Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 2870–2874, May 1985
Genetics

Intron-mediated recombination may cause a deletion in an α1 type I collagen chain in a lethal form of osteogenesis imperfecta

(gene deletions/procollagen)

GREGORY S. BARSH*, CHRISTINE L. ROUSH†, JEFFREY BONADIO*, PETER H. BYERS*‡§, AND RICHARD E. GELINAS†¶

Departments of *Pathology and ‡Medicine, and §The Center for Inherited Disease, University of Washington, Seattle, WA 98195; and †Department of Tumor Biology, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104

Communicated by Arno G. Motulsky, December 31, 1984

ABSTRACT **To understand the nature of the mutation in type I collagen genes in cells from an infant with the perinatal lethal form of osteogenesis imperfecta (type II), we cloned and sequenced almost 2 kilobases of a normal α1(I) collagen gene and the corresponding region of a mutant α1(I) gene from cell strain CRL 1262. The mutant gene had undergone recombination between two non-homologous introns, which resulted in the loss of three exons coding for 84 amino acids in the triple-helical domain. The deletion predicted the loss of amino acid residues surrounding and including the methionine at the junction between the CNBr peptides α1(I) CB8 and α1(I) CB3, a result confirmed by analysis of the cleavage peptides from the product of the mutant gene. Although large deletions from collagen genes are uncommon causes of the osteogenesis imperfecta type II phenotype, analysis of the *de novo* change in gene structure in this cell strain suggests that similar rearrangements may have occurred during the evolution of the large collagen genes.**

Osteogenesis imperfecta (OI) is a genetically and biochemically heterogeneous group of disorders affecting both mineralized and nonmineralized connective tissue (1–3). One form of this disease, the perinatal lethal form or OI type II (2), is caused by structural mutations in the proα1(I) (4, 5) or proα2(I) (6) polypeptide chains which form type I procollagen, [proα1(I)]$_2$proα2(I) (7). To further understand the molecular pathogenesis of this heterogeneous disorder, we cloned and sequenced genomic DNA from the abnormal α1(I) gene from a patient with OI type II, for which evidence for a deletion of coding sequence has been presented at the protein and mRNA levels (4, 8–11). The mutant α1(I) gene from the OI type II cells has undergone recombination between two non-homologous introns, which resulted in a 651-base-pair (bp) deletion which removes three exons coding for 84 amino acids in the triple-helical domain. Because this event changed the size of the collagen gene, we suspect that similar events may have occurred during the molecular evolution of interstitial collagen genes (12).

MATERIALS AND METHODS

Genomic Cloning Strategy. DNA from OI cells (CRL 1262, fetal fibroblasts, frozen at passage 9; American Type Culture Collection) was digested with *Bam*HI; the 5- to 7-kilobase (kb) size fraction was then purified by preparative agarose gel electrophoresis and ligated to *Bam*HI arms of λ vector L47.1 (13). After *in vitro* packaging, 10^5 recombinant bacteriophage were obtained from 0.1 μg of human DNA. This library was screened with the 2.0-kb *Bgl* II–*Hind*III probe (indicated in Fig. 1) and one of the six positive clones was plaque-purified and found to contain a 6.0-kb *Bam*HI fragment. Standard techniques for purification of DNA from fibroblasts, for genomic or bacteriophage hybridizations, and for library construction, screening, and maintenance all have been described (13, 14).

DNA Sequencing. DNA sequence was determined by the chain-termination method (15) with M13 templates (16). The 2.8-kb *Bam*HI–*Bgl* II fragment from the deletion chromosome and the 2.0-kb *Xho* I–*Bgl* II fragment from a normal chromosome were each subcloned in M13mp18 (17). To facilitate sequence analysis, an *in vitro* deletion series which started from the *Bgl* II site of each clone was generated with exonuclease III and nuclease S1 (18). *Sal* I and *Sph* I were used to digest the 2.8-kb *Bam*HI–*Bgl* II subclone and *Xma* I plus *Sph* I was used to digest the 2.0-kb *Xho* I–*Bgl* II subclone. The sequence generated from each series of M13 clones was derived from the noncoding strand; data generated from the abnormal gene series covers coordinates 30–221 and 873–1946; data generated from the normal gene series covers coordinates 1–1320 and 1630–1720. DNA sequences were compiled and analyzed with the aid of a VAX 11/780 computer.

Peptide Mapping Studies. CRL 1262 and normal fetal fibroblasts were incubated with [^{3}H]proline for 16 hr, the intracellular proteins were digested with pepsin, and two-dimensional cyanogen bromide peptide mapping was performed, all as previously described (4, 5).

RESULTS AND DISCUSSION

To devise a strategy for isolating the abnormal segment of the α1(I) gene from CRL 1262 cells (4, 7–9), we first analyzed the α1(I) genes in a series of genomic blotting experiments. Because the deletion removed a *Bam*HI site (11), digestion with *Bam*HI produced fragments of 2.2 and 4.3 kb from the normal allele but a single fragment of 6.0 kb from the abnormal allele (Fig. 1*A*). We used a genomic hybridization probe that mapped 1.0 kb to the right of the deletion to select a recombinant λ clone that contained the abnormal 6.0-kb *Bam*HI fragment. A portion of this fragment was subcloned in the single-stranded bacteriophage M13mp18 (17), and the DNA sequence was determined as described in the legends to Figs. 1 and 2. DNA sequence from the normal α1(I) gene was determined over the region that included the deletion, using a fragment subcloned from an α1(I) cosmid (Fig. 1; ref. 19).

By comparing the genomic sequences with the α1(I) cDNA sequence (21) and with the consensus sequence for "*Alu*" repetitive elements (22), we determined the intron/exon structure, the location of an *Alu* middle-repetitive ele-

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; bp, base pair(s); kb, kilobase(s).

¶To whom reprint requests should be addressed.

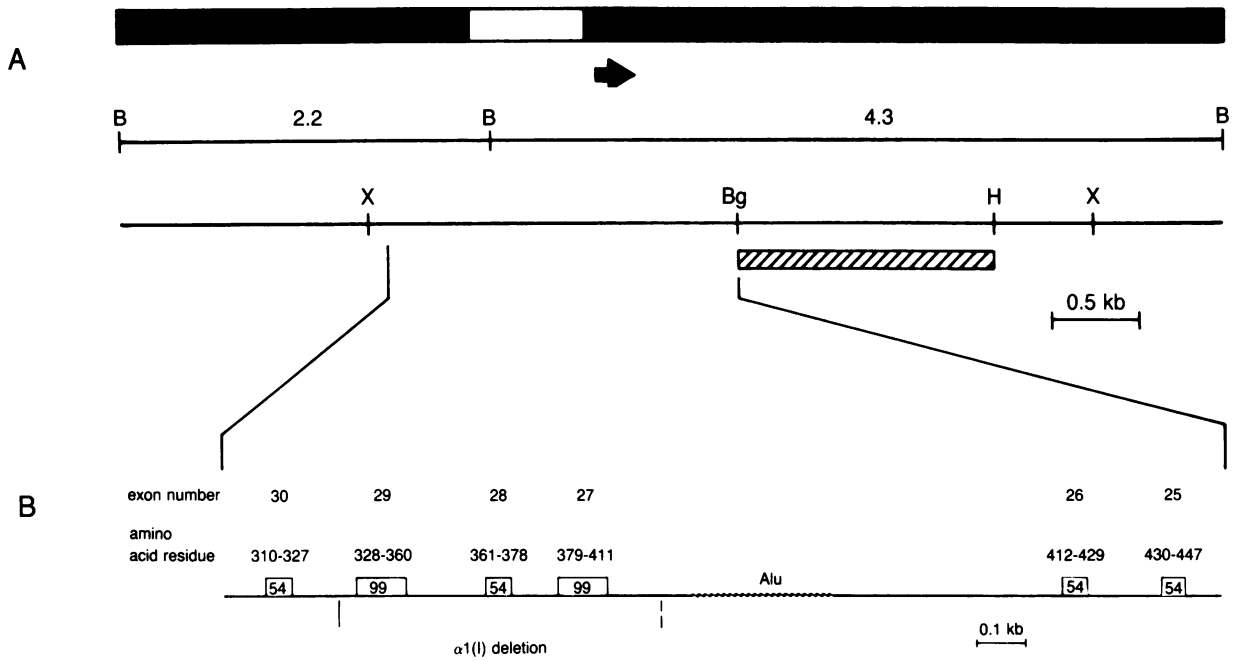


FIG. 1. Cloning strategy and exon structure of an α1(I) collagen deletion in a lethal form of osteogenesis imperfecta. (*A*) The restriction map summarizes the results from a series of genomic blot-hybridization experiments with normal human placental DNA and DNA purified from CRL 1262 cells (4, 8, 10). The bar at top represents two *Bam*HI fragments that comprise 6.5 kb from the central portion of the normal α1(I) gene, the unshaded part of the bar represents the deletion (which removes a *Bam*HI site), the hatched bar represents the hybridization probe prepared from a 4.3-kb *Xho* I subclone from an α1(I) cosmid (19), the short arrow represents an *Alu* middle-repetitive element. The region of DNA sequenced lies between the *Xho* I site and the *Bgl* II site. B, Bg, H, and X represent the restriction sites for *Bam*HI, *Bgl* II, *Hin*dIII, and *Xho* I, respectively. (*B*) Exons, introns, and the *Alu* middle-repetitive element are schematically indicated with open boxes, solid lines, and a zig-zag line, respectively. The size of each exon (in bp) is indicated within the box; each exon number (20) and the amino acid residues for which it codes (21) are indicated above. The deletion endpoints are marked with vertical lines.

ment, and the endpoints of the deletion. Over the 1946-nucleotide region sequenced, there are six exons; four are 54 bp long and two are 99 bp long (20). The sequence of these exons is identical to that determined from an α1(I) cDNA for the region coding for amino acid residues 310–447 (21), and the inferred intron/exon boundaries adhere to the consensus sequence for acceptor and donor splicing signals (23). A 280-bp *Alu* middle-repetitive element is located between exons 26 and 27 in the same transcriptional orientation as the α1(I) gene. It differs by about 10% from an *Alu* consensus sequence (22), is terminated by an imperfect 36-bp poly(A) tract, and is flanked by an imperfect 17-bp direct repeat (Fig. 2). Middle-repetitive elements have been described within the human and chicken α2(I) gene (20, 24) as well as within the mouse α1(I) gene (25), but their relation to collagen gene expression, if any, is unknown.

The exon/intron structure predicted by our DNA sequence is similar to the corresponding region of the chicken α2(I) gene (26). The number and lengths of exons coding for amino acid residues 328–429 in the chicken gene have been determined by DNA sequencing (12, 24, 26) and correspond exactly to exons 30–26 of the human α1(I) gene (Fig. 1*B* and Fig. 2). Our data indicate that amino acid residues 430–447 of the human α1(I) gene are encoded by a 54-bp exon (number 25, Fig. 1*B*), but heteroduplex studies have suggested that these residues are encoded by part of a 99-bp exon in the chicken α2(I) gene (24, 26). This apparent discrepancy may be resolved when more DNA sequence of the latter gene is available; however, it is also possible that a difference in the exon/intron structure of α1(I) and α2(I) has arisen since their divergence from a common ancestor more than 600 million years ago (4, 27, 28).

Both endpoints of the deletion in the abnormal α1(I) gene from CRL 1262 cells occur within introns (Fig. 1*B* and Fig. 2). The left-hand endpoint is 43 bp upstream of exon 29 and the right-hand endpoint is 105 bp downstream of exon 27; thus, exons 29, 28, and 27, which code for amino acids 328–411, are removed in their entirety. The *Alu* family member is intact in the mutant DNA and begins 64 bp downstream from the right-hand deletion endpoint.

Deletion of exons 29, 28, and 27 leaves the codon reading frame intact and the Gly-Xaa-Yaa tripeptide structure of the triple helical domain unaltered (albeit 84 residues shorter) and should allow normal molecular assembly. Analysis of type I procollagen produced by the OI cells (4, 10) has supported these predictions. Because one of the deleted exons includes the codon for the methionine residue at position 402, cyanogen bromide digestion of the mutant α1(I) chain should provide a fused peptide. We tested this prediction by examining the cyanogen bromide peptides (CBs) of the shortened proα1(I) chain produced by the OI cells and found that α1(I) CB8 and α1(I) CB3 were missing from the mutant α1(I) chain and were replaced with the fusion peptide α1(I) CB8-3(del) (Fig. 3). These results confirm the location of the deletion. Assuming that assortment of the normal and mutant proα1(I) gene products was random, the α1(I) chain carrying the deletion would be incorporated into 3/4 of all type I procollagen molecules synthesized. Because few of those molecules are secreted (4), the effect of the deletion is to decrease extracellular accumulation of normal type I collagen to about 1/4 the usual amount, while the secreted abnormal molecules interfere with normal fibrillogenesis (4, 10).

Although large deletions from the genes for type I collagen are uncommon as the molecular basis of the OI type II phenotype (5–7, 29–31), the findings here provide the conceptual framework for understanding the pathophysiologic mechanism for defective bone formation. In all the cell strains derived from patients with OI type II studied to date, there is defective secretion of an abnormal type I collagen molecule and increased intracellular retention and overmodification of the abnormal molecules to interfere with normal fibrillogenesis and fibril stabilization (31). This occurs regardless of the

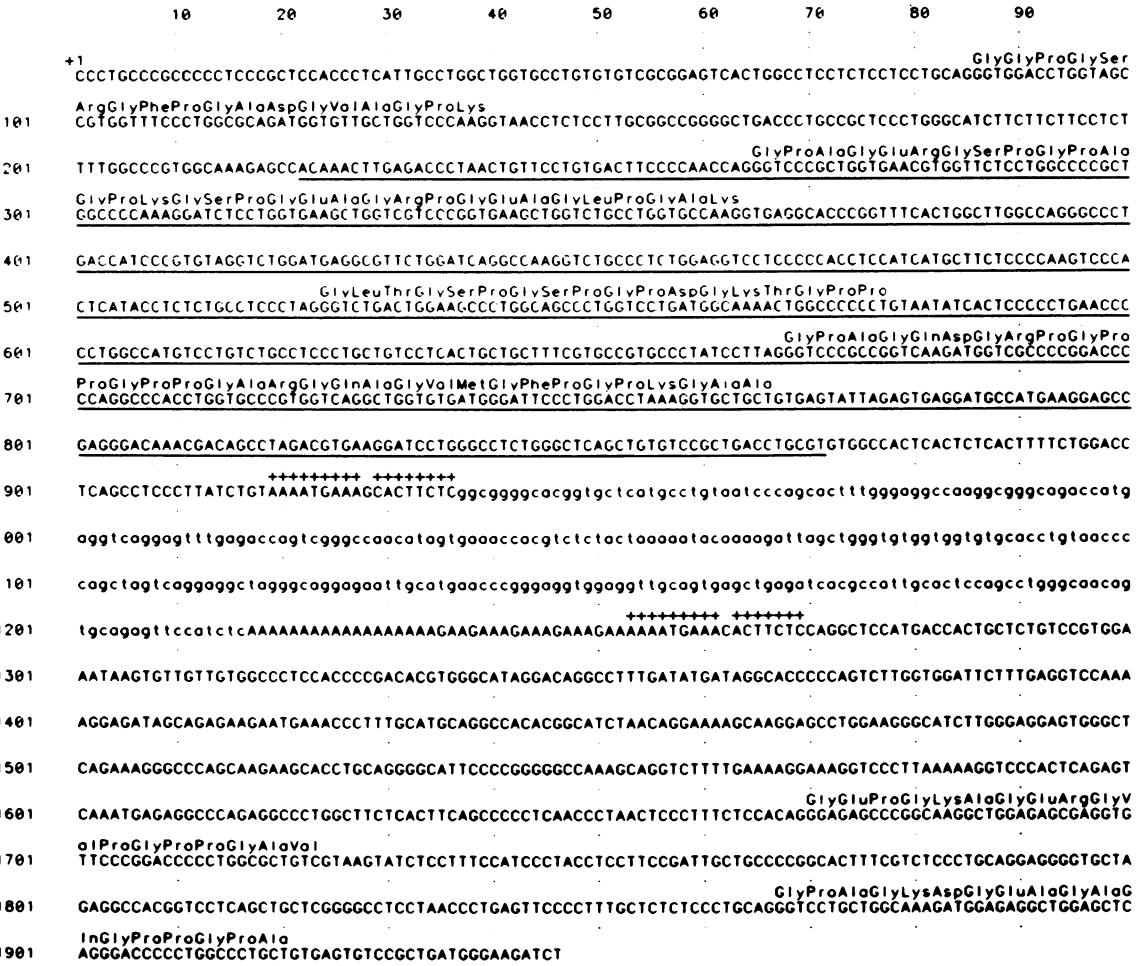
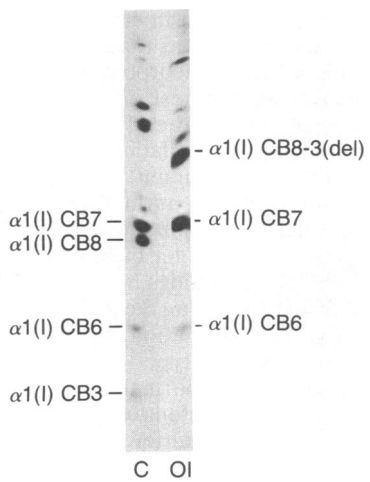


FIG. 2. DNA sequence of the α1(I) collagen gene surrounding the osteogenesis imperfecta deletion. The 1946-nucleotide sequence is a composite of information derived from the deletion chromosome and a normal α1(I) collagen gene (19). Exon boundaries were determined by comparison with the sequences of α1(I) cDNA clones Hf-677 and Hf-404 (21). The deletion removes nucleotides 222–872 and is underlined. The *Alu* middle-repetitive element was identified by comparison with a published consensus sequence (23), and its sequence is given in lower-case letters; the imperfect direct repeat on either side of the *Alu* element is indicated with plus signs. The deduced amino acid sequences for exons 29, 28, and 27 are given above the exon nucleotide sequences.

underlying mutation (7), and it is the combination of decreased secretion of normal collagen and secretion of an abnormal molecule which results in the OI type II phenotype.

To determine the mechanism by which α1(I) deletion could have occurred, we searched the sequence of the entire 1946-bp region (as well as shorter sequences at the deletion endpoints) for direct- and inverted-repeat sequence elements. We found examples of both types of features (Fig. 4).

The most striking directly repeated sequence elements present in this portion of the α1(I) gene, as revealed by dot-matrix analysis (32), were the exons themselves (Fig. 4*A*). Examination of the sequences at the deletion points showed no long (about 20 bp, Fig. 4*A*) or short (about 5 bp, Fig. 4*B*) direct repeats that might have facilitated recombination by



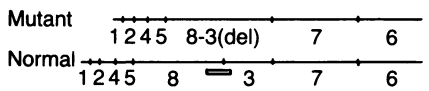


FIG. 3. Comparison of the CNBr-derived peptides from the normal and mutant collagen chain. (*Left*) CRL 1262 and control cells were incubated with [^{3}H]proline in the presence of ascorbic acid (4, 5, 29) for 16 hr, and the intracellular proteins were harvested and treated with pepsin. The collagenous proteins were separated by electrophoresis in NaDodSO$_4$/5% polyacrylamide gels. The chains were then digested in the gels with CNBr (29) and the peptides so generated were separated in a NaDodSO$_4$/12.5% polyacrylamide gel and visualized by radioautofluorography. All the large peptides from the control α1(I) chain are readily apparent (lane C), whereas the digest of the mutant chain (lane OI) lacks α1(I) CB8 and α1(I) CB3 but contains a larger peptide, α1(I) CB8-3(del). (*Right*) A diagrammatic representation of the normal and mutant α1(I) chains.

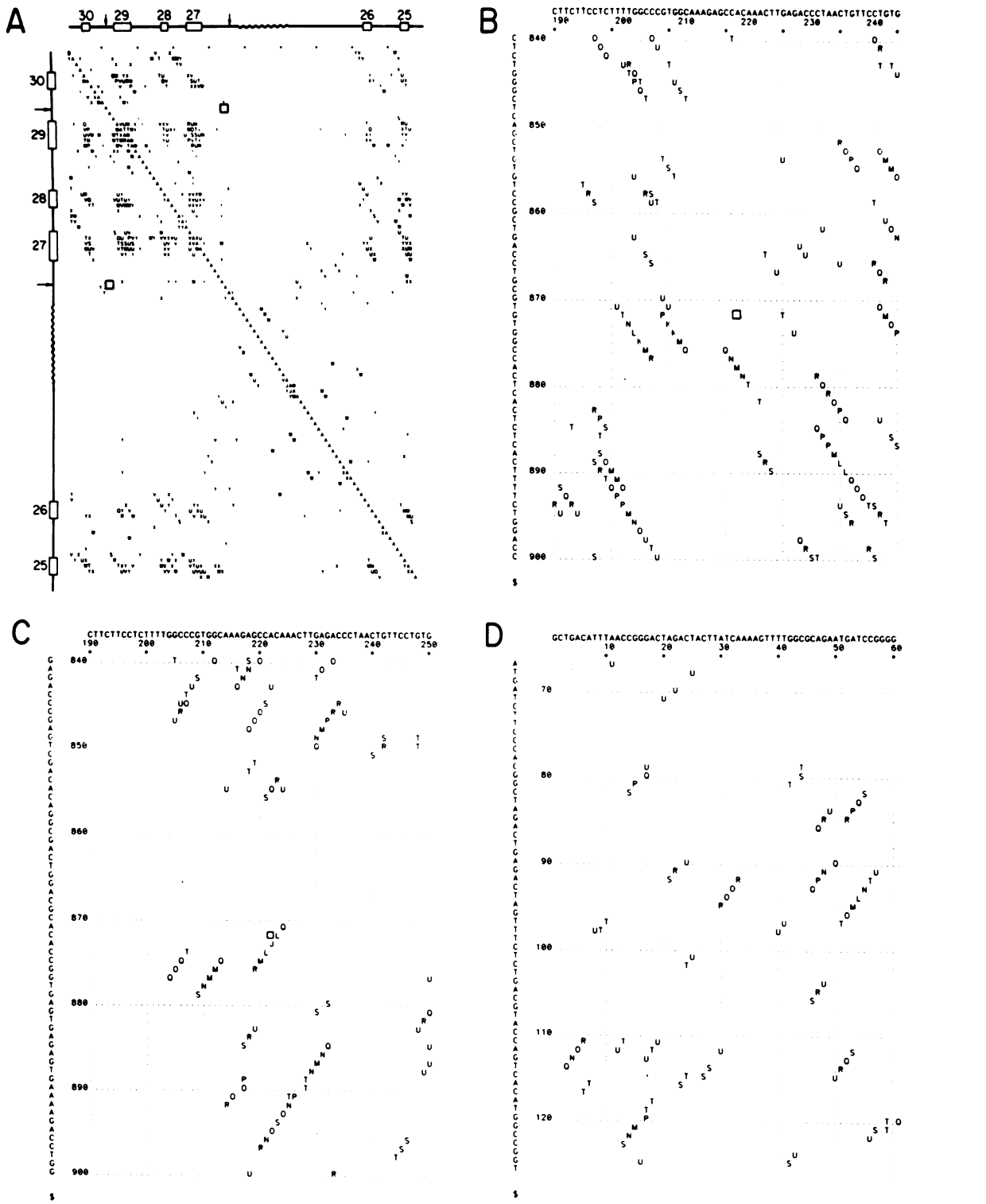


FIG. 4. Dot-matrix analysis of the DNA sequence surrounding the α1(I) collagen deletion. The matrix homology program of Pustell and Kafatos (32) was used, which indicates different levels of homology with different letters in the alphabet. With a compression factor of 1 and a range of N, the letter A indicates a perfect alignment over 2N + 1 nucleotides; the letter Y indicates an alignment over 2N + 1 nucleotides wherein the percentage of identical nucleotides at aligned positions is approximately equal to the minimum value plotted (see ref. 32 for details). The nucleotide coordinates are the same as in Fig. 2. (*A*) The 1946-nucleotide sequence from Fig. 2 compared to itself. Exons, introns, and the *Alu* repeat are schematically indicated (see Fig. 1*B*) along the abscissa and ordinate. Characters which do not lie on the diagonal line of identity represent homologous alignments; these are most prominent between different exons where they appear as a cluster of letters. The box indicates the breakpoints and alignments of the deletion chromosome; the relative absence of characters at the box indicate that a mechanism of homologous but unequal exchange is unlikely to have produced the deletion. Range, 20; scale, 0.99; minimum value plotted, 53; compression factor, 20. (*B*) the 60-bp sequence surrounding the left-hand breakpoint compared to the 60-bp sequence surrounding the right-hand breakpoint to display direct repeats. The box indicates the precise alignment of each breakpoint at coordinates 221 and 872. (*C*) The 60-bp sequence surrounding the left-hand breakpoint compared to the complement of the 60-bp sequence surrounding the right-hand breakpoint to display inverted repeats. The features of this plot are similar to those of *A* and *B* except that characters indicate potential inverted repeats. (*D*) Comparison between two 60-bp random sequences at the same level of stringency used in *B* and *C*. Program settings for B–D: range, 5; scale, 0.70; minimum value plotted, 60; compression factor, 1. With these settings, the character U indicates an 11-bp alignment in which approximately 60% of the nucleotides in aligned positions are identical.

homologous exchange (33, 34) or "slippage" during DNA replication (33, 35). The deletion endpoints are embedded in an inverted-repeat element 6 bp long (Fig. 4*C*). It has been postulated that similar short inverted repeats could lead to asymmetric mispairing of single-stranded molecules during DNA replication (33, 35). Because of the short length of this repeat (6 bp), the long distance separating its constituents (651 bases), and the occurrence of several similar features in the surrounding DNA as well as in a random sequence (Fig. 4*D*), we think that its presence at the deletion endpoints is probably coincidental and that it has no mechanistic significance. We suspect that this *de novo* recombination event probably occurred by simple breakage and reunion, unrelated to specific sequence elements, in a manner which appears to be responsible for several deletions in the β-like globin gene cluster (35, 36).

Yamada *et al.* (12) have suggested that serial duplication of a primordial 54-bp exon led to the highly repetitive structure of interstitial collagen genes. At least two different mechanisms could account for such duplication events: homologous exchange between adjacent exons, mediated by chromosome pairing, or breakage and reunion in non-homologous introns. Neither pathway easily explains the observation that many collagen exons are not multiples of 54 bp (20, 26). Nonetheless, the α1(I) mutant presented here shows that a change in exon number may result from recombination between nonhomologous introns.

Note Added in Proof. At the conclusion of the work reported here, we learned that a similar study was being concluded in the laboratory of F. Ramirez; we appreciate having been given the opportunity to compare results prior to their publication (see ref. 37).

We thank Ann Kritzberger and Toni Higgs for their help in preparing the manuscript. This work was supported in part by grants from the National Institutes of Health (AM21557, AM07171, GM07266, GM15253, and AM31232), a Clinical Research Grant from the March of Dimes Birth Defects Foundation (6-298), and a grant from the American Heart Association (83-1298). G.S.B. was a predoctoral trainee of the Medical Scientist Training Program at the University of Washington (GM07266), J.B. is supported by a post-doctoral fellowship from the National Institutes of Health (AM07171), and P.H.B. was an Established Investigator of the American Heart Association during this investigation.

1. McKusick, V. A. (1972) *Heritable Disorders of Connective Tissue* (Mosby, St. Louis, MO), pp. 292–371.
2. Sillence, D. O., Senn, A. S. & Danks, D. M. (1979) *J. Med. Genet.* **16,** 101–106.
3. Smith, R., Francis, M. J. O. & Houghton, G. F. (1983) *Brittle Bone Disease: Osteogenesis Imperfecta* (Butterworths, London).
4. Barsh, G. S. & Byers, P. H. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5142–5146.
5. Steinmann, B., Rao, V. H., Vogel, A. Bruckner, P., Gitzelmann, R. & Byers, P. H. (1984) *J. Biol. Chem.* **259,** 11129–11138.
6. deWet, W. J., Pihlajaniemi, T., Myers, J., Kelly, T. E. & Prockop, D. J. (1983) *J. Biol. Chem.* **258,** 7721–7728.
7. Byers, P. H., Bonadio, J. F. & Steinmann, B. (1984) *Am. J. Med. Genet.* **17,** 429–436.
8. Heller, R. H., Winn, K. J. & Heller, R. M. (1975) *Am. J. Obstet. Gynecol.* **121,** 572–573.
9. Penttinen, R. P., Lichtenstein, J. R., Martin, G. R. & McKusick, V. A. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 586–589.
10. Williams, C. J. & Prockop, D. J. (1983) *J. Biol. Chem.* **258,** 5915–5921.
11. Chu, M.-L., Williams, C., Pepe, G., Hirsch, J., Prockop, D. & Ramirez, F. (1983) *Nature (London)* **304,** 78–80.
12. Yamada, Y., Avvedimento, V., Mudryji, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. & deCrombrugghe, B. (1980) *Cell* **22,** 887–892.
13. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
14. Barsh, G. S., Seeburg, P. H. & Gelinas, R. E. (1983) *Nucleic Acids Res.* **11,** 3939–3958.
15. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
16. Messing, J. (1983) *Methods Enzymol.* **101,** 20–78.
17. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–265.
18. Henikoff, S. (1984) *Gene* **28,** 351–359.
19. Barsh, G. S., Roush, C. L. & Gelinas, R. E. (1984) *J. Biol. Chem.* **259,** 14906–14913.
20. Chu, M.-L., deWet, W., Bernard, M., Ding, S.-F., Morabito, M., Myers, J., Williams, C. & Ramirez, F. (1984) *Nature (London)* **310,** 337–340.
21. Bernard, M. P., Chu, M.-L., Myers, J., Ramirez, F., Eikenberry, E. & Prockop, D. (1983) *Biochemistry* **22,** 5213–5223.
22. Jelinek, W. R. & Schmid, C. W. (1982) *Annu. Rev. Biochem.* **51,** 813–844.
23. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.
24. Wozney, J., Hanahan, D., Morimoto, R., Boedtker, H. & Doty, P. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 712–716.
25. Monson, J. M., Friedman, J. & McCarthy, B. J. (1982) *Mol. Cell. Biol.* **2,** 1362–1371.
26. Tate, V., Finer, M., Boedtker, H. & Doty, P. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47,** 1039–1049.
27. Bornstein, P. & Traub, W. (1979) *The Proteins*, eds. Neurath, H. & Hill, R. L. (Academic, New York), 3rd Ed., Vol. 4, pp. 417–632.
28. Mathews, M. B. (1980) in *Biology of Collagen*, eds. Viidik, A. & Vuust, J. (Academic, New York), pp. 193–215.
29. Bonadio, J., Holbrook, K. A., Jacob, J., Gelinas, R. E. & Byers, P. H., (1985) *J. Biol. Chem.* **260,** 1734–1742.
30. Bateman, J. F., Mascara, T., Chan, D. & Cole, W. G. (1984) *Biochem. J.* **217,** 103–115.
31. Byers, P. H. & Bonadio, J., in *Butterworth's International Medical Reviews, Pediatrics*, eds. Scriver, C. R. & Lloyd, J. (Butterworths, London), Vol. 5, in press.
32. Pustell, J. & Kafatos, F. C. (1982) *Nucleic Acids Res.* **10,** 4765–4782.
33. Efstradiatis, A., Posakony, J., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R., DeRiel, J., Forget, B., Weissman, S., Slightom, J., Blechl, A., Smithies, O., Baralle, F., Shoulders, C. & Proudfoot, N. (1980) *Cell* **21,** 653–668.
34. Flavell, R. A., Kooter, J. M., deBoer, E., Little, P. F. R. & Williamson, R. (1978) *Cell* **15,** 25–41.
35. Shen, S., Slightom, J. L. & Smithies, O. (1981) *Cell* **26,** 191–203.
36. Vanin, E. F., Henthorn, P. S., Kioussis, D., Grosveld, F. & Smithies, O. (1983) *Cell* **35,** 701–709.
37. Chu, M.-L., Gargiulo, V., Williams, C. & Ramirez, F. (1985) *J. Biol. Chem.* **260,** 691–694.