

# Gene structure for the $\alpha 1$ chain of a human short-chain collagen (type XIII) with alternatively spliced transcripts and translation termination codon at the 5' end of the last exon

(intron-exon structure/extracellular matrix)

LIISA TIKKA\*<sup>†</sup>, TAINA PIHLAJANIEMI<sup>‡</sup>, PIRKKO HENTTU\*, DARWIN J. PROCKOP<sup>†</sup>, AND KARL TRYGGVASON\*

\*Biocenter and Department of Biochemistry, University of Oulu, 90570 Oulu, Finland; <sup>‡</sup>Collagen Research Unit, Biocenter and Department of Medical Biochemistry, University of Oulu, 90230 Oulu, Finland; and <sup>†</sup>Department of Biochemistry and Molecular Biology, Jefferson Institute of Molecular Medicine, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107

Communicated by Elizabeth D. Hay, July 1, 1988 (received for review April 15, 1988)

**ABSTRACT** Two overlapping human genomic clones that encode a short-chain collagen, designated  $\alpha 1$ (XIII), were isolated by using recently described cDNA clones. Characterization of the cosmid clones that span  $\approx 65,000$  base pairs (bp) of the 3' end of the gene established several unusual features of this collagen gene. The last exon encodes solely the 3' untranslated region and it begins with a complete stop codon. The 10 adjacent exons vary in size from 27 to 87 bp and two of them are 54 bp. Therefore, the  $\alpha 1$ -chain gene of type XIII collagen has some features found in genes for fibrillar collagens but other features that are distinctly different. Previous analysis of overlapping cDNA clones and nuclease S1 mapping of mRNAs indicated one alternative splicing site causing a deletion of 36 bp from the mature mRNA. The present study showed that the 36 bp is contained within the gene as a single exon and also that the gene has a 45-bp -Gly-Xaa-Xaa- repeat coding exon not found in the cDNA clones previously characterized. Nuclease S1 mapping experiments indicated that this 45-bp exon is found in normal human skin fibroblast mRNAs. Accordingly, the data demonstrate that there is alternative splicing of at least two exons of the type  $\alpha 1$ (XIII)-chain gene.

The collagen family of proteins constitutes the major structural supportive elements of the body. There are at least 13 types of collagens that can be divided into three major groups based on similarities in their protein and gene structure: the fibrillar collagens, the large nonfibrillar collagens, and the short-chain collagens (1-3).

The fibrillar collagens are a highly homologous group of proteins that share features such as a similar biosynthetic pathway, a common size for  $\alpha$  chains ( $M_r$ ,  $>95,000$ ), and a large uninterrupted triple-helical domain with a repeating -Gly-Xaa-Xaa- amino acid sequence. Based on these features, the group of fibrillar collagens includes types I, II, III, and V. Type XI collagen also appears to belong to this group (4). The fibrillar collagens also share an unusual and highly conserved gene structure in terms of the intron-exon pattern. Most of the exons encoding the triple-helical domain begin with a complete codon for glycine. Also, the exon sizes have a characteristic 54-base-pair (bp) pattern with most of the exons being 54, 108, and 162 bp and the rest being 45 or 99 bp. In addition, the sizes of specific exons are highly conserved among types I, II, III, and V collagen genes both within a given species and across a broad spectrum of species (5-8).

The large nonfibrillar collagens are characterized by  $\alpha$  chains of  $M_r$   $>95,000$  and by interruptions in the repeating -Gly-Xaa-Xaa- sequences of their triple-helical domains (1,

9). This group includes types IV, VI, and VIII collagens. The genes for the type IV collagen  $\alpha 1$  and  $\alpha 2$  chains lack the regular 54-bp exon pattern (10-14).

The group of short-chain collagens are characterized by  $\alpha$  chains of  $M_r$  considerably less than 95,000. The group includes types IX, X, and XII collagens (2, 15). The gene for type IX collagen (16) has only a few 54-bp exons (17), and the gene for the type X collagen has no intervening sequences that interrupt the triple helix coding sequences in the gene (18).

We have previously described a short-chain collagen that we here refer to as the  $\alpha 1$  chain of type XII collagen (3). At present it is not known whether this collagen type contains more than one type of  $\alpha$  chain. The predicted amino acid sequence (245 residues) at the C terminus of this polypeptide chain differs from that of the  $\alpha$  chains of types IX and X (18, 19) collagens as well as from the recently described type XII collagen (2). Antibodies against synthetic peptides derived from the predicted amino acid sequence detected two bacterial collagenase-sensitive polypeptides ( $M_r$  67,000 and  $M_r$  62,000) synthesized by a human fibrosarcoma cell line (HT-1080) and human fibroblasts. A unique feature with one of the cDNA clones encoding this protein was the presence of a 36-bp segment that was absent from the corresponding region of an overlapping clone. Nuclease S1 mapping demonstrated the presence of two different mRNAs apparently generated by alternative splicing of the primary transcript (3). Alternative splicing is well recognized for several proteins (20) including the matrix components fibronectin (21, 22) and elastin (23). However, alternative RNA splicing has not been reported for any collagen.

Here we have isolated and partially characterized the gene for the  $\alpha 1$  chain of type XIII collagen.<sup>§</sup> The data show that the 36 bp missing from some of the cDNA clones are contained within a complete exon. They also demonstrate the presence of another exon in the gene that is not found in most cDNA clones. In addition, the last exon of the gene was found to begin with a stop codon.

## MATERIALS AND METHODS

Two previously described cDNA clones, HT-125 and HT-127 (3), were labeled by nick-translation and used as probes to screen a human genomic library generously provided by Yun-Fai Lau (Division of Medical Genetics and Molecule Hematology, Department of Medicine, University of California, San Francisco). The library was constructed in the cosmid vector pJB8 with a 3.4-kilobase (kb) fragment of the

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: nt, nucleotide(s).

<sup>§</sup>The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J04085).

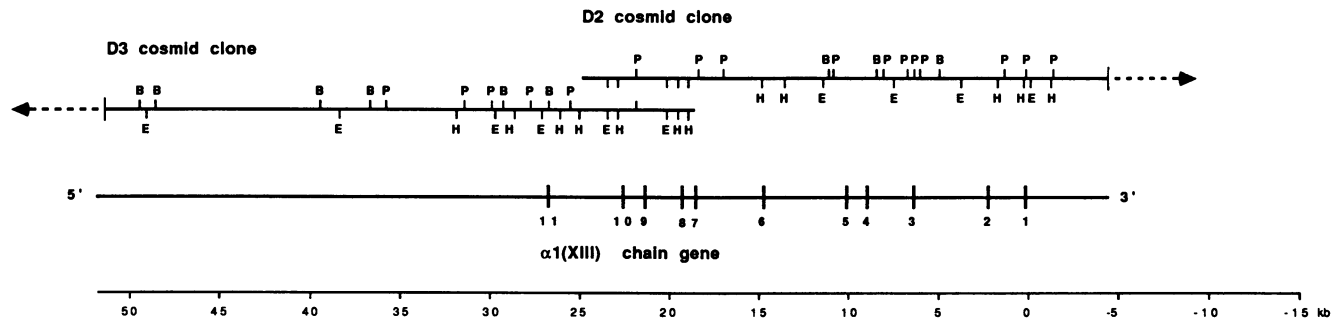


FIG. 1. Partial restriction map of two overlapping human  $\alpha 1(\text{XIII})$ -chain genomic cosmid clones and the exon-intron pattern at the 3' end of the gene. (Top) The restriction sites of *EcoRI* (E), *BamHI* (B), *PstI* (P), and *HindIII* (H) in the region studied in the D2 and D3 genomic clones are indicated. (Middle) The locations of the first 11 exons (vertical lines) from the 3' end of the gene are shown. The numbering of exons starts from the 3' end. (Bottom) The scale in kb is depicted.

herpes virus thymidine kinase gene inserted into a *Bgl* II site. Partially digested human DNA was inserted into the *Bam*HI site of the vector (24). The library screening was performed as described elsewhere (25). After preliminary restriction mapping of positive clones, fragments of two overlapping cosmid clones D2 and D3 were subcloned into *Eco*RI, *Hind*III, *Bam*HI, and *Pst* I sites of pBR322 for further mapping by endonuclease analysis and Southern blotting (25). Selected restriction endonuclease fragments containing coding sequences were cloned into M13mp18 and/or M13mp19 (26) for nucleotide sequence analysis by the dideoxy chain-termination method (27). Sequences at exon-intron boundaries were determined in most cases by using the M13 universal primer or oligonucleotide primers derived from the cDNA sequence. Oligonucleotide primers were synthesized on an Applied Biosystems DNA synthesizer at the Depart-

ment of Biochemistry, University of Oulu. Occasionally, a sequential series of overlapping clones was constructed by T4 DNA polymerase exonuclease digestion of an M13 recombinant clone using the cyclone system by International Biotechnologies (New Haven, CT). The series of M13 subclones with increasing size of deletion were sequenced by using the M13 universal primer.

A nuclease S1 protection assay was used to test human skin fibroblast mRNA for the presence of the 45-bp sequence that encoded five Gly-Xaa-Xaa- repeats upstream of exon 10. RNA was purified by the guanidine thiocyanate-CsCl procedure from cultured human skin fibroblasts and poly(A)<sup>+</sup> RNA was prepared by oligo(dT)-cellulose chromatography. The synthetic 45-nucleotide (nt) antisense oligonucleotide probe was 5'-end-labeled by polynucleotide kinase with [ $\gamma$ -<sup>32</sup>P]ATP and was used for hybridization to 10–30  $\mu$ g of

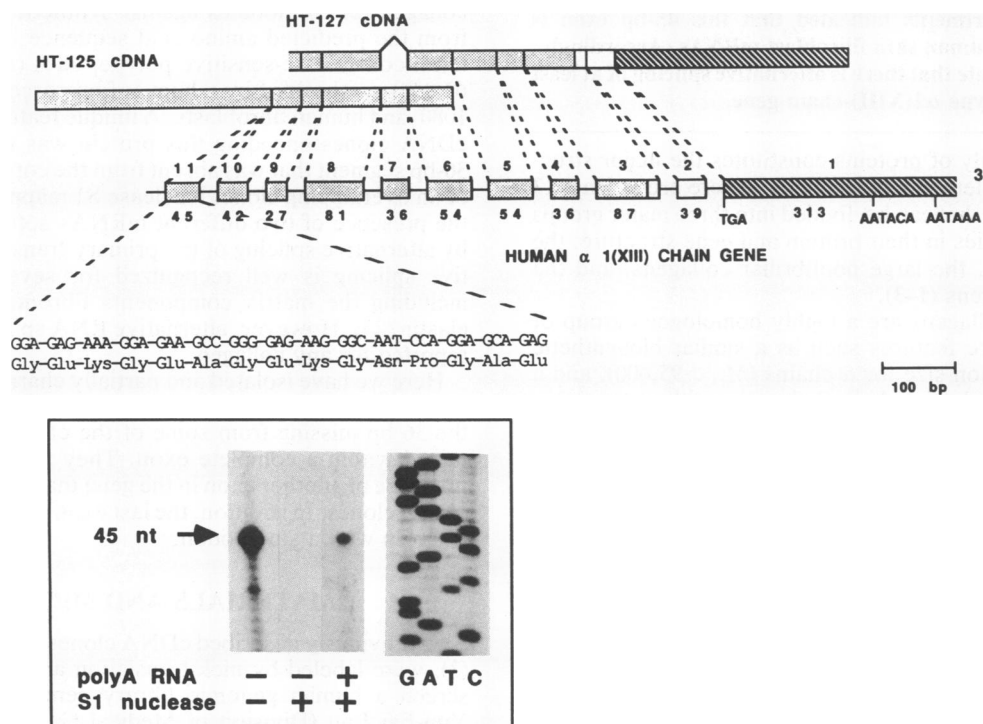


FIG. 2. Diagram showing the sizes of the first 11 exons at the 3' end of the  $\alpha 1(\text{XIII})$  collagen gene and their relationship with the HT-125 and HT-127 cDNA clones. The two cDNA clones are aligned at the top and the individual exons (boxes) of the gene are shown below on the same scale (bp). Shaded boxes indicate -Gly-Xaa-Xaa- repeat coding sequences, open boxes represent the noncollagenous domain, and hatched boxes represent 3' untranslated region coding sequences. The exon numbers are shown above and the exon sizes (bp) are indicated below. The translation termination codon (TGA) and the locations of the two polyadenylation signals (AATACA and AATAAA) are indicated. The complete sequence of exon 11 that is not present in the HT-125 cDNA is shown. Autoradiograph of a nuclease S1 protection assay is shown at the bottom. The 5'-end-labeled 45-nt oligomer probe complementary to the exon 11 transcript sequence was hybridized to fibroblast RNA (10  $\mu$ g) or tRNA (30  $\mu$ g) and digested with nuclease S1. Protected fragments were fractionated on an 8% polyacrylamide sequencing gel.

poly(A)<sup>+</sup> RNA or 30 μg of tRNA in the presence of 250 mM NaCl/40 mM Pipes, pH 6.4/0.2% sodium dodecyl sulfate. The total amount of RNA was adjusted to 30 μg with tRNA. Nuclease S1 digestion was performed as described (28). Protected fragments were analyzed on an 8% polyacrylamide sequencing gel.

**RESULTS**

**Identification of Genomic Clones.** Two overlapping cosmid clones (D2 and D3) that span ≈65,000 bp of human genomic DNA were isolated with two previously characterized cDNA clones, HT-125 and HT-127 (3). The cDNA clones contain 1000 bp that encode a 3' untranslated region, an 18-residue-long C-terminal noncollagenous domain, and 227 residues of a collagenous domain of the α1 chain of type XIII collagen. The two genomic clones isolated were characterized by restriction mapping and Southern analysis by using the cDNA probes (Fig. 1). Relevant segments of the genomic clones were subcloned and sequenced to determine the exon-intron structure. The exon nucleotide sequences were primarily determined with synthetic oligonucleotide primers that were designed based on the nucleotide sequence known from the cDNA clones but also by sequencing genomic fragments generated by the preparation of deletion clones. The sequence of 11 consecutive exons covering 812 bp, or ≈25% of the mRNA, was determined. These exons were contained in ≈25,000 bp of genomic DNA, with the introns varying between 1000 and 5000 bp.

**Exon Structure.** The sequence analysis demonstrated a unique exon size pattern in this collagen gene. The exon sizes and their relationship with the two HT-125 and HT-127 cDNA clones are depicted in Fig. 2. Exon 1 (as counted from the 3' end) encodes solely the 3' untranslated region of the

mRNA, starting with the translation termination codon TGA. The sequence extends downstream of the last polyadenylation signal found in the corresponding cDNA clone but the exact termination of the gene has not been determined. Exon 2 contains 39 bp encoding the C-terminal noncollagenous domain and exon 3 is an 87-bp junction exon that encodes five residues of the noncollagenous domain and the last eight Gly-Xaa-Xaa- repeats of the collagenous domain. Exons 4–11 encode a collagenous domain and have sizes of 36, 54, 54, 36, 81, 27, 42, and 45 bp (Fig. 2). Of the 10 protein coding exons studied here, 9 start with a complete codon for glycine and 1 (exon 10) starts with a complete codon for valine at a site of an interruption in the -Gly-Xaa-Xaa- repeat sequence.

A 45-bp exon-like -Gly-Xaa-Xaa- coding sequencing (exon 11) with complete intron junction consensus sequences at the 5' and 3' ends was found upstream of exon 10 (Fig. 2). This exon-like sequence is in the region of the gene encoding the HT-125 cDNA but is not found in that clone.

**Alternative Splicing.** The presence of alternative splicing of the primary transcript of the gene was previously suggested by the 36-bp difference in the sequence of the overlapping region of the two cDNA clones HT-125 and HT-127 (Fig. 2) and by nuclease S1 mapping (3). The present study demonstrates that this difference is caused by an alternative splicing of exon 7, which encodes the 36-bp sequence absent from the cDNA clone HT-127 (Fig. 2). The finding of a 45-bp -Gly-Xaa-Xaa- repeat coding exon-like sequence (Fig. 2, exon 11) not present in HT-125 suggested that still another alternative splicing event may occur in transcripts of this gene. To determine whether this 45-nt sequence is present in mature RNA, a nuclease S1 mapping experiment was carried out. A 45-nt-long 5'-end-labeled oligonucleotide complementary sequence (antisense) was hybridized to fibroblast RNA and the hybrids were digested with nuclease S1. Analysis of the digestion products showed full protection of the probe with

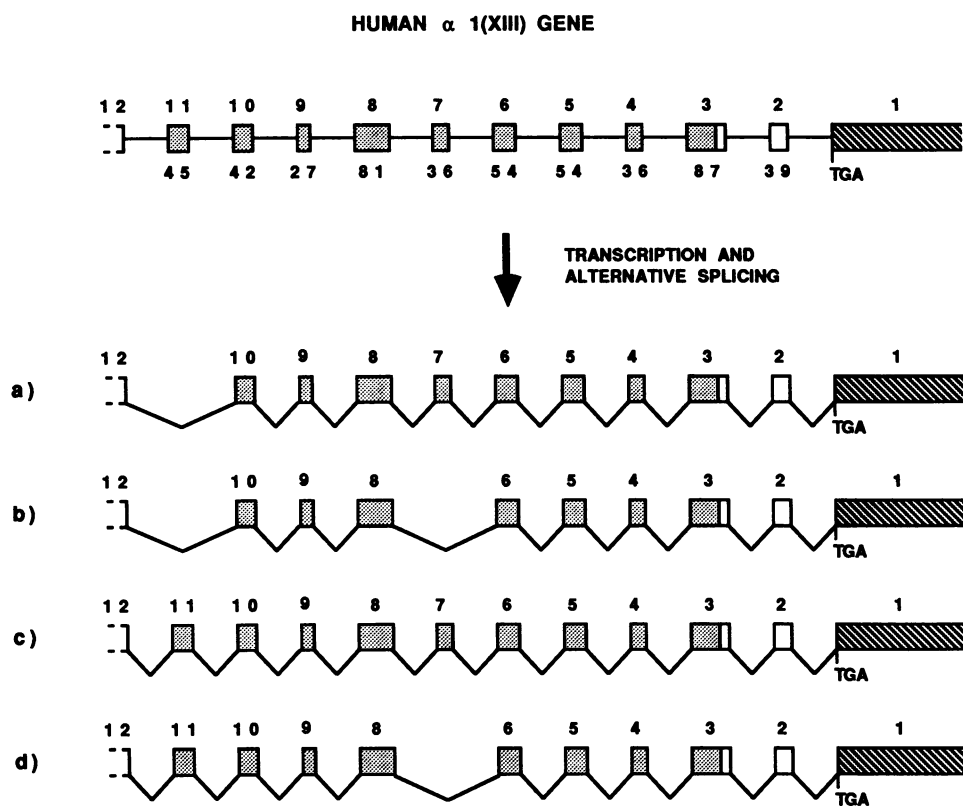


FIG. 3. Alternative splicing pattern at the 3' end of the α1(XIII) collagen gene transcripts. The four different mRNAs generated by alternative usage of the collagenous domain encoding exons 7 and 11 are indicated (a–d). (Upper) A gene diagram is shown with introns represented by thin lines and exons as boxes (as in Fig. 2).

the fibroblast RNA but not with tRNA (Fig. 2). The different transcripts generated by alternative splicing of exons 7 and 11 are shown in Fig. 3.

One explanation for different exon skipping is inefficient splicing because of unusual consensus sequences at the 5' and 3' splice sites and the putative branch point regions for lariat formation (Fig. 4). The analyses showed that seven of nine sequences determined had the 5' intron splice site consensus sequence GTRAG (where R is purine) (29). The two sequences with minor differences were GTACA and GTATG flanking exons 11 and 3, respectively, exon 11 not being found in cDNA clones HT-125 or HT-127. At the boundaries of the 3' splice sites the consensus sequence (Y)<sub>11</sub>NCAG (where Y is pyrimidine) (30) was identified in all but one case. The one exception, (Y)<sub>11</sub>TTAG, was at the boundary of exon 10 that has not been shown to be spliced out. Sequences that can be identified as putative lariat branch point consensus sequences YNYTRAY (for review, see ref. 31) were found upstream from the 3' cleavage sites (Fig. 4). The pattern of diversion from the consensus sequence did not correlate with alternative splicing of exon sequences. However, the tentative consensus sequence upstream from exon 7, which appears to be the least used, was least well conserved. It is also noteworthy that there is a 59-bp-long extremely pyrimidine-rich sequence (not shown) just preceding the putative branch point consensus sequence of that intron (number 8).

## DISCUSSION

The results presented here establish several unusual features of the human gene for the  $\alpha 1$  chain of type XIII collagen. Although two of the exons are 54 bp, none of the other exons are either 54 bp or any multiple thereof. Also, all the coding exons analyzed begin with a complete codon for an amino

acid. Hence, the  $\alpha 1$ (XIII) gene has some features found in genes for fibrillar collagens (5–7) but also features that are distinctly different. A similar pattern of exons of 54 bp and other sizes has also been observed in the gene for type IX collagen (17). One striking feature of the  $\alpha 1$ (XIII) collagen gene structure is that the last exon begins with a stop codon. No other collagen gene has a stop codon at the 5' end of its last exon, and we know of no other gene that has this arrangement. The presence of a stop codon at the beginning of the last exon obviously provides a gene structure that can generate a large number of translation products by alternative splicing of the last exon to different exons more 5' in the gene. The potential for alternatively spliced RNA transcripts and for generating protein products of different lengths is significant in a gene such as the  $\alpha 1$ (XIII) gene in which each of the interior exons ends in a complete codon for an amino acid. To date there is no evidence, however, of alternative splicing that involves the last exon, but type XIII collagen is unique among collagens in that alternative splicing of other exons clearly occurs.

The present study demonstrates that alternative splicing is a prominent feature of the gene for the  $\alpha 1$ (XIII) chain. Alternative splice of  $\alpha 1$ (XIII) gene transcripts was first suggested by the observation that one cDNA clone (HT-125) contained 36 bp of coding sequence not found in a second cDNA clone (HT-127) from the same library (3). RNA protection experiments with nuclease S1 directly demonstrated alternative splicing in that three fragments were generated with the cDNA containing the 36 bp and labeled at the 3' end of the antisense strand: a fully protected fragment of 510 nt, a shorter fragment of 426 nt, and a shorter fragment of 271 nt (figure 4 in ref. 3). The shorter fragment of 426 nt is explained by alternative splicing of the 36 bp shown here to be present in exon 7. The second mismatch could not be defined earlier because of the lack of overlapping clones.

	Intron 3' splice consensus	EXON	Intron 5' splice consensus
	yyyyyyyyyyncag/		/gtuagt
5'			3'
	++++++	*	
	caggctttcacaaggctgcatctcatgcagtggtgtttgtgccacag/GGA-GAG-AAA...11...GGA-GCA-GAG/gtacctgaga	<u>Gly-Glu-Lys</u> Gly-Ala-Glu	
	--++++		
	tcaggaggaggccacttcaaatgtaacagggtccctttgcttttag/GTT-CCT-GGG...10...GGA-CCT-CCG/gtaagtttg	Val-Pro-Gly Gly-Pro-Pro	
	-+++++		
	tgtaactagagatcagttctaatgcacctgtctttattccaatgccag/GGG-CTC-CAA...9...GGA-CCA-AG/gtaagagaag	Gly-Leu-Gln Gly-Pro-Pro	
	-+----		
	ttgtctgggacatgtccttgtgtgatctgtccctctcccttctccag/GGG-GAA-GCA...8...GGA-CTA-CCC/gtaagtacct	Gly-Glu-Ala Gly-Leu Pro	
	++-+-	*	
	tccaactctctctccccttctttgtgcaactgccccttctgctcag/GGA-GCT-TCA...7...GGG-CCA-CCG/gtgagtgctg	Gly-Ala-Ser Gly-Pro-Pro	
	-+++++		
	accagggtcccagagtggtggacctcactctgctttcttccctccag/GGT-ACT-CCA...6...GGC-GAG-AGG/gtgagtgca	Gly-Thr-Pro Gly-Glu-Arg	
		N.D./GGC-AGC-AAA...5...GGG-CTT-CCT/gtgagtctct	
		Gly-Ser-Lys Gly-Leu-Pro	
	++++++		
	tgccctgtgtgctcattaacatgaccttctctctgatctcttgag/GGT-TTA-CAT...4...GGA-AAC-CGG/N.D.	Gly-Leu-His Gly-Asn-Arg	
	++++++		
	atgatctaaagttttattgtttttctcggactaaaccattattacag/GGG-GAG-AGG...3...TGC-CCA-TTG/gtatgtttt	Gly-Glu-Arg Cys-Pro-Leu	
	-++++-		
	ggtggaccattgcaggagtgctgagccctctgcttcttctcttgag/GGG-CAA-GAT...2...TGG-AAC-AGG/gtaaggcttc	Gly-Gln-Asp Trp-Asn-Lys	
	++++++		
	ctgcagaccactgaaactaacagaacattgtttgtttgtttgttcag/TGATGCCTCTAA...1... (3' uncoding )		

FIG. 4. Nucleotide sequences at the exon-intron boundaries and flanking exons 1–11 in the  $\alpha 1$ (XIII) collagen gene. Intron sequences are indicated by lowercase and exon sequences by capital letters. Stop codons in exon 1 are underlined. The lariat branch point consensus sequence YNYTRAY is shown by + and -. Exons that are used alternatively are indicated by an asterisk. u, Purine; y, pyrimidine; ND, not determined.

Here, the nucleotide S1 experiment showed that the shorter fragment of 271 nt can be explained by a second splice involving the 45-bp exon 11 that is not present, however, in the cDNA. Further evidence for alternative splicing of exon 11 was recently obtained by analysis of cDNA clones from an endothelial cell line in that several cDNA clones contained the 45 bp of exon 11 and several did not (T.P. and M. Tamminen, unpublished observations). Therefore, the sequence of both exons 7 and 11 can be alternatively spliced to generate four different mRNAs as depicted in Fig. 3.

Alternative splicing of DNA transcripts is well established for a number of genes. Among matrix proteins, alternative splicing occurs extensively with fibronectin in which alternative splicing inserts or removes domains carrying out specific functions such as cell binding or binding to other matrix components (21, 22). Alternative splicing has also been reported for transcripts of the elastin gene (23). Here, several alternatively spliced transcripts were demonstrated for type XIII collagen, but the physiological significance of this finding is as yet unknown. The regions involved do not contain any sequences known to have potential biological roles such as Arg-Gly-Asp for integrin receptor binding or Asn-Xaa-Thr for carbohydrate attachment. Another important aspect to consider is that since all the hydrogen bonds in the triple helix are interchain, the -Gly-Xaa-Xaa- repeat sequences of the three  $\alpha$  chains must be of equal length to form a stable triple helix. This is particularly important for fibrillar collagens in which the exactly right length of the molecule is also important for fibril formation (32–34). Thus, several spontaneous mutations leading to osteogenesis imperfecta were caused by the production of shorter pro- $\alpha$  chains. The shorter chains associated with and became disulfide-bonded to normal pro- $\alpha$  chains. Because molecules containing one or two shorter pro- $\alpha$  chains did not form a stable triple helix, the trimeric complexes were degraded shortly after secretion from fibroblasts. In the case of nonfibrillar collagens like type XIII, the length of the molecules may not be critical. Alternative splicing that gives rise to  $\alpha$  chains with collagenous domains of different length in the same cell will probably give rise to molecules that have segments with -Gly-Xaa-Xaa- repeats extending into a domain that is not triple helical. It is possible that such nontriple-helical domains have an important biological function. Alternatively, the mechanisms for splicing may be so precisely regulated that an appropriate number of  $\alpha$  chains of the same length are consistently produced and therefore all the -Gly-Xaa-Xaa- sequences associate into a triple helix.

We thank Björn Olsen for providing us with unpublished data on the structure of the type IX collagen gene. This work was supported in part by grants from the National Research Council for Natural Sciences, Academy of Finland, The Finnish Cultural Foundation, and from the National Institutes of Health (AR-38188 and AR-38923).

1. Miller, E. J. & Gay, S. (1987) *Methods Enzymol.* **144**, 3–41.
2. Gordon, M. K., Gerecke, D. R. & Olsen, B. R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6040–6044.
3. Pihlajaniemi, T., Myllylä, R., Seyer, J., Kurkinen, M. & Prockop, D. J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 940–944.
4. Morris, N. P. & Bächinger, H. P. (1987) *J. Biol. Chem.* **262**, 11345–11350.
5. Boedtker, H., Finer, M. & Aho, S. (1985) *Ann. N.Y. Acad. Sci.* **460**, 85–116.
6. de Crombrughe, B., Schmidt, A., Liao, G., Setoyama, C., Mudryj, M., Yamada, Y. & McKeon, C. (1985) *Ann. N.Y. Acad. Sci.* **460**, 154–162.
7. Ramirez, F., Bernard, M., Chu, M.-L., Dickson, L., Sangiorgi, F., Weil, D., de Wet, W., Junien, C. & Sobel, M. (1985) *Ann. N.Y. Acad. Sci.* **460**, 117–129.
8. Weil, D., Bernard, M., Gargano, S. & Ramirez, R. (1987) *Nucleic Acids Res.* **15**, 181–198.
9. Timpl, R. & Dziadek, M. (1986) *Int. Rev. Exp. Pathol.* **29**, 1–112.
10. Kurkinen, M., Bernard, M. P., Barlow, D. P. & Chow, L. T. (1985) *Nature (London)* **317**, 177–179.
11. Soininen, R., Chow, L., Kurkinen, M., Tryggvason, K. & Prockop, D. J. (1986) *EMBO J.* **5**, 2821–2823.
12. Soininen, R., Tikka, L., Chow, L., Pihlajaniemi, T., Kurkinen, M., Prockop, D. J., Boyd, C. D. & Tryggvason, K. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1568–1572.
13. Sakurai, Y., Sullivan, M. & Yamada, Y. (1986) *J. Biol. Chem.* **261**, 6654–6657.
14. Hostikka, S. L. & Tryggvason, K. (1987) *FEBS Lett.* **224**, 297–305.
15. Olsen, B. R., Ninomiya, Y., Lozano, G., Konomi, H., Gordon, M., Green, G., Parsons, J., Seyer, J., Thompson, H. & Vasios, G. (1985) *Ann. N.Y. Acad. Sci.* **460**, 141–153.
16. Lozano, G., Ninomiya, Y., Thompson, H. & Olsen, B. R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4050–4054.
17. Olsen, B. R., Gerecke, D., Gordon, M., Green, G., Kimura, T., Konomi, H., Nuragaki, Y., Ninomiya, Y., Niskimura, I. & Fugrue, S. (1988) in *Collagen: Biochemistry, Biotechnology and Molecular Biology*, eds. Olsen, B. R. & Nimni, M. (CRC, Boca Raton, FL), Vol. 4, in press.
18. Ninomiya, Y. & Olsen, B. R. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3014–3018.
19. Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T. & Olsen, B. R. (1986) *J. Biol. Chem.* **261**, 5041–5050.
20. Breitbart, R. E., Andreadis, A. & Nadal-Ginard (1987) *Annu. Rev. Biochem.* **56**, 467–495.
21. Schwarzbauer, J. E., Tamkun, J. W., Lemischka, I. R. & Hynes, R. O. (1983) *Cell* **35**, 421–431.
22. Kornblihtt, A. R., Umezawa, K., Vibe-Pedersen, K. & Baralle, F. E. (1985) *EMBO J.* **4**, 1755–1759.
23. Indik, Z., Yeh, H., Ornstein-Goldstein, N., Sheppard, P., Anderson, N., Rosenbloom, J. C., Peltonen, L. & Rosenbloom, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5680–5684.
24. Lau, Y. F. & Kan, Y. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5225–5229.
25. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
26. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
28. Berk, A. J. & Sharp, P. A. (1977) *Cell* **12**, 721–732.
29. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) *Cell* **38**, 317–331.
30. Mount S. (1982) *Nucleic Acids Res.* **10**, 459–472.
31. Padgett, R. A., Grabowski, P. J., Kanarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150.
32. Prockop, D. J. & Kivirikko, K. I. (1984) *N. Engl. J. Med.* **311**, 376–386.
33. Byers, P. H. & Bonadio, J. F. (1985) in *Genetic and Metabolic Diseases in Pediatrics*, eds. Lloyd, J. & Scriver, C. R. (Butterworths, London), pp. 56–90.
34. Prockop, D. J., Kadler, K. E., Hojima, Y., Constantinou, C. D., Dombrowski, K. E., Kuivaniemi, H., Tromp, G. & Vogel, B. (1988) in *Cell and Molecular Biology of Hard Tissue*, CIBA Foundation Symposium No. 136, ed. Harnett, S. (CIBA Found., Pitman, London), in press.