Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 6045–6047, August 1986
Genetics

Lethal osteogenesis imperfecta resulting from a single nucleotide change in one human proα1(I) collagen allele

(mutation/gene cloning/DNA sequence analysis/extracellular matrix)

DANIEL H. COHN*†, PETER H. BYERS*‡§, BEAT STEINMANN¶, AND RICHARD E. GELINAS||

Departments of *Pathology and ‡Medicine and §Center for Inherited Disease, University of Washington, Seattle, WA 98195; ¶Department of Pediatrics, Division of Metabolism, University of Zurich, Zurich, Switzerland; and ||Fred Hutchinson Cancer Research Center, Seattle, WA 98195

Communicated by Earl P. Benditt, April 28, 1986

ABSTRACT **We have characterized a mutation in a proα1(I) procollagen gene (*COL1A1*) that results in lethal (type II) osteogenesis imperfecta. The mutation is a single base change that results in a cysteine-for-glycine substitution at position 988 of the triple-helical portion of half of the α1(I) chains of type I collagen. The mutation thus disrupts the (Gly-Xaa-Yaa)$_n$ pattern necessary for triple-helix formation, where Xaa and Yaa are other amino acids. These experiments establish the minimal mutation in a type I collagen gene capable of producing lethal disease, and the lethality demonstrates a selective mechanism for the stringent maintenance of the collagen gene structure.**

Osteogenesis imperfecta (OI) is a heterogeneous group of connective tissue disorders characterized by abnormal bone formation (1–3). The most severe form of OI is the perinatal lethal form (type II), which has been shown to result from defects in type I collagen structure. The types of mutations that can result in the lethal phenotype include large deletions from genes encoding α1(I) or α2(I) chains (*COL1A1* or *COL1A2*) (4–6) and an insertion in *COL1A1* (unpublished data), but in the majority of patients there are subtle changes in the primary structure of type I collagen. The type I collagen primary translation products of cells from infants in the latter group are of uniform size. However, after post-translational processing, two populations can be resolved. One population is normal, while the other is excessively modified (increased hydroxylation and glycosylation) following translation, is poorly secreted, and has decreased thermal stability (7–10). To better understand the specific nature of the mutations in such infants, we have isolated and determined nucleotide sequences from the relevant portions of the normal and mutant proα1(I) procollagen alleles from one affected individual.

MATERIALS AND METHODS

Cloning of Genomic DNA. Thirty micrograms of fibroblast DNA from the affected individual was digested with *Bam*HI and separated by agarose gel electrophoresis. The region of the gel carrying the desired size class of molecules was resected, and 1.6 μg of DNA was purified. The DNA was ligated into *Bam*HI arms of the bacteriophage λ vector L47; 300,000 recombinants were screened by plaque-hybridization (11) with a proα1(I) genomic probe (12), and five positive clones were isolated.

DNA Sequencing. The 2.4-kilobase (kb) *Bam*HI–*Xho* I fragments from three clones were purified and subcloned into phage M13 mp18 (13). Complete digestion of replicative form DNA from these clones with *Eco*RI (in the M13 polylinker) and *Apa* I followed by repair, ligation, and transfection generated the clones used as substrates for DNA sequence analysis. DNA sequence determination was by the dideoxy chain-termination method (14).

RESULTS AND DISCUSSION

We chose cells from an infant in which we knew there had been a *de novo* appearance of a cysteine within α1(I)CB6 (the carboxyl-terminal cyanogen bromide-cleavage fragment of the triple-helical domain, containing 190 residues) in one α1(I) allelic product (8). The father was normal, while the mother had the Marfan syndrome. Because the coding sequences for the region containing the new cysteine residue are dispersed over ≈2 kb of genomic DNA, we first attempted to sequence the α1(I) mRNA directly by primer extension in the presence of dideoxynucleoside triphosphates (unpublished data). These experiments suggested that, in about half of the codons encoding triple-helical position 988, there had been a first-position base change resulting in a cysteine-for-glycine substitution that was not present in the parental cells (data not shown). However, because of the amount of nonspecific termination and because of heterozygosity, the results were equivocal. Therefore, we isolated a portion of each proα1(I) allele by molecular cloning and, using the RNA sequence information as a guide, determined their sequences separately.

The 2.0-kb region of interest is completely contained within an 11.5-kb *Bam*HI fragment (Fig. 1). Five recombinant clones carrying the fragment were identified in a bacteriophage λ library of DNA from the affected individual. We subcloned and determined sequences from three clones; two contained the normal sequence and one contained the mutant sequence (Fig. 2). The sequence data confirmed that the mutation was a single base change (G→T) in which a cysteine at triple-helical position 988 replaced a glycine (Fig. 2). The mutation disrupts the (Gly-Xaa-Yaa)$_n$ triplet necessary for triple-helix formation and substitutes a residue normally excluded from the triple helix of type I collagen.

Type I collagen is a heterotrimer that contains two α1(I) chains and one α2(I) chain. Thus, a mutation in one proα1(I) allele, such as we have described here, would result in 75% of all type I collagen molecules being abnormal by virtue of carrying at least one altered α1(I) chain. Substitution of cysteine for glycine in the triple helix would be expected to affect the way in which the propagation of triple-helix formation proceeds from the carboxyl end of the molecule beyond the substituted residue (15). Indeed, Steinmann *et al.* (8) have shown that abnormal molecules from this cell strain are overmodified for the full length of the triple helix, are secreted inefficiently, have a lower melting temperature than normal, and are unstable within the cell. As a result, a

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.

Abbreviations: OI, osteogenesis imperfecta; kb, kilobase(s).

†To whom correspondence should be addressed at: Department of Pathology, SM-30, University of Washington, Seattle, WA 98195.

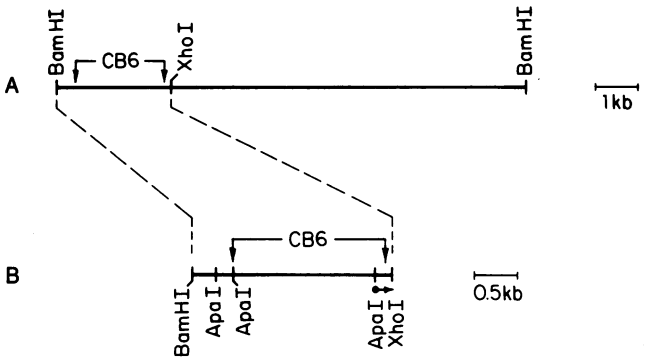
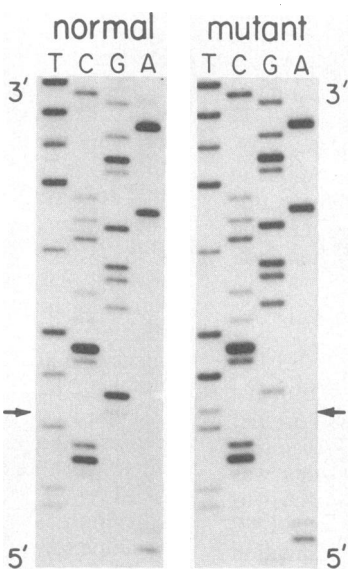


FIG. 1. Cloning and sequencing strategy used to identify the mutation. (*A*) The 11.5-kb genomic fragment containing the region of interest. The region encoding the cyanogen bromide fragment at the carboxyl terminus of the triple helix (CB6) is indicated. (*B*) The direction and extent of sequence derived from each clone is indicated by the arrow.

reduced amount of procollagen is secreted, and a significant proportion of the secreted procollagen is overmodified. The presence of overmodified molecules within the matrix possibly interferes with bone formation (24). Since about half of the assembled molecules have a reactive sulfhydryl group (because they contain only one mutant chain), these molecules could interact with other molecules within cells or in the extracellular matrix to alter bone formation. Because the presence of cysteine in the α1(I)CB6 domain in molecules that are not overmodified results in the mild OI type I phenotype (16, 17), cysteine alone is not sufficient to be lethal.

We think that the lethal phenotype results from two factors: secretion of a markedly reduced amount of type I procollagen and secretion of a significant amount of unstable and overmodified type I procollagen. We cannot assess the possible contribution of the Marfan allele to the observed phenotype because we do not know if the infant inherited the allele from his mother. However, cells from the mother do not show the observed effects on the molecular behavior of type I collagen seen in cells from the infant (8), and her cells do not contain the mutant allele (RNA sequence, data not shown). Although we cannot exclude the possibility that a single base change results in lethality only if it occurs on a permissive genetic background, substitution for a glycine would be predicted to disrupt triple-helix formation and propagation (15, 18) and, thus, can account for all of the observed biochemical changes seen in the structure and handling of type I procollagen by cells from the affected infant. Finally, we predict that, in other infants with OI type II in which no large alterations of the type I collagen genes can be found (7, 9, 10), the effects on type I collagen structure and function will usually result from heterozygosity for single amino acid substitutions for glycine within the triple-helical domain of proα1(I) or from other subtle mutations that similarly disrupt the Gly-Xaa-Yaa sequence.

Although it has been suggested that OI type II is an autosomal recessive disorder (1, 19), our results and the accumulated biochemical and genetic data (4–6, 8–10, 20–22) indicate that the phenotype often results from new dominant mutations. Recently Bonadio and Byers (10) proposed that most infants with OI type II synthesize type I collagen molecules that have disruptions in the triple-helical structure within the carboxyl-terminal half of the triple-helical domain. In conjunction with our findings in the infant described here, it appears that the proα1(I) gene of type I collagen (*COL1A1*) may present a large target for phenotypically observable mutations because any change in the first two positions of the repeated GGN-NNN-NNN nucleotide sequence that encodes the triple-helical tripeptide Gly-Xaa-Yaa is likely to be lethal if it occurs within the portion of the gene encoding the carboxyl-terminal half of the triple helix. Similar mutations in other regions of the proα1(I) gene or in the proα2(I) gene



988

EcoRI pro ***gly*** *pro arg gly arg thr gly asp ala*

5′ GAATTC CCT GGT CCT CGC GGT CGC ACT GGT GAT GCT 3′ normal

5′ GAATTC CCT TGT CCT CGC GGT CGC ACT GGT GAT GCT 3′ mutant

EcoRI pro ***cys*** *pro arg gly arg thr gly asp ala*

FIG. 2. DNA sequence derived from the normal and mutant alleles of the affected individual. Arrows mark the site of the lesion. The DNA sequences are summarized below. The last nucleotide of the *Eco*RI site marks the first nucleotide of collagen sequence.

(*COL1A2*) may have milder but recognizable phenotypic effects (3). The relatively high number of potentially lethal mutation sites in the proα1(I) gene could explain the incidence of the OI type II phenotype [at least 1/60,000 births and perhaps higher (1, 21) if, as we think, most instances are due to new dominant mutations (3, 10, 20)]. The lethal effect of this mutation provides a partial explanation for the stringency with which the coding sequences of this large gene have been maintained through evolution (23).

This work was supported in part by grants from the National Institutes of Health (AM-21557, GM-15253, AM-31232), the American Heart Association (83-1298), a Clinical Research Grant from the March of Dimes Birth Defects Foundation (6-298), and a Swiss National Foundation Grant (3.905.083). D.H.C. is the recipient of a National Institutes of Health Individual Postdoctoral Fellowship (AM-07713).

1. Sillence, D. O., Senn, A. & Danks, D. M. (1979) *J. Med. Genet.* **16,** 101–106.
2. Smith, R., Francis, M. J. O. & Houghton, G. R. (1983) *The Brittle Bone Syndrome: Osteogenesis Imperfecta* (Butterworths, London).
3. Byers, P. H. & Bonadio, J. F. (1985) in *Genetic and Metabolic Disease in Pediatrics*, eds. Lloyd, J. & Scriver, C. R. (Butterworths, London), pp. 56–90.
4. Chu, M.-L., Gargiulo, V., Williams, C. & Ramirez, F. (1985) *J. Biol. Chem.* **260,** 691–694.
5. Barsh, G. S., Roush, C. L., Bonadio, J., Byers, P. H. & Gelinas, R. E. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2870–2874.
6. Byers, P. H., Bonadio, J. F., Steinmann, B., Barsh, G. S., Holbrook, K. A., Greenberg, C., Rowe, D. W. & Gelinas, R. (1983) *Am. J. Hum. Genet.* **35,** 39A (abstr.).
7. Bateman, J. F., Mascara, T., Chan, D. & Cole, W. G. (1984) *Biochem. J.* **217,** 103–115.
8. Steinmann, B., Rao, V. H., Vogel, A., Bruckner, P. Gitzelmann, R. & Byers, P. H. (1984) *J. Biol. Chem.* **259,** 11129–11138.
9. Bonadio, J., Holbrook, K. A., Gelinas, R. E., Jacob, J. & Byers, P. H. (1985) *J. Biol. Chem.* **260,** 1734–1742.
10. Bonadio, J. & Byers, P. H. (1985) *Nature (London)* **316,** 363–366.
11. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
12. Barsh, G. S., Roush, C. L. & Gelinas, R. E. (1984) *J. Biol. Chem.* **259,** 14906–14913.
13. Messing, J., Crea, R. & Seeburg, P. H. (1984) *Nucleic Acids Res.* **9,** 309–321.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
15. Bachinger, H. P., Bruckner, P., Timpl, R. & Engel, J. (1978) *Eur. J. Biochem.* **90,** 605–613.
16. Nicholls, A. C., Pope, F. M. & Craig, D. (1984) *Br. Med. J.* **288,** 112–113.
17. Steinmann, B., Nicholls, A. C. & Pope, F. M. (1985) *Pediatr. Res.* **19,** 1076 (abstr.).
18. Bornstein, P. & Traub, W. (1979) in *The Proteins*, eds. Neurath, H. & Hill, R. L. (Academic, New York), Vol. 4, pp. 412–632.
19. Sillence, D. O., Barlow, K. K., Garber, A. P., Hall, J. G. & Rimoin, D. L. (1984) *Am. J. Med. Genet.* **17,** 407–423.
20. Byers, P. H., Bonadio, J. F. & Steinmann, B. (1984) *Am. J. Med. Genet.* **17,** 429–436.
21. Tsipouras, P., Bonadio, J. F., Schwartz, R. C., Horwitz, A. & Byers, P. H. (1985) *Am. J. Hum. Genet.* **37,** 79A (abstr.).
22. deWet, W. J., Pihlajaniemi, T., Myers, J., Kelly, T. E. & Prockop, D. J. (1983) *J. Biol. Chem.* **258,** 7721–7728.
23. Cheah, K. S. E. (1985) *Biochem. J.* **229,** 287–303.
24. Toole, B. P., Kang, A. H., Trelstad, R. L. & Gross, J. (1972) *Biochem. J.* **127,** 715–720.