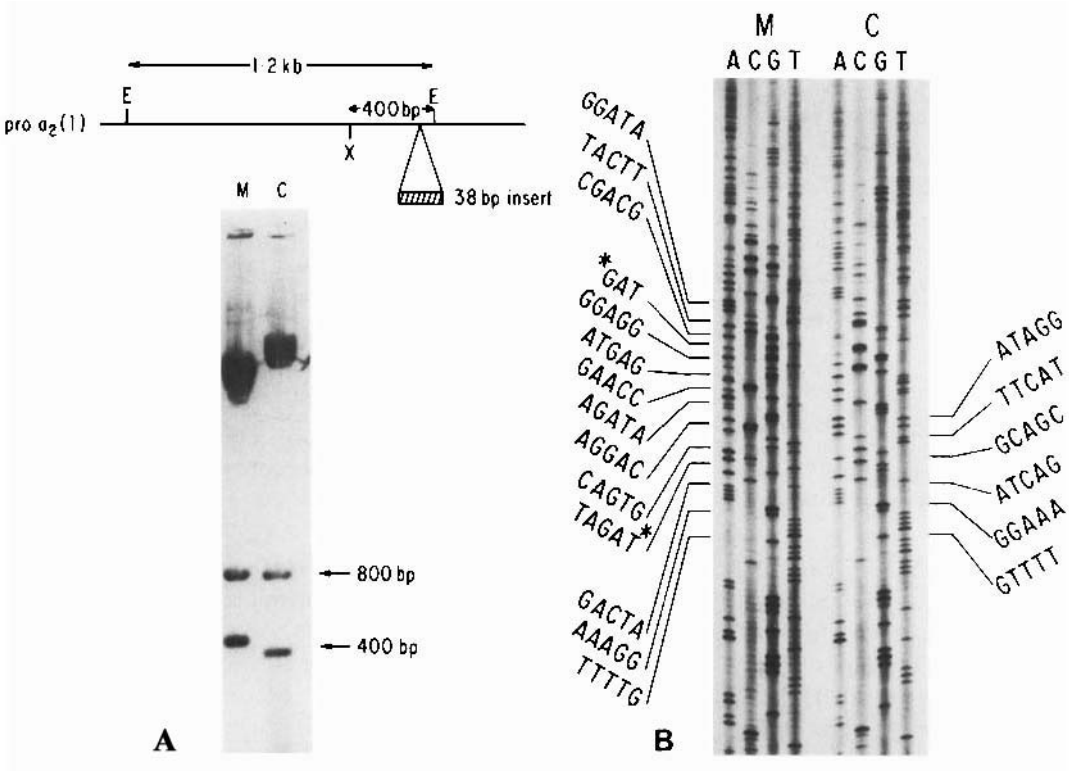

Fig. 2. A) Comparison of the 1.2 kb EcoRI fragments isolated from λHM-1 and λHproα2(I)1. Marfan (M) and control (C) 1.2 kb EcoRI fragments were digested with XbaI (X) and EcoRI (E) and separated on 1.3% agarose gel. B) partial DNA sequence of the 400 bp EcoRI-XbaI fragments from control (C) and Marfan (M) proα2(I) collagen genes demonstrating the 38 bp insertion in the Marfan DNA. The position of the 38 bp insertion in the negative DNA strand sequence of the Marfan gene is indicated by asterisks at the appropriate position beside the radioautograph of the sequencing gel. The sequence of the complementary, transcribed strand is shown here. The asterisks illustrate the position of the 38 bp insertion which is underlined below.

5′CCTATATGAAGCTGC*CTACCTCCTACTCCTTGGTCTATTCCTGGTCACATGTA*CTGATTTTCCAAAAC 3′

Futhermore, we compared the sequence of the EcoRI-XbaI fragment with the cDNA sequence, which includes this region of the gene [13]. Again, no significant homology was found. We therefore conclude that the insertion is not located in a usual exon and that the altered protein is not explained by this mechanism.

A second possible way of explaining the abnormality in this patient is that the insertion results in an alteration in precursor collagen RNA splicing. We examined the insertion for typical consensus splice site sequences and found no homology for either a donor or acceptor consensus splice site (requiring 60% homology). However, we have identified a 54 bp exon 89 bases 3′ to the site of the 38 bp insertion. We therefore searched the region surrounding the insertion for any cryptic acceptor splice sites. Five potential sites were identified, but none had a high degree of homology with the consensus acceptor sequence $(\frac{T}{C})11\frac{T}{C}N$ AG/G. It is possible that the insertion, by altering the secondary structure of the RNA, causes utilization of an alternative acceptor splice site thereby increasing the size of the resultant messenger RNA. These

possibilities can only be confirmed by functional analysis of collagen RNA derived from the patient we have studied. In general, however, mutations that affect RNA splicing, as in the β-thalassemias [14], have resulted in an unstable messenger RNA and reduced synthesis of the protein product. In the analysis of Byers et al [7], the elongated and the normally sized α2(I) chains were present in approximately equal amounts.

DNA insertions have now been reported in mouse immunoglobin genes [15], a Drosophila cuticle protein gene [16], and maize Adh1 genes [17] and have been shown to interfere with normal gene expression. A large DNA insertion has also been described in the human apolipoprotein AI gene [18] of a patient with premature atherosclerosis, although its role, if any, in the production of the disease is not known. Likewise, the alteration we describe in this gene cannot be demonstrated to have an etiologic basis for the elongated α2(I) collagen chain or the Marfan syndrome in this patient. However, the insertion may have been the result of a genetic event that introduced other changes in this gene or may be totally unrelated to the abnormality in the α2(I) collagen chain. In either case this insertion may represent an insertion polymorphism, and, as such, it could represent a useful linkage marker for the diagnosis of the Marfan syndrome if it is on the same chromosome as the mutation responsible for the Marfan phenotype.

We have attempted to determine whether or not the insertion can be detected by genomic blotting studies. We find that, though variation in the EcoRI-XbaI fragment mobility in agarose gels can be detected on cloned fragments, Southern blotting analysis of genomic DNA does not reproducibly detect differences. Furthermore, no new useful restriction endonuclease site is produced by the insertion; although new BstNI, EcoRII, MnlI, RsaI, and ScrI sites are produced, the new fragments are all less than 150 bp and are not identifiable on genomic blotting studies for technical reasons. The insertion should be identifiable with a homologous synthetic oligodeoxynucleotide probe, corresponding to the insertion sequence, if it is not a repeated sequence. Using techniques used in this study it should be possible to determine by family studies the relationship between this DNA variant and the Marfan syndrome.

ACKNOWLEDGMENTS

We would like to thank Peter Byers for providing the fibroblast line from the patient. Furthermore, we would like to thank Drs. D. Bastia and J. Ting for thoughtful comments on the manuscript and J. Cowdery for her patience in its preparation. E.H. is a recipient of a Dermatology Foundation Fellowship. R.K. is a Searle Scholar and is supported in part by a Basil O'Connor Award from the March of Dimes. These studies were also supported by NIH grants AM31672 and AM17128 and a Basic Research Grant from the March of Dimes.

REFERENCES

1. Byers PH, Shapiro JR, Rowe DW, David KE, Holbrook KA: J Clin Invest 71:689,1983.
2. Barsh GS, David KE, Byers PH: Proc Natl Acad Sci USA 79:3838, 1980.
3. Williams CJ, Prockop DJ: J Biol Chem 258:5915, 1983.
4. Scheck M, Siegel RC, Parker J, Chang Y, Fu JCC: J Anat 129:645, 1979.
5. Priest RE, Moinuddin JF, Priest JH: Nature 245:264, 1973.

6. Boucek RJ, Noble NL, Gunja-Smith Z, Butler WT: N Engl J Med 305:988, 1981.
7. Byers PH, Siegel RC, Peterson KE, Rowe DW, Holbrook KA, Smith LT, Chang Y-H, Fu JCC: Proc Natl Acad Sci USA 78:7745, 1981.
8. Myers JC, Dickson LA, deWet WJ, Bernard MP, Chu M, DiLiberto M, Pepe G, Sangiorgi FO, Ramirez F: J Biol Chem 258:10128, 1983.
9. Tajima S, Ting JPY, Pinnell SR, Kaufman RE: J Invest Derm 82:3,265, 1984.
10. Southern EM: J Mol Biol 98:503, 1975.
11. Mears JC, Ramirez R, Leibowitz R, Bank A: Cell 15:15, 1978.
12. Sanger F, Nicklen S, Coulson AR: Proc Natl Acad Sci USA 74:5463, 1977.
13. Bernard M, Myers JC, Chu M, Ramirez F, Eikenberry EF, Prockop D: Biochemistry 22:1139, 1983.
14. Dobkin C, Pergolizzi RG, Bahre P, Bank A: Proc Natl Acad Sci USA 80:1184, 1983.
15. Hawley RG, Shulman MJ, Murialdo H, Gibson DM, Hozumi N: Proc Natl Acad Sci USA 79:7425, 1982.
16. Snyder MP, Kimbrell D, Hunkapiller M, Hill R, Fristrom J, Davidson N: Proc Natl Acad Sci USA 79:7430, 1982.
17. Strommer JN, Hake S, Bennetzen J, Taylor WC, Freeling M: Nature 300:542, 1982.
18. Karathanasis SK, Zannis VI, Breslow JL: Nature 305:823, 1983.