Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 1568–1572, March 1986
Biochemistry

Large introns in the 3′ end of the gene for the proα1(IV) chain of human basement membrane collagen

(collagen genes/exon structure/DNA sequence)

RAIJA SOININEN*, LIISA TIKKA*, LOUISE CHOW†, TAINA PIHLAJANIEMI*, MARKKU KURKINEN*‡, DARWIN J. PROCKOP*, CHARLES D. BOYD*§, AND KARL TRYGGVASON*¶

Departments of *Biochemistry, ¶Pathology, ‡Medicine, and §Surgery, University of Medicine and Dentistry of New Jersey–Rutgers Medical School, Piscataway, NJ 08854; and †Department of Biochemistry, University of Rochester, School of Medicine and Dentistry, Rochester, NY 14642

Communicated by Elizabeth D. Hay, October 28, 1985

ABSTRACT **Using a recently characterized cDNA clone (HT-21) coding for the proα1(IV) chain of human type IV procollagen, we have isolated three clones from a bacteriophage λ Charon 4A library of human genomic DNA. The intron/exon structure of the proα1(IV) genomic clones was analyzed by heteroduplex electron microscopy and nucleotide sequencing. The analysis showed that the introns separating exons 2–9 are large and have a total length of over 12,000 base pairs (bp). Six of seven exons at the 3′ end of the gene coded for -Gly-Xaa-Yaa- repeats of the collagenous part of the chain. Five of the -Gly-Xaa-Yaa- coding exons (numbers 5–9) varied in size between 72 bp and 134 bp, and none of them were 54 bp or multiples thereof. A sixth exon (exon 4) was a junction exon containing 71 bp coding for -Gly-Xaa-Yaa- sequences and 142 bp coding for the carboxyl-terminal noncollagenous domain (NC-1). The seventh exon (exon 3, 178 bp) coded for sequences of the NC-1 domain. Five of the six -Gly-Xaa-Yaa- coding exons began with the second base coding for glycine, and only one exon began with a complete glycine codon at the 5′ end. The results (*i*) suggest that the gene for the proα1(IV) chain of human basement membrane collagen is significantly larger than the genes for fibrillar collagens and (*ii*) show that it lacks the 54-bp exon repeats characteristic of fibrillar collagen genes.**

Collagenous proteins are major constituents of the connective tissue and form the primary structural and supportive elements of the body (for review, see refs. 1 and 2). To date, at least 10 different collagen types have been identified. Collagens of types I, II, and III and probably also type V are a homologous group of proteins in that they are present in interstitial connective tissues in the form of fibrils (1–3). Type IV collagen, in contrast, is a nonfibrillar protein found only in basement membranes, where it forms a network-like structure (4). Other, more recently identified collagen types, such as VI (5), VII (6), VIII (7), IX (8), and X (9), do not appear to form cross-striated fibrils.

The genes coding for fibrillar types I, II, and III collagens have been extensively characterized from several species (10–23). These genes share a remarkable homology in the structure of exons coding for the -Gly-Xaa-Yaa- repeats of the triple helix. Each exon begins with a complete codon for glycine, and there is a marked preference for U and C in the third position of codons for glycine, proline, and alanine. About 74% of the exons of the α1(I) and α2(I) vertebrate genes have sizes of 54 base pairs (bp) or 108 bp (11). One exon is 162 bp long, and another, 36 bp (11, 12). The rest of the exons are 45 or 99 bp long. With one minor exception, the same pattern of exon sizes is present in the four genes for types I, II, and III collagen (10–23). The same pattern of exon sizes and codon preference has been conserved across species ranging from chicken to mouse and to man (10–24). These observations have led to the hypothesis that the genes of interstitial collagens evolved by duplications of an ancestral 54-bp sequence unit (10, 11).

The distinctive exon structure found in interstitial collagen genes is not, however, present in all collagen genes. Studies on two invertebrate collagen genes have revealed a distinctly different genomic structure. A *Drosophila melanogaster* gene encoding a collagen-like protein revealed no evidence of a 54-bp exon pattern (25). Instead, coding sequences for -Gly-Xaa-Yaa- repeats were found in two large exons. Also, the codon preference for U and C was not as marked as in the genes for interstitial collagens. Similarly, a collagen gene from the nematode *Caenorhabditis elegans* has no intervening sequences in the coding sequences for the characteristic -Gly-Xaa-Yaa- repeat of collagen (26). Further, >95% of third bases in codons for glycine and proline are A. More recently, Lozano *et al.* (27) demonstrated that the genes coding for the α1 and α2 chains of the chicken type IX collagen, a nonfibrillar collagen from cartilage, do not contain 54-bp exons coding for -Gly-Xaa-Yaa- repeats. The type IX genes showed no marked codon preference for glycine or proline.

Type IV collagen differs from the fibrillar interstitial collagens with respect to structure, assembly, and tissue distribution. The molecule is a trimer composed of two kinds of chains, termed proα1(IV) and proα2(IV), of molecular weight 185,000 and 170,000, respectively (4, 28). This heterotrimer has a 400-nm-long triple-helical portion with a globular domain at the carboxyl terminus (4). In contrast to interstitial collagens, type IV procollagen does not appear to be substantially processed extracellularly before it is incorporated into the matrix structure. A striking feature of this collagen type is that the -Gly-Xaa-Yaa- sequences of the triple helix are frequently interrupted by other sequences (29). Several groups have recently isolated cDNA clones coding for mouse and human type IV procollagen (30–35). Kurkinen *et al.* (35) recently reported that three exons in a genomic clone for the α2(IV) from mouse were 64, 123, and 182 bp, without any evidence of a 54-bp repeat. In the present work, we have explored further whether the differences between basement membrane and fibrillar collagen proteins are reflected in differences of their gene structures. We used a cloned cDNA for the human proα1(IV) chain (33) to isolate three overlapping genomic clones coding for the proα1 chain of human type IV collagen and examined the intron sizes and nucleotide sequence around seven exons at the 3′ end of the gene.

MATERIALS AND METHODS

Materials. Restriction endonucleases and T4 DNA ligase were purchased from International Biotechnologies (New

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

Haven, CT) and New England Biolabs and were used according to the suppliers' recommendations. The M13 mp18 cloning vectors were supplied by Bethesda Research Laboratories. DNA polymerase I (Klenow fragment), the universal primer for M13 sequencing, ^{32}P-labeled dCTP, ^{35}S-labeled dATP, and γ-^{32}P-labeled ATP were purchased from Amersham. Nitrocellulose filters were bought from Schleicher & Schuell.

DNA Clones. A library of human genomic DNA fragments, generated by digestion with *Alu* I and *Hae* III and cloned in bacteriophage λ Charon 4A vector, was kindly provided by T. Maniatis (Harvard University, Cambridge, MA). The screening of the library was performed according to Maniatis *et al.* (36). Briefly, the DNA from the bacteriophage plaques was transferred to duplicate nitrocellulose filters and then denatured, neutralized, and fixed by baking the filters at 80°C for 2 hr. The filters were prehybridized and hybridized at 68°C in 5× Denhardt's solution/0.1% $NaDodSO_4$/750 mM sodium chloride/75 mM sodium citrate, pH 7.0. The probe was the plasmid pBR322 containing an insert of the cDNA clone HT-21, which was previously shown to contain DNA sequence coding for a part of the proα1(IV) chain or human type IV procollagen (33). The probe was labeled with ^{32}P by nick-translation, and the filters were washed in 300 mM sodium chloride/0.1% $NaDodSO_4$/30 mM sodium citrate, pH 7.0, at 68°C and then were autoradiographed.

DNA Sequencing. DNA sequence determinations were carried out either by using the chemical modification procedure of Maxam and Gilbert (37) or by M13 cloning (38) and dideoxynucleotide sequencing strategies (39).

For dideoxynucleotide sequencing, the universal primer or 17-mer oligonucleotide primers synthesized in an Applied Biosystems DNA synthesizer were used. Protecting groups on the oligonucleotides were removed by incubating them in 13.3 N ammonium hydroxide at 55°C overnight. The solution was evaporated under vacuum, the sample was suspended in ethanol, and it was again evaporated three times. The sample was incubated in 80% (vol/vol) glacial acetic acid in water at room temperature for 30 min and dried. It was then suspended in ethanol and evaporated three times. The sample was dissolved in formamide and purified by preparative gel electrophoresis in 7 M urea in a 20% polyacrylamide gel. The major band detected under UV light was cut out, electroeluted into a solution containing 90 mM Tris·HCl, 90 mM sodium borate, and 20 mM EDTA at pH 8, and precipitated with ethanol and 10 mM $MgCl_2$. Four nanograms of each oligonucleotide primer was used in each annealing reaction.

Heteroduplex Analysis. DNA purified from a genomic DNA-containing recombinant λ phage, H4C-31, was digested with *Kpn* I. DNA from the genomic DNA clone H4C-30 was used as such. The cDNA clone pHT-21 was linearized with *Cla* I enzyme. The phage DNAs and the cDNA were denatured with NaOH, neutralized with Tris·HCl, and renatured in 50% (vol/vol) formamide/0.1 M Tris·HCl buffer/0.01 M EDTA, pH 8.5, at room temperature. The resulting heteroduplexes were prepared for electron microscopy as described (40, 41). The single-stranded cDNA molecules of length 7 kilobases (kb) in the same negative were used as length standards in measuring the lengths of exons and introns in the heteroduplexes.

RESULTS

Identification of Human α1(IV) Genomic Clones. The human *Alu* I/*Hae* III genomic library was screened with the cDNA clone HT-21 (33), and three clones, H4C-16, H4C-30, and H4C-31, were isolated. Digestion of H4C-31 with *Eco*RI resulted in gene fragments of 5.0, 4.2, 4.2, 1.4, and 0.25 kb (Fig. 1). Southern blot hybridization of these restriction fragments with the HT-21 cDNA showed that the 1.4- and 0.25-kb and both of the 4.2-kb fragments contained sequences homologous with HT-21. The 0.25-kb *Eco*RI restriction fragment was then subcloned into the *Eco*RI site of pBR322 and sequenced from the *Hin*dIII site by using the Maxam–Gilbert chemical sequencing technique. The resulting sequence contained 230 nucleotides, including a 72-bp-long sequence identical with bases −371 to −300 in the HT-21 cDNA clone (exon 7, see Fig. 4). Therefore the H4C-31 genomic clone was part of the human proα1(IV) type IV collagen gene. The two other clones overlapped with H4C-31. Partial restriction maps of the three clones are shown in Fig. 1. The clones H4C-16 and H4C-30 contained more base sequences from the 3′ end of the gene than did H4C-31. Accordingly, the three clones covered a total of about 19.5 kb of genomic DNA. Based on hybridization studies, the 5′ end of the HT-21 cDNA clone was found in the 3′ half of the genomic sequence (Fig. 1).

Heteroduplex Analysis. The heteroduplexes between genomic clone H4C-31 and cDNA clone pHT-21 revealed six intervening sequences and seven exons (Fig. 2). The measured lengths of the exons and introns are presented in Table 1. The orientation of the heteroduplexes was deduced from the asymmetric location of the cDNA insert in the linearized cDNA clone; i.e., the 5′ end of the cDNA is 3.6 kb from one end, and the 3′ end of the cDNA is 0.77 kb from the other end. The lengths of the single-stranded pHT-21 flanking the heteroduplex region were measured. One arm was 3.6 kb, suggesting that the genomic DNA spanned the 5′ end of the cDNA. The other arm measured about 2.6 kb, and it consisted of unhybridized cDNA and pBR322 sequences. The other two single-stranded arms flanking the heteroduplexes were also measured, and the locations of the exons in

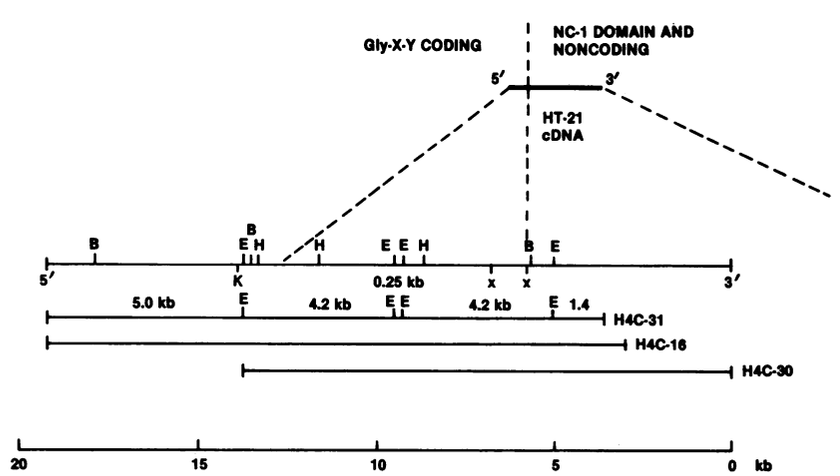

FIG. 1. Restriction map of three overlapping human proα1(IV) chain genomic clones and their relationship with cDNA clone HT-21 coding for proα1(IV). The 2.6-kb cDNA clone HT-21 is shown at the top. The 5′ end contains 0.55 kb coding for -Gly-Xaa-Yaa- repeats; the rest of the cDNA codes for the NC-1 domain and the 3′ nontranslated sequence. The next line down represents the 19 kb of genomic DNA contained in the three genomic clones. The *Eco*RI (E), *Bam*HI (B), *Hin*dIII (H), *Kpn* I (K), and *Xho* I (X) restriction sites are indicated. The sizes of the *Eco*RI (E) restriction fragments of H4C-31 are given.

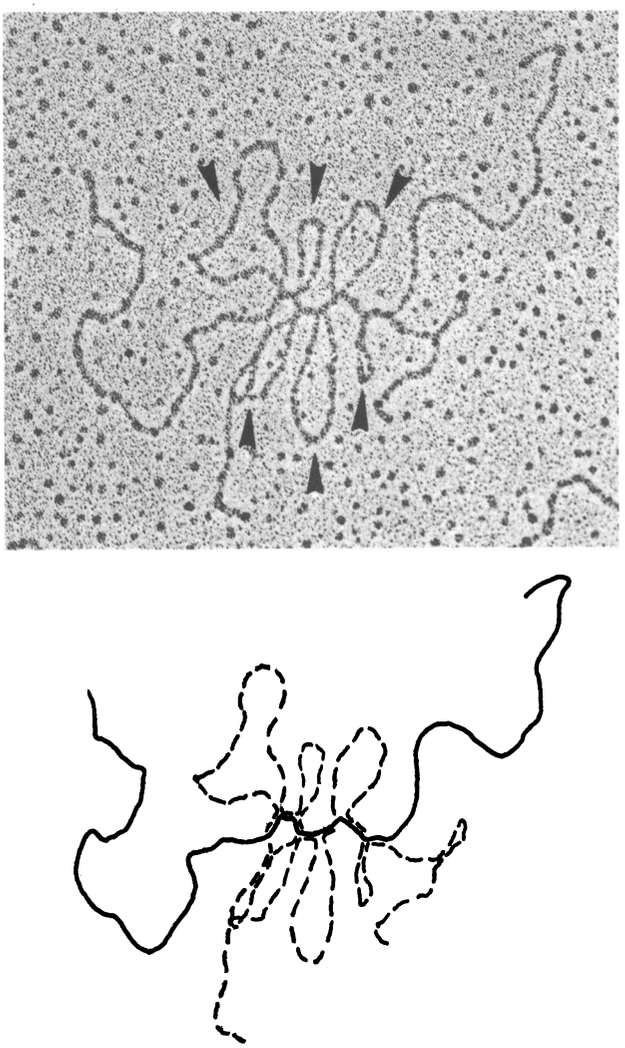

FIG. 2. Electron micrograph and interpretative tracing of a heteroduplex between cDNA pHT-21 and genomic DNA clone H4C-31. Genomic and cDNA clones were digested with restriction enzymes as described in *Materials and Methods*. Heteroduplexes were prepared as described previously (40, 41). The lengths of the exons and introns are presented in Table 1. Arrowheads point to the 6 introns. —, pHT-21 cDNA (including pBR322 sequences); ----, H4C-31 genomic DNA. The largest intron at the left is intron 8.

the H4C-31 clone were determined relative to the *Kpn* I restriction site (Fig. 2, diagram). The sizes of introns 3–8 varied in size between 960 and 2670 bp (Table 1). Selected segments of the genomic DNA clone H4C-31 were sequenced as described below. The lengths of the exons determined by DNA sequencing and analysis of the heteroduplexes are presented in Table 1. The two sets of data correlate extremely well. It can also be seen in Fig. 2 that the genomic DNA 5′ of exon 9 (see Fig. 4) contained short sequences that were inverted duplicates of others located in intron 6, a common phenomenon among eukaryotic genes. Similar experiments were performed using cDNA pHT-21 and the genomic clone H4C-30, which contained 3.5 kb of genomic DNA extending beyond the 3′ end of the clone H4C-31. The heteroduplexes had exon and intron patterns identical with those seen for H4C-31 and pHT-21 (data not shown). Thus, it appeared that intron 2 is ≥3.5 kb long, and exon 2 is located further downstream from the DNA included in clone H4C-30.

DNA Sequence Analysis. The two 4.2-kb, the 0.25-kb, and the 1.4-kb *Eco*RI fragments from the H4C-31 clone were subcloned into pBR322. The more 3′ fragment of the two 4.2-kb fragments was cut with *Eco*RI and *Xho* I and then subcloned into M13 mp18. The other fragments were subcloned directly into M13 mp18 for DNA sequencing by the dideoxy method. The resulting M13 recombinants were checked for their identity by Southern hybridization with the HT-21 cDNA. About 300 bp at the ends of the inserts were sequenced using the universal primer. After the first two exons (numbers 3 and 7 in Fig. 4) were identified with this approach, adjacent exons were sequenced by "exon hopping" with synthetic oligonucleotides identical or complementary to coding sequences in the cDNA. For example, a 17-mer complementary to the coding sequences beyond the 5′ boundary of exon 3 was used to prime the M13 insert containing exon 4 (the 3′ 4.2-kb fragment in Fig. 1). Similarly, coding sequences adjacent to the coding sequences of exon 7 were used to design appropriate primers for determining the sequences of exons 6, 8, and 9. The boundaries of exon 5 were defined by establishing the boundaries of exons 4 and 6. Heteroduplex analysis (Fig. 2) showed that the coding sequences assigned to exon 5 were in fact in one exon.

Table 1. Sizes of exons and introns in a part of the human proα1(IV) collagen gene, as determined by heteroduplex and DNA sequence analyses

Gene segment	Length, bp: Heteroduplex analysis	Length, bp: DNA sequencing
Intron 2	>3500*	
Exon 3	180	178
Intron 3	960	
Exon 4	240	213
Intron 4	1210	
Exon 5	110	99
Intron 5	1500	
Exon 6	130	129
Intron 6	1390	
Exon 7	70	72
Intron 7	960	
Exon 8	70	73
Intron 8	2670	
Exon 9	155	134

The heteroduplex analysis was carried out using the HT-21 proα1(IV) chain cDNA clone (33) and the H4C-31 genomic clone (see text and Fig. 2).
*Obtained from heteroduplexes with the H4C-30 genomic clone (see text).

The 5′ end of exon 9 contained 23 nucleotides that extended beyond the 5′ end of the HT-21 cDNA clone (Fig. 3). These nucleotides coded for the same amino acid sequence that was found by amino acid sequencing of the human α1(IV) chain from placenta (29). One nucleotide difference was found between the genomic sequence (exon 9) and the cDNA clone. Instead of a G in position −553 of the cDNA, an A was observed in the genomic DNA. Both codons coded for proline.

DISCUSSION

The results presented here establish several features of the α1(IV) gene. One is that the introns separating exons 2–9 are consistently large. The smallest intron is 960 bp, and the largest, >3500 bp. The total length of introns 2–8 is >12,000 bp. In contrast, comparable introns of the human genes for fibrillar types I, II, and III collagen chains are considerably shorter. Introns 2–8 of the human proα1(III) gene are about 3500 bp, and the same introns of genes for proα1(I), proα2(I), and proα1(II) are even smaller (refs. 18, 21, and 22; Fig. 4). Another feature is that no 54-bp pattern is apparent in the six exons (numbers 4–9) coding for -Gly-Xaa-Yaa- repeats of the

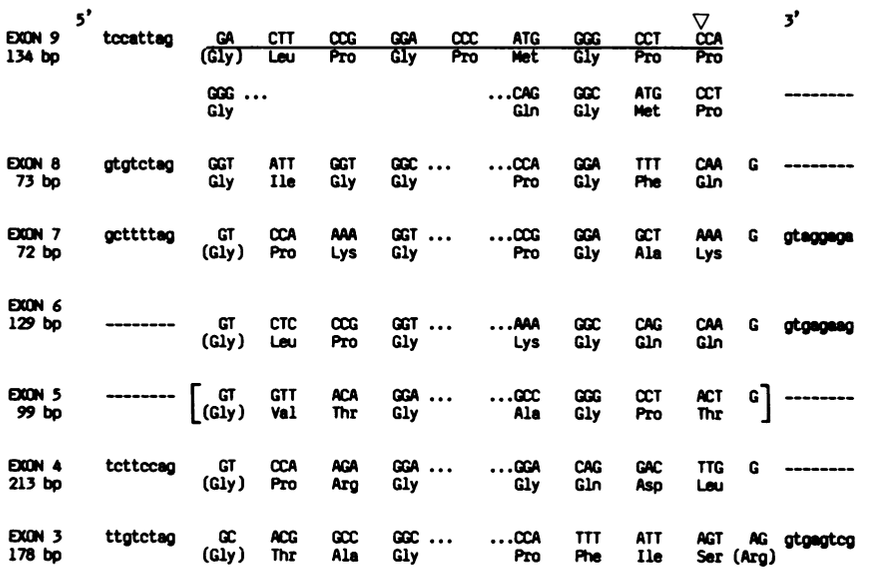

FIG. 3. DNA sequence at intron–exon junctions of exons 3–9 (see Fig. 4) of the human proα1(IV) collagen gene. Lowercase letters are intron nucleotide sequences. The underlined nucleotide sequence in exon 9 is located 5′ of the HT-21 cDNA (33) and has not been published previously. Amino acids in parentheses indicate translation at sites of truncated codons. The sequence of exon 5 is shown in brackets because it was predicted from the heteroduplex analysis and sequencing of adjacent exons (see text). Broken lines indicate intron sequences not determined. The 5′ end of the HT-21 cDNA clone is indicated by the open triangle.

collagen domain of the protein. Exons 5–9 vary in size between 72 bp and 134 bp, and an additional 71 bases code for -Gly-Xaa-Yaa- in the junction exon (number 4, Fig. 4). None of the exons have the same size as the three exons of 64, 123, and 182 bp found in the mouse gene for the proα2(IV) chain (35). However, it is not known whether the exons from the mouse α2(IV) gene code for the same region of the protein as any of the exons examined here. Thus, it remains to be seen whether or not the exon sizes in the two chains are conserved as they are in interstitial collagen genes (11, 12, 14–22). The exon patterns in basement membrane and fibrillar collagen genes are different, but the overall coding capacity within comparable regions of the genes is much the same. Exons 3–9 of the human proα2(IV) gene contain a total of 898 bp, whereas exons 3–9 of the human proα2(I) and proα1(III) genes total 877 and 918 bp, respectively. Therefore, approximately the same amount of coding information is present in seven exons of basement membrane and fibrillar collagen genes, but the length of the region in the proα1(IV) gene is almost 3 times that of the same region in the genes of fibrillar collagens.

Still another feature of the proα1(IV) gene is that five out of the six exons begin with the second base coding for a glycine (Fig. 3). Only one, exon 8, begins with a complete codon for glycine. Therefore, there is no preservation for complete glycine codons at the 5′ end of most exons as seen in genes for fibrillar collagens (10, 11, 14–22). Further, the gene contains a junction exon at the 3′ end of the repeat-coding region of the gene, similar to the junction exons for fibrillar collagens. The -Gly-Xaa-Yaa- coding sequences in the junction exon account for about one-third of the base pairs in this exon; codons for -Gly-Xaa-Yaa- account for about one-fifth of the junction exons in fibrillar collagens. As reported previously (32, 33), the amino acid sequence of the NC-1 domain of the α1(IV) chain has unusual homology between the first half and the second half of the structure. The homology involves all 12 cysteine residues and divides the protein into six subdomains, with the first three highly homologous to the second three. The results presented here show that the first of the subdomains is entirely within the junction exon, exon 4. The second subdomain and part of the third are within exon 3. These results suggest that the homology in the two halves of the NC-1 domain will not be reflected in a distinctive exon pattern, but further information will be required to establish this point.

The absence of a 54-bp exon pattern in type IV collagen as well as in type IX (27) and in the *D. melanogaster* (25) and *C. elegans* (26) collagen genes further demonstrates that the 54-bp exon suggested to be a primordial exon for all -Gly-Xaa-Yaa- coding sequences is not retained in all collagen genes. The 54-bp exon may in fact be a characteristic only of collagens that form distinctive fibrils. As discussed elsewhere

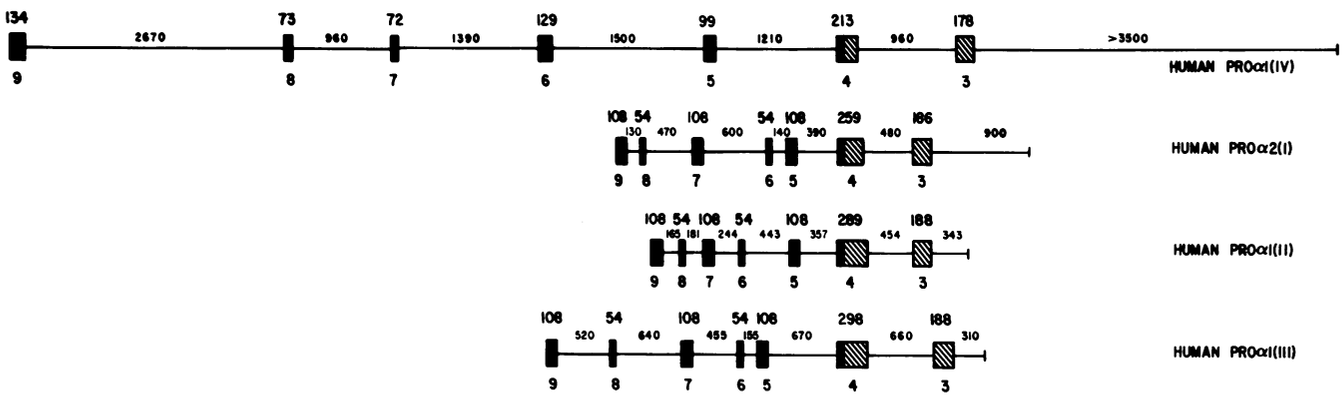

FIG. 4. Relationship of introns and exons of the human proα1(IV) collagen gene, and the patterns in comparable regions of the human proα2(I), proα1(II), and proα1(III) collagen genes. Black boxes represent -Gly-Xaa-Yaa- repeat-coding exons, hatched boxes represent NC-1 domain-coding exons, and the solid lines represent intron sequences. The numbering of exons of the proα1(IV) gene is based on the numbering of exons of interstitial collagen genes (10, 11, 13–22) and is shown beneath the exons. The small numbers above the solid lines indicate intron sizes in bp, and the larger numbers above exons represent sizes of exons. Exon 4 of the proα1(IV) gene is a junction exon containing 71 bp coding for -Gly-Xaa-Yaa- sequences and 142 bp coding for the NC-1 domain.

(24), repetitive coding sequences, such as the coding sequences for collagen, are unusually prone to unequal-crossover mutations during meiosis or mitosis. In the case of fibrillar collagens, an unequal-crossover event that produced a longer or shorter protein would be very likely to prevent assembly of the monomer into a longitudinal fibril. As illustrated by the mutations found in several patients with osteogenesis imperfecta and Ehlers–Danlos syndrome, such mutations can also prevent correct processing of the amino terminus of the protein by procollagen N-proteinase (42, 43) and destabilize the triple-helical structure itself (44). In the case of nonfibrillar collagens such as basement membrane collagen, mutations that change the length of the protein are much less likely to affect the triple-helical structure, since this is already interrupted by noncollagenous sequences (4). Mutations that change the length of the protein are also less likely to have any effect on the network-like structures formed by nonfibrous collagens.

We gratefully acknowledge the assistance of Dr. Linda Sandell in the screening of the genomic library. This work was supported in part by National Institutes of Health Grants AM16516 and AM22051.

1. Piez, K. A. (1984) in *Extracellular Matrix Biochemistry*, eds. Piez, K. A. & Reddi, A. H. (Elsevier, New York), pp. 1–39.
2. Miller, E. J. (1984) in *Extracellular Matrix Biochemistry*, eds. Piez, K. A. & Reddi, A. H. (Elsevier, New York), pp. 41–81.
3. Broek, D. L., Madri, J., Eikenberry, E. F. & Brodsky, B. (1985) *J. Biol. Chem.* **260**, 555–562.
4. Timpl, R., Wiedemann, H., Van Delden, V., Furthmayr, H. & Kühn, K. (1981) *Eur. J. Biochem.* **120**, 203–211.
5. Jander, R., Rauterberg, J., Voss, B. & von Bassewitz, D. B. (1981) *Eur. J. Biochem.* **114**, 17–25.
6. Bentz, H., Morris, N. P., Murray, L. W., Sakai, L. Y., Hollister, D. W. & Burgeson, R. E. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3168–3172.
7. Sage, H., Trüeb, B. & Bornstein, P. (1983) *J. Biol. Chem.* **258**, 13391–13401.
8. van der Rest, M., Mayne, R., Ninomiya, Y., Seidah, N. G., Chretien, M. & Olsen, B. R. (1985) *J. Biol. Chem.* **260**, 220–225.
9. Schmid, T. M. & Linsenmayer, T. F. (1985) *J. Cell Biol.* **100**, 598–605.
10. Yamada, Y., Avvedimento, V. E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. & de Crombrugghe, B. (1980) *Cell* **22**, 887–892.
11. Tate, V., Finer, M., Boedtker, H. & Doty, P. (1982) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1039–1049.
12. Tate, V. E., Finer, M. H., Boedtker, H. & Doty, P. (1983) *Nucleic Acids Res.* **11**, 91–104.
13. Boyd, C. D., Tolstoshev, P., Schafer, M. P., Trapnell, B. C., Coon, H. C., Kretschmer, P. J., Nienhuis, A. W. & Crystal, R. G. (1980) *J. Biol. Chem.* **255**, 3212–3220.
14. Chu, M.-L., de Wet, W. J., Bernard, M. P., Ding, J.-F., Morabito, M., Myers, J., Williams, C. & Ramirez, F. (1984) *Nature (London)* **310**, 337–340.
15. Myers, J. C., Dickson, L. A., de Wet, W. J., Bernard, M. P., Chu, M.-L., Di Liberto, M., Pepe, G., Sangiorgi, F. O. & Ramirez, F. (1983) *J. Biol. Chem.* **258**, 10128–10135.
16. Monson, J. M. & McCarthy, B. J. (1981) *DNA* **1**, 59–69.
17. Sandell, L. J., Yamada, Y., Dorfman, A. & Upholt, W. B. (1983) *J. Biol. Chem.* **258**, 11617–11621.
18. Cheah, K. S. E., Stoker, N. G., Griffin, J. R., Grosveld, F. G. & Solomon, E. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2555–2559.
19. Sangiorgi, F. O., Benson-Chanda, V., de Wet, W. J., Sobel, M. E. & Ramirez, F. (1985) *Nucleic Acid Res.* **13**, 2815–2826.
20. Yamada, Y., Liau, G., Mudryj, M., Obici, S. & de Crombrugghe, B. (1984) *Nature (London)* **310**, 333–337.
21. Chu, M.-L., Weil, D., de Wet, W. J., Bernard, M., Sippola, M. & Ramirez, F. (1985) *J. Biol. Chem.* **260**, 4357–4363.
22. Ramirez, F., Bernard, M., Chu, M.-L., Dickson, L., Sangiorgi, F., Weil, D., de Wet, W. J., Junien, C., Myers, J. & Sobel, M. (1985) *Ann. N. Y. Acad. Sci.*, in press.
23. Bernard, M. P., Chu, M.-L., Myers, J. C., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983) *Biochemistry* **22**, 5213–5223.
24. Prockop, D. J. & Kivirikko, K. I. (1984) *N. Engl. J. Med.* **311**, 376–386.
25. Monson, J. M., Natzle, J., Friedman, J. & McCarthy, B. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1761–1765.
26. Kramer, J. M., Cox, G. N. & Hirsh, D. (1982) *Cell* **30**, 599–606.
27. Lozano, G., Ninomiya, Y., Thompson, H. & Olsen, B. R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4050–4054.
28. Tryggvason, K., Gehron Robey, P. & Martin, G. R. (1980) *Biochemistry* **19**, 1284–1289.
29. Babel, W. & Glanville, R. W. (1984) *Eur. J. Biochem.* **143**, 545–556.
30. Kurkinen, M., Barlow, D. P., Helfman, D. M., Williams, J. G. & Hogan, B. L. M. (1983) *Nucleic Acid Res.* **11**, 6199–6209.
31. Wang, S.-Y. & Gudas, L. J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5880–5884.
32. Oberbaümer, I., Laurent, M., Schwarz, U., Sakurai, Y., Yamada, Y., Vogeli, G., Voss, T., Siebold, B., Glanville, R. W. & Kühn, K. (1985) *Eur. J. Biochem.* **147**, 217–224.
33. Pihlajaniemi, T., Tryggvason, K., Myers, J., Kurkinen, M., Lebo, R., Cheung, M.-C., Prockop, D. J. & Boyd, C. D. (1985) *J. Biol. Chem.* **260**, 7681–7687.
34. Brinker, J. M., Gudas, L. J., Loidl, H. R., Wang, S.-Y., Rosenbloom, J., Kefalides, N. A. & Myers, J. C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3649–3653.
35. Kurkinen, M., Bernard, M. P., Barlow, D. P. & Chow, L. T. (1985) *Nature (London)* **317**, 177–179.
36. Maniatis, T., Fritsch, E. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
37. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
38. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
39. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
40. Chow, L. T., Broker, T. R. & Lewis, J. B. (1979) *J. Mol. Biol.* **134**, 265–303.
41. Lopata, M. A., Havercroft, J. C., Chow, L. T. & Cleveland, D. W. (1983) *Cell* **32**, 713–724.
42. Steinmann, B., Tuderman, L., Peltonen, L., Martin, G. R., McKusick, V. A. & Prockop, D. J. (1980) *J. Biol. Chem.* **255**, 8887–8893.
43. Sippola, M., Kaffe, S. & Prockop, D. J. (1984) *J. Biol. Chem.* **259**, 14094–14100.
44. Williams, C. J. & Prockop, D. J. (1983) *J. Biol. Chem.* **258**, 5915–5921.