Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 6565–6568, September 1990
Genetics

Single base mutation in the type II procollagen gene (*COL2A1*) as a cause of primary osteoarthritis associated with a mild chondrodysplasia

LEENA ALA-KOKKO*, CLINTON T. BALDWIN*, ROLAND W. MOSKOWITZ†, AND DARWIN J. PROCKOP*‡

*Department of Biochemistry and Molecular Biology, Jefferson Institute of Molecular Medicine, Jefferson Medical College, Thomas Jefferson University, Philadelphia, PA 19107-6799; and †Department of Medicine, Division of Rheumatic Diseases, Case Western Reserve University and University Hospitals of Cleveland, Cleveland, OH 44106

Communicated by Victor A. McKusick, June 11, 1990

ABSTRACT **A cosmid clone was isolated that contained an allele for the type II procollagen gene previously shown to be coinherited with primary generalized osteoarthritis in a large family. Affected members of the family had evidence of a mild chondrodysplasia, but they developed progressive osteoarthritic changes in many joints that had no epiphyseal deformities. The clone contained 52 of the 54 exons of the gene. Nucleotide sequencing of >20,000 base pairs from the clone demonstrated that all the coding sequences and all the intron–exon boundaries were normal except for a single base mutation that converted the codon for arginine at position 519 of the α1(II) chain to a codon for cysteine, an amino acid not found in type II collagen from humans or a variety of other species. The mutation was found in all affected members of the family but not in unaffected members or in 57 unrelated individuals.**

Osteoarthritis is a common disease that produces joint pain and stiffness together with radiologic evidence of progressive degeneration of joint cartilage (1). Some forms of osteoarthritis are secondary to events such as trauma, infections, metabolic disorders, or congenital or heritable conditions that deform the epiphyses or related structures. In most patients, however, there is no readily identifiable cause of osteoarthritis. At the same time, inheritance in a Mendelian dominant manner has been demonstrated in some families with primary generalized osteoarthritis, the idiopathic form of the disease characterized by the appearance of osteophytes around the distal joints of the fingers (Heberden's nodes) and progressive degeneration of cartilage in the knees, hips, and most other joints (2, 3). Two recent reports on three unrelated families (4, 5) demonstrated coinheritance of primary generalized osteoarthritis with specific alleles of the gene for type II procollagen (*COL2A1*), the precursor of the major protein of cartilage (6, 7). Here we have cloned the allele for type II procollagen previously shown to be coinherited with primary generalized osteoarthritis and a mild chondrodysplasia in one family (5). Affected members of the family began to develop joint pain and stiffness in the second and third decades and thereafter developed Heberden's nodes and typical radiographic evidence of osteoarthritis in hips, knees, shoulders, wrists, and hands (5, 8). In addition, affected members of the family had flattening of some metatarsal heads and spinal changes consisting of irregular endplates, flattening of vertebrae, and wedge-shaped deformities indicating a mild chondrodysplasia. The osteoarthritic changes, however, developed progressively in many joints that did not have any epiphyseal deformities or other evidence of a chondrodysplasia. Common causes of secondary osteoarthritis were excluded and the disease was best classified as primary generalized osteoarthritis associated with a mild chondrodysplasia. Here, we demonstrate that the allele linked to the disease had a single base mutation that converted the codon for arginine at position 519 in the α1(II) gene to a codon for cysteine, an amino acid not found in type II collagen from humans or a variety of other species (6, 7).

MATERIALS AND METHODS

Analysis by restriction fragment length polymorphisms previously demonstrated (5) that the disease in the family was coinherited with one allele for the type II procollagen gene. The allele contained a polymorphism that generated two 7-kilobase (kb) *Hin*dIII fragments from the middle of the gene instead of a single 14-kb *Hin*dIII fragment (5). For the work described here, genomic DNA was isolated from 2×10^8 cultured skin fibroblasts (9) from one affected member of the family (5). The genomic DNA was partially digested by *Sph* I, the partial *Sph* I digest was fractionated by gel electrophoresis, fragments of 25–35 kb were electroeluted from the gel, and the DNA was cloned into a cosmid vector engineered to receive *Sph* I–*Sph* I fragments of 25–35 kb (10). More than 70 positive clones were obtained by filter hybridization. One clone (see Fig. 1) containing the allele with the two 7-kb *Hin*dIII fragments (5) was plaque-purified and the DNA was isolated by cesium chloride centrifugation (11). To prepare subclones, the cosmid DNA was digested with *Sph* I and two fragments of 12 and 14 kb were isolated by agarose gel electrophoresis (10). The *Sph* I fragments were then digested with *Bam*HI or *Eco*RI. The new fragments of 2–6 kb were isolated by agarose gel electrophoresis and subcloned into the plasmids pUC19 (Boehringer Mannheim), pBluescript (Stratagene), or pBS (Stratagene). The plasmid subclones were used for double-stranded DNA sequencing by the dideoxynucleotide method (12) with specifically designed oligonucleotide primers based on cDNA (13) and genomic (14–17) sequences.

Hybridization with allele-specific oligonucleotides (ASOs) was carried out (18) on DNA synthesized by the polymerase chain reaction (19) using as a template genomic DNA that was isolated either from cultured skin fibroblasts (20), from peripheral blood leukocytes (9), or from samples of saliva (gift of H. Kuivaniemi and G. Tromp). For the polymerase chain reaction, the 5′ primer was an 18-mer that was complementary to the sequence in intron 30 of the human type II procollagen gene and the 3′ primer was a 17-mer with sequence contained in intron 32 (17). The products were blotted onto nylon-reinforced nitrocellulose membranes (Optibind; Schleicher & Schuell) and hybridized with either of two 18-mer oligonucleotides, one containing part of the normal sequence of exon 31 (ASO-Arg^{519}) and the other containing the same sequence with the single base mutation in the codon for amino acid 519 (ASO-Cys^{519}). The blots were prehybridized for 2 hr at 65°C in 6× SSC (1× SSC is 0.15 M

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: ASO, allele-specific oligonucleotide.

‡To whom reprint requests should be addressed.

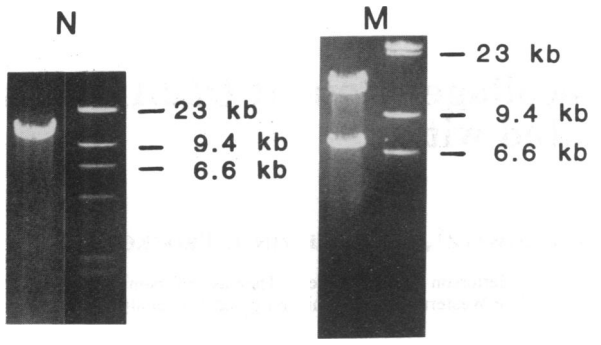

FIG. 1. Analysis of isolated cosmid clones for the type II procollagen gene from an affected member of the family. Two cosmid clones were isolated and the DNA was digested with *Hin*dIII (see text). The DNA digests were electrophoresed in an agarose gel containing 1 μg of ethidium bromide per ml. (*Left*) First lane, *Hin*dIII fragments from a clone of one allele; second lane, marker *Hin*dIII fragments from bacteriophage γ. (*Right*) First lane, *Hin*dIII fragments from a clone of the allele used here for DNA sequencing; second lane, *Hin*dIII marker fragments from bacteriophage γ. The cosmid clone without the polymorphic *Hin*dIII site (*Left*; labeled N for normal) generates three 14-kb fragments, one from the modified vector (9) and two from the type II procollagen gene (5, 13). The cosmid clone with the polymorphic *Hin*dIII site (*Right*, labeled M for mutant) generates two 14-kb fragments and two 7-kb fragments.

NaCl/0.015 M sodium citrate, pH 6.8) containing 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin, 0.5% SDS, and 0.25 mg of denatured salmon sperm DNA per ml. Hybridization was performed at 37°C for 15 hr after addition of the ^{32}P-end-labeled oligonucleotides to the same solution (11). The blots were washed three times in 2× SSC for 15 min each at room temperature (15 min per wash) and then three times in 2× SSC at 60°C (2 min per wash). The filters were exposed to x-ray film.

RESULTS

A cosmid library was prepared with genomic DNA extracted from cultured skin fibroblasts from one affected member of the family (II-5 in ref. 5). A 26-kb clone for the allele that was coinherited with the disease in the family was isolated (Fig. 1) and subcloned, and critical nucleotide sequences were analyzed. A total of >20 kb from the allele were sequenced. The sequence included all the nucleotides for exons 2B to 52. Therefore, the sequence included most of the coding sequence for the N-terminal propeptide and all the coding sequence for the N-terminal telopeptide, the triple-helical domain, the C-terminal telopeptide, and the C-terminal propeptide of the proα1(II) chain (13). The data included all 1488 codons of the gene except for the 69 codons in exon 2A that undergo alternative splicing (32) and the 27 codons in exon 1. Seven single-base differences were found between the exon sequences and the recently published sequences of overlapping cDNAs covering all the coding sequences for human type II procollagen (13). With two exceptions, the differences were third-base changes that did not alter the amino acids encoded. One exception was a difference in the codon ACC for threonine at position 869 of the α1(II) chain. The codon was found to be GCC, coding for alanine. Resequencing of the original cDNA clone (13) indicated that the original assignment of a threonine at this position was incorrect that the correct codon was GCC for alanine. The second exception was a single base change that converted the codon CGT for arginine at position 519 to TGT, a codon for cysteine (Fig. 2).

The data also provided >40 base pairs (bp) of the 5′ boundaries and >40 bp of 3′ boundaries of exons 2B to 51 in addition to >40 bp of the 5′ boundary of exon 52. All the intron–exon boundaries of the 52 exons contained normal consensus sequences for RNA splicing.

To confirm that the mutation changing the codon for Arg^{519} to a codon for cysteine was coinherited with the disease, DNA samples from affected and unaffected members of the family were examined by amplifying genomic DNA with the polymerase chain reaction and using the products for ASO hybridization (18). As indicated in Fig. 3 and Table 1, all nine affected members of the family had the mutation that converted the codon for Arg^{519} to a codon for cysteine. Ten unaffected members of the family did not have the mutation.

In addition, genomic DNA was examined from 57 individuals who were not related either to the family studied here or to one another. As indicated in Table 1, the mutation con-

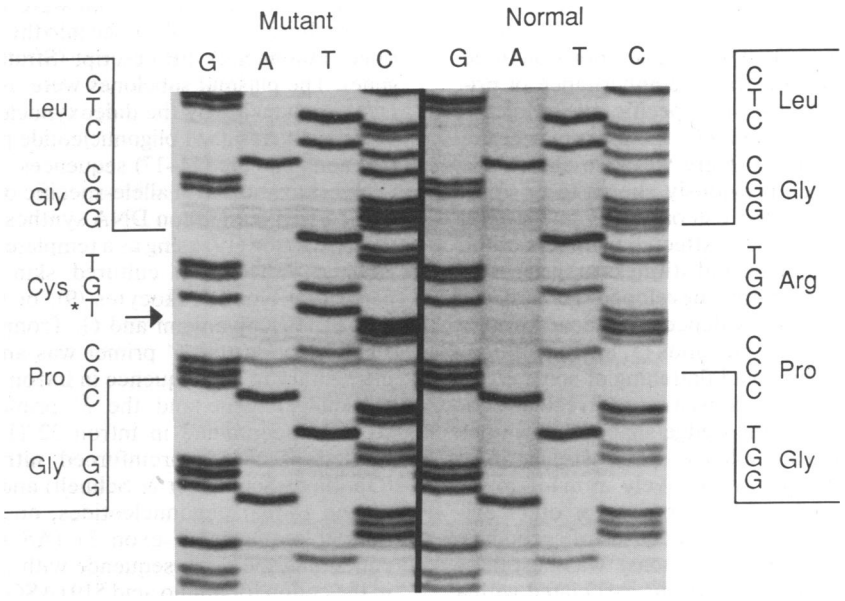

FIG. 2. DNA sequencing gel of an appropriate region of a cosmid clone containing the mutation from an affected member of the family and of the same region of a normal cosmid clone. An asterisk marks the single base change that converts the codon CGT for arginine at position 519 of the α1(II) chain to TGT for cysteine. Amino acid positions in the α1(II) chain can be converted to positions in the preproα1(II) chain by adding 131 (12).

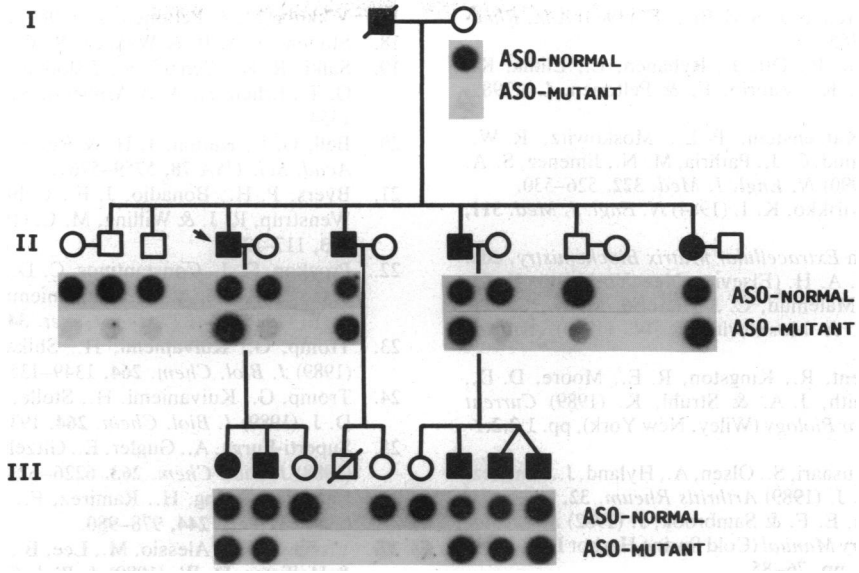

FIG. 3. ASO hybridization assay for presence of the normal sequence (ASO-normal: ASO-Arg^{519}) and the mutated sequence (ASO-mutant: ASO-Cys^{519}) in affected and unaffected members of generations I–III of the family. Genomic DNA from family members was used as a template for the polymerase chain reaction. The products of the polymerase chain reactions were then blotted on nitrocellulose filters and hybridized with the ASO probes.

verting the codon for Arg^{519} to a codon for cysteine was not found in the 114 chromosomes from the 57 individuals.

DISCUSSION

A series of single base mutations and partial gene deletions in the type I procollagen gene were shown to cause osteogenesis imperfecta and related heritable disorders of connective tissue (6, 21, 22). Similarly, a series of single base mutations (23, 24) and partial gene deletions (25) in the type III procollagen gene were shown to cause the severe variant of Ehlers–Danlos syndrome (type IV) that is associated with sudden death because of rupture of the aorta or other hollow organs. More recently, a partial gene deletion in the type II procollagen gene was found in a family with a rare chondrodystrophy (spondyloepiphyseal dysplasia) characterized by short stature and pleiotropic involvement of skeletal and occular systems (26). Also, a glycine-to-serine substitution in type II procollagen was found in a proband with a lethal form of short-limbed dwarfism (27). In addition, coinheritance of alleles for type II procollagen and the disease phenotype have been reported in several families with the Stickler syndrome (2, 28), a heritable disorder characterized by progressive blindness and joint degeneration.

The family examined here developed signs and symptoms characteristic of primary generalized osteoarthritis in the second to third decades of life (5, 8). Previous linkage studies on three generations of the family demonstrated that one allele for the type II procollagen gene was coinherited with the phenotype with a logarithm of the odds (lod) score of 3.56 at a recombination map distance of zero (5). Since a lod score (29) of >3.0 indicates a >1000:1 probability of coinheritance between a disease and an allele, the results provided a statistically significant demonstration that a mutation causing the disease was present in the type II procollagen gene or some other gene located within about 2000 kb of the gene.

Table 1. ASO hybridization for alleles of the type II procollagen gene with oligonucleotides containing the normal codon (Arg^{519}) and the mutated codon (Cys^{519})

Source of genomic DNA	Positive signals	
	ASO-Arg^{519}	ASO-Cys^{519}
Family members		
Affected	9/9	9/9
Unaffected	10/10	0/10
Unrelated individuals	57/57	0/57

The results here established that all 9 affected members of the family had an allele for type II procollagen that contained a single base mutation converting the codon for Arg^{519} of the α1(II) chain to a codon for cysteine. The mutation was not found in 10 unaffected members of the same family. Also, it was not found in 114 chromosomes from 57 individuals unrelated to the family or to one another. The presence of cysteine in a collagen molecule is readily demonstrable because the cysteines in adjacent α chains form disulfide bonds, and the resulting dimers of disulfide-linked α chains migrate as an easily identified band when the collagen is analyzed by polyacrylamide gel electrophoresis (6, 7, 30). Such analyses, as well as extensive data developed by amino acid and nucleotide sequencing, demonstrated that cysteine is not present in type II collagen from humans or a variety of species (6, 7), including species as primitive as the lamprey (31). Therefore, the mutation that introduces a cysteine into type II collagen is apparently analogous to the previously identified single base mutations that convert glycine residues in the triple-helical domain of type I procollagen to cysteine and cause either moderately severe or lethal variants of osteogenesis imperfecta (6, 21, 22, 30).

We are grateful to Dr. Sergio A. Jimenez for many helpful discussions and for providing cultures of skin fibroblasts. We thank Dr. Robert G. Knowlton for providing DNA from affected and unaffected members of the family. We also thank Dr. Knowlton as well as Drs. Helena Kuivaniemi and Gerard Tromp for providing DNA from unrelated individuals. Mr. Anthony Savarese provided expert technical assistance for much of the work reported here. The research contributions of Dr. Paul L. Katzenstein to the clinical family study are acknowledged. This research was supported in part by National Institutes of Health Grants AR39740 and AR20618.

1. Moskowitz, R. W. (1989) in *Arthritis and Allied Conditions: A Textbook of Rheumatology*, ed. McCarty, D. J. (Lea & Febiger, Philadelphia), 11th Ed., pp. 1605–1630.
2. Stecher, R. M., Hersh, A. H. & Hauser, H. (1953) *Am. J. Hum. Genet.* **5**, 46–60.

3. Kellgren, J. H., Lawrence, J. S. & Bier, F. (1963) *Ann. Rheumatic. Dis.* **22,** 237–255.
4. Palotie, A., Vaisanen, P., Ott, J., Ryhanen, L., Elima, K., Vikkula, M., Cheah, K., Vuorio, E. & Peltonen, L. (1989) *Lancet* **ii,** 924–927.
5. Knowlton, R. G., Katzenstein, P. L., Moskowitz, R. W., Weaver, E. J., Malemud, C. J., Pathria, M. N., Jimenez, S. A. & Prockop, D. J. (1990) *N. Engl. J. Med.* **322,** 526–530.
6. Prockop, D. J. & Kivirikko, K. I. (1984) *N. Engl. J. Med.* **311,** 376–386.
7. Miller, E. J. (1984) in *Extracellular Matrix Biochemistry*, eds. Piez, K. A. & Reddi, A. H. (Elsevier, New York), pp. 41–82.
8. Katzenstein, P. L., Malemud, C. J., Pathria, M. N., Carter, J. R., Sheon, R. P. & Moskowitz, R. W. (1990) *Arthritis Rheum.* **33,** 674–684.
9. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1989) *Current Protocols in Molecular Biology* (Wiley, New York), pp. 1:2.2.1–1:2.2.3.
10. Ala-Kokko, L., Kontusaari, S., Olsen, A., Hyland, J., Jimenez, S. A. & Prockop, D. J. (1989) *Arthritis Rheum.* **32,** 584.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 76–85.
12. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
13. Baldwin, C. T., Reginato, A. M., Smith, C., Jimenez, S. A. & Prockop, D. J. (1989) *Biochem. J.* **262,** 521–528.
14. Sangiorgi, F. O., Benson-Chanda, V., de Wet, W. J., Sobel, M. E., Tsipouras, P. & Ramirez, F. (1985) *Nucleic Acids Res.* **13,** 2207–2225.
15. Cheah, K. S. E., Stoker, N. G., Griffin, J. R., Gosveld, F. G. & Solomon, E. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2555–2559.
16. Su, M.-W., Benson-Chanda, V., Vissing, H. & Ramirez, F. (1989) *Genomics* **4,** 438–441.
17. Vikkula, M. & Peltonen, L. (1989) *FEBS Lett.* **250,** 171–174.
18. Studencki, A. B. & Wallace, R. B. (1984) *DNA* **3,** 7–15.
19. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230,** 1350–1354.
20. Bell, G. I., Karam, J. H. & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5759–5763.
21. Byers, P. H., Bonadio, J. F., Cohn, D. H., Starman, B. J., Wenstrup, R. J. & Willing, M. C. (1988) *Ann. N.Y. Acad. Sci.* **543,** 117–128.
22. Prockop, D. J., Constantinos, C. D., Dombrowski, K. E., Hojima, Y., Kadler, K. E., Kuivaniemi, H., Tromp, G. & Vogel, B. E. (1989) *Am. J. Med. Genet.* **34,** 60–67.
23. Tromp, G., Kuivaniemi, H., Shikata, H. & Prockop, D. J. (1989) *J. Biol. Chem.* **264,** 1349–1352.
24. Tromp, G., Kuivaniemi, H., Stolle, C., Pope, M. & Prockop, D. J. (1989) *J. Biol. Chem.* **264,** 19313–19317.
25. Superti-Furga, A., Gugler, E., Gitzelmann, R. & Steinmann, B. (1988) *J. Biol. Chem.* **263,** 6226–6232.
26. Lee, B., Vissing, H., Ramirez, F., Rogers, D. & Rimoin, D. (1989) *Science* **244,** 978–980.
27. Vissing, H., D'Alessio, M., Lee, B., Ramirez, F., Godfrey, M. & Hollister, D. W. (1989) *J. Biol. Chem.* **264,** 18265–18267.
28. Knowlton, R. G., Weaver, E. J., Struyk, A. F., Knobloch, W. H., King, R. A., Norris, K., Shamban, A., Uitto, J., Jimenez, S. A. & Prockop, D. J. (1989) *Am. J. Hum. Genet.* **45,** 681–688.
29. Morton, N. E. (1955) *Am. J. Hum. Genet.* **7,** 277–318.
30. Steinmann, B., Nicholls, A. & Pope, F. M. (1986) *J. Biol. Chem.* **261,** 8958–8964.
31. Eikenberry, E. F., Childs, B., Sheren, S. B., Parry, D. A. D., Craig, A. S. & Brodsky, B. (1984) *J. Mol. Biol.* **176,** 261–277.
32. Ryan, M. C. & Sandell, L. J. (1990) *J. Biol. Chem.* **265,** 10334–10339.