GENOMICS **41,** 236–242 (1997)
ARTICLE NO. GE974638

Complete Primary Structure of Two Splice Variants of Collagen XII, and Assignment of α1(XII) Collagen (COL12A1), α1(IX) Collagen (COL9A1), and α1(XIX) Collagen (COL19A1) to Human Chromosome 6q12–q13

DONALD R. GERECKE,*[,1] PAMELA F. OLSON,* MANUEL KOCH,* JOAN H. M. KNOLL,†
REGINALD TAYLOR,‡ DAVID L. HUDSON,* MARIE-FRANCE CHAMPLIAUD,*
BJORN R. OLSEN,§ AND ROBERT E. BURGESON*

**Cutaneous Biology Research Center, Massachusetts General Hospital and Department of Dermatology, Harvard Medical School, Building 149, 13th Street, Charlestown, Massachusetts 02129-2017; †Division of Genetics, Children's Hospital and Department of Pathology, Beth Israel Deaconess Medical Center/Harvard Medical School 300 Longwood Avenue, Boston, Massachusetts 02115; ‡The University of Alabama, School of Dentistry, 1919 Seventh Avenue South, Birmingham, Alabama 35294-0007; and §Department of Cellular Biology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115*

Received November 1, 1996; accepted January 29, 1997

Overlapping cDNA clones that encode the full-length human α1(XII) collagen polypeptides were isolated. The long variant molecule cDNA of 9750 nucleotides (nt) contains a 9189-nt open reading frame encoding 3063 amino acid residues. The short variant molecule cDNA of 6258 nt contains a 5697-nt open reading frame encoding 1899 amino acid residues. At the amino terminus of each variant is a 24-residue signal peptide that is followed by the mature polypeptides of 3039 amino acid residues with a calculated molecular mass of 330,759 Da for the long variant and 1875 amino acid residues with a calculated molecular mass of 203,163 Da for the short variant polypeptide. The human collagen XII chains are predicted to have all the structural domains described for the molecules in chicken and mouse, including, fibronectin type III repeats, von Willebrand factor A domains, and two triple-helical domains similar to those of all the other collagen family members. The amino acid residue sequence of human α1(XII) collagen showed 92% identity to the mouse chain and 78% identity to the chicken chain. The sequence of three peptide fragments of collagen XII isolated from human placenta was identical to the sequence predicted from the deduced cDNA sequence and confirms that the cDNA encodes human α1(XII) collagen. An isolated genomic clone was used to map the locus of the COL12A1 gene to chromosome 6q12–q13, very close to the locus of the FACIT collagen genes COL9A1 and COL19A1. RT-PCR on a variety of cDNAs demonstrates that both variant transcripts appear in human amnion, chorion, skeletal muscle, small intestine, and in cell cultures of human dermal fibroblasts, keratinocytes, and endothelial cells. Only the small variant transcript is apparent in human lung, placenta, kidney, and a squamous cell carcinoma cell line. These results confirm the previous observations showing that collagen XII is found in collagen I-containing tissues. © 1997 Academic Press

INTRODUCTION

Type XII collagen was independently discovered as a pepsin-resistant fragment (Yamauchi *et al.*, 1986) and as a distinct cDNA (Gordon *et al.*, 1987). The final protein is a homotrimer consisting of three α1 (XII) polypeptide chains (Dublet *et al.*, 1989). Analysis of the full-length sequence by overlapping cDNA clones has been completed in both chicken (Yamagata *et al.*, 1991)) and mouse (Bohme *et al.*, 1995), enabling the α1(XII) polypeptide chains to be subdivided into two collagen triple-helical domains (COL1 and COL 2) and three non-triple-helical domains (NC1, NC2, and NC3). Additionally, the globular amino-terminal domain (NC3) can be assigned several distinct subdomains homologous to domains found in other molecules. These include fibronectin type III repeats (FN), von Willebrand factor A domains (vW), and the amino-terminal globular domain found in α1(IX) collagen (called NC4 in collagen IX).

There are two different sizes of collagen XII polypeptides that are presumably formed by alternative splicing of the primary transcripts at the 5′ end (Treub and Treub, 1992). The short variant has 2 vW subdomains

Sequence data reported in this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. U73778 and U73779.

[1] To whom correspondence should be addressed at the Cutaneous Biology Research Center, Massachussetts General Hospital, Building 149 13th Street, Charlestown, MA 02129-2017. Telephone: (617) 726-4188. Fax: (617) 726-4453. E-mail: donald_gerecke@cbrc.mgh.harvard.edu.

0888-7543/97 $25.00

and 10 FN subdomains whereas the long variant has 2 additional vW subdomains and 8 additional FN subdomains at the amino-terminal end (Yamagata *et al.*, 1991; Trueb and Trueb, 1992). The final tissue form of collagen XII may contain homotrimers of either the longer or the shorter variant or any combination of long and short variant chains (Koch *et al.*, 1995). Only the long variant α1(XII) collagen is a proteoglycan, containing glycosaminoglycan side chains. Both variants bind heparin, but the long variant has additional potential heparin-binding sites (Koch *et al.*, 1995).

Collagen XII has been assigned to the FACIT (fibril-associated collagens with interrupted triple helices) family of collagens, which also includes collagens IX, XIV, XVI, and XIX (Olsen *et al.*, 1995). Immunohistochemical studies in chick (Sugrue *et al.*, 1989; Walchli *et al.*, 1994), bovine (Keene *et al.*, 1991), and mouse (Oh *et al.*, 1993) and gold-labelling antibody studies (Keene *et al.*, 1991; Koch *et al.*, 1995) support the hypothesis that collagen XII colocalizes with collagen I.

The association between collagen I fibrils and collagen XII is well established; however, the exact function of collagen XII is still unknown. Both are found in a variety of connective tissues, and one hypothesis is that collagen XII somehow modifies the interactions of collagen I fibrils with the surrounding matrix. This could in turn affect the biomechanical properties of the matrix and thus affect fibril and matrix density. This hypothesis is supported by *in vitro* experiments that have shown that collagen XII promotes fibroblast-mediated collagen I gel contraction and alters the compressability of collagen gels in the absence of cells (Nishiyama *et al.*, 1994).

In the present study we have characterized cDNA clones that encode the entire primary structure of both variants of human α1(XII) collagen polypeptides and compare them to the structure of the molecule in other species. We have used genomic clones for human α1(IX), α1(XII), and α1(XIX) FACIT collagens to map the chromosomal location of the genes to the same locus on chromosome 6. Finally, we have looked at the expression of the collagen XII variant transcripts by reverse transcriptase-polymerase chain reaction (RT-PCR) analysis in several tissues and cell types and have found that there are tissue-specific differences in expression of the two collagen XII variants.

MATERIALS AND METHODS

Isolation of total and poly(A)$^+$ RNA. RNA isolation for cells in culture was accomplished as follows: 1 to 10 Costar T-225 flasks were seeded with squamous carcinoma cells (SCC), primary foreskin fibroblasts, primary foreskin keratinocytes, or endothelial (HUVEC) cell line and allowed to grow until subconfluent. Medium was removed and the cells were trypsinized, transferred to a centrifuge tube, and washed 3× with PBS (centrifuging between washes); the final pellet was resuspended in 1 ml Trizol reagent (Gibco BRL). The whole tissues were processed as follows: to 1 g of fresh human placenta, amnion, or chorion 10 ml Trizol reagent was added and homogenized with a Polytron homogenizer. After the cell culture or tissue Trizol samples were incubated for 5 min at room temperature (RT), they were centrifuged for 1 min at 10,000*g*, and the supernatants were transferred to new tubes. To each Trizol reagent sample (either tissue or cells) was added 0.2 ml chloroform per milliliter of supernatant. Samples were vigorously hand shaken, incubated for 3 min at RT, and centrifuged at 12,000*g* for 15 min at 4°C, and the clear aqueous phase was removed. To the aqueous was added 0.5 ml isopropanol per milliliter of solution, and samples were incubated for 10 min at RT and centrifuged at 12,000*g* for 10 min at 4°C; the supernatant was discarded. Pellets were washed with 75% ethanol, vortexed, and centrifuged at 7500*g* for 5 min at RT, and the supernatant was discarded. Pellets were air dried for 10 min and resuspended in 0.5 ml sterile RNase-free water, and the total RNA was quantitated by spectrophotometry. The lung RNA, small intestine RNA, and skeletal muscle RNA were purchased from Clontech, Inc.

cDNA cloning. A Clontech first-strand cDNA synthesis kit was used to synthesize cellular or tissue cDNA from total RNA using either oligo(dT), random, or specific human type XII collagen antisense primers following the manufacturer's protocol. RT-PCR-based technology was used to generate overlapping clones complementary to most of the α1(XII) collagen mRNA. Specific primers were deduced from the previously published partial human genomic clone (Oh *et al.*, 1992), chicken collagen XII sequence (Yamagata *et al.*, 1991), or mouse collagen XII sequence (Bohme *et al.*, 1995), or from specific human sequences determined from novel clones developed during the course of these studies. The Gibco BRL anchor primer (5′AP) from the 5′ RACE kit and adapter primer (3′AP) from the 3′ RACE kit were also used. All primers were used at a concentration of 1 μM. Resulting PCR products were directly cloned into the PCR II or PCR 2.1 vectors using reagents and conditions suggested by the manufacturer (Invitrogen). The following 16 plasmids (Fig. 1) were obtained from amplifications using the indicated primer pairs and any of the various cDNAs as templates as described above (the first primer listed is a direct sequence and the second primer listed is the reverse complement in every case): cDNA 72 (5′AP and human nt 382–406); cDNA 71 (5′AP and human nt 861–885); cDNA 60 (mouse nt 131–154 and mouse nt 1273–1296); cDNA 59 (mouse nt 1032–1055 and human nt 2211–2235); cDNA 58 (mouse nt 1378–1401 and human nt 2211–2235); cDNA 56 (mouse nt 1945–1968 and mouse nt 2749–2772); cDNA 70 (5′AP and human nt 3606–3630); cDNA 67 (chick nt 3647–3673 and chick nt 4187–4213); cDNA 42 (human nt 3963–3986 and human nt 7518–7542); cDNA 31 (human nt 4880–4904 and human nt 6969–6993); cDNA 68 (5′AP and human nt 8058–8082); cDNA 66 (human nt 7811–7836 and human nt 8352–8379); cDNA 64 (human nt 8252–8273 and 3′AP); cDNA 62 (human nt 8322–8346 and 3′AP); cDNA 73 (primers human nt 78–99 and human nt 4023–4045); and cDNA 74 (primers human nt 114–135 and human nt 4023–4045).

To obtain the initiation codon and the 5′ untranslated region of the long variant, a 5′ RACE kit was employed following company protocols (Gibco BRL). To obtain the complete 3′ untranslated region to the poly(A)$^+$ tail, a 3′ RACE kit was employed following the manufacturer's protocols (Gibco BRL). In all cases using the RACE procedure, three rounds of internal, human-specific nested primers were used with the adapter primers provided by the company to generate increasingly smaller, specific PCR products. For the 5′ end of the long variant, the three antisense nested internal primers were human nt 1219–1239; human nt 1173–1193; and human nt 382–406. For the 3′ end, the three sense nested internal primers were human nt 7111–7135, human nt 7811–7836, and human nt 8322–8346. The 5′ end of the short variant was determined using the primer pairs described for cDNAs 73 and 74 above, using lung cDNA as a template. The final PCR products were directly ligated into the TA cloning vectors as described previously.

Nucleotide sequencing. The nucleotide sequences were determined by the dideoxy chain termination technique of Sanger *et al.* (1977) using a Sequenase 2.0 kit (U.S. Biochemicals). Sequence reactions employed the M13 forward or reverse primers and specific primers synthesized in our laboratory (generally 18–21mers). Nucleotide sequence was determined from both the sense and antisense strands, and DNA compressions were resolved using inosine in place of guano-

sine. The sequence data were assembled and manipulated using the fragment assembly system (Genetics Computer Group).

Protein sequencing. Human collagen XII was purified from placenta as previously described by our laboratory and analyzed by Western immunoblotting using our polyclonal antibody (Lunstrum *et al.*, 1991) against α1 (XII) collagen. Tryptic peptides were generated from the immunoreactive 220-kDa band and sequenced as described previously (Gerecke *et al.*, 1994).

RT-PCR analysis. Random-primed cDNAs made from RNA isolated as above was used in all cases to compare the distribution of the two variants of collagen XII in various cell and tissue samples. Human nt 3819–3843 was used as a common antisense primer to detect both variants of collagen XII. To amplify the specific short variant, human nt 114–135 was used as a sense primer. The product size would be 248 bp. To amplify the specific long variant, human nt 3255–3279 was used as a sense primer. The product size would be 586 bp. The PCRs were carried out with equal amounts of cDNA and both sense primers and the common antisense primers in the same tube. A Perkin–Elmer Geneamp 9600 PCR system was used to amplify the samples with *Taq* polymerase enzyme (Perkin–Elmer) under the following conditions: 30 cycles at 94°C for 60 s, 55°C for 30 s, 72°C for 60 s, and 7 min extension at 72°C. The resulting products were analyzed on a 1% agarose gel containing ethidium bromide.

Chromosomal location. Fluorescence *in situ* hybridization (FISH) was performed using purified DNAs from phage clones 8B211 [human α1(IX) collagen, Warman *et al.*, 1993], RT5-1 [human α1(XII) collagen; Oh *et al.* (1992)], and αYt9E [human α1(XIX) collagen; Inoguchi *et al.* (1995)]. The clones were labeled by nick-translation with either biotin-16–dUTP or digoxigenin-11–dUTP. The probes (100 ng of each) were simultaneously hybridized to interphase preparations from peripheral blood lymphocytes of a normal individual. Three-probe two-color hybridizations and detection were performed as previously described (Knoll and Lichter, 1994). Biotin-labeled probes were detected with fluorescein (FITC) conjugated avidin (Vector Laboratories, 7 μg/ml), and digoxigenin-labeled probes were detected with rhodamine conjugated digoxigenin antibody (Boehringer Mannheim, 3 μg/ml). Nuclear DNA was counterstained with 4′,6-diamidino-2-phenylindeole (DAPI, 0.1 μg/ml). Cells were viewed with a Zeiss Axiophot epifluorescence microscope equipped with a FITC/rhodamine dual bandpass filter set (Omega Optical, Inc.) and DAPI filter set (Zeiss). The sequence order was determined by scoring the color order of the three probes on interphase cells. In one experiment, 8B211 and αYt9E were detected with rhodamine and RT5-1 was detected with FITC while in the other simultaneous hybridization experiment 8B211 and RT5-1 were detected with FITC and αYt9E was detected with rhodamine. At least 60 hybridized interphase chromosomes were scored per experiment.

RESULTS

Molecular Cloning and Domain Structures of α1(XII) Collagen Polypeptides

A single initial human genomic clone has previously been described as containing three exons for a portion of COL2 and NC3 (Oh *et al.*, 1992). By a combination of RT-PCR using primers based on the published chicken and mouse nucleotide sequences and on 3′ RACE and 5′ RACE protocols, we extended the sequence in both directions to over 9 kb from the translational start codon to the poly(A)$^+$ tail. Figure 1 shows a diagram of the various domains for human α1(XII) collagen as well as the overlapping cDNAs covering the entire translated region for both variants of collagen XII polypeptides.

The overlapping cDNAs for the long variant span 9750 nucleotides including a 113-nt 5′ untranslated region, an open reading frame of 9189 nucleotides encoding 3063 amino acid residues, and a 3′ untranslated region of 418 nt (data reviewed, but not shown). The open reading frame begins at the translation initiation signal ATG at nucleotide 114 as determined according to the method of Kozak (1989). The first 24 amino acid residues encode a typical signal peptide, characteristic of secreted proteins, and is the same size as that in mouse (1 amino acid longer than in the chicken). The signal peptide cleavage site is predicted to occur after the sequence glu-ala-glu, in agreement with the rules of von Heijne (1986). The predicted, processed long variant of human type XII collagen begins with valine as do the mouse and chicken chains. The mature, longer form polypeptide has a calculated molecular mass of 330,759 Da. The long variant of α1(XII) collagen may be divided into 4 von Willebrand factor A-like domains (vWA, vWB, vWC, and vWD), 18 fibronectin type III repeats (FN 1-18), 1 NC4 α1(IX) collagen homology domain (IXP), 2 triple-helical domains (COL 1-2) and 2 non-triple-helical domains (NC 1-2). The short variant is 1164 amino acid residues shorter than the long variant. It is missing domains vWA, vWB, and FN 1-8. Both forms share the same 5′ untranslated region and a common signal peptide, which is apparently spliced to amino acid residue 1189 in the case of the short variant. The amino acid sequence of the long form contains 6 consensus sequence sites for the putative attachment of N-linked oligosaccharides (asn-X-ser/thr, four of which are retained in the short variant). Three potential glycosaminoglycan glycosylation sites (ser-gly-X-gly) are found only in the long variant. The sequence of amino acid residues 1280 to 1295, 1782 to 1801, and 2906 to 2916 matches exactly the sequence of three peptides isolated from human collagen XII, confirming that the cDNA encodes authentic human α1(XII). The translation stop codon, TAA, starting at nucleotide 9303, is followed by the 3′ untranslated region of 418 nucleotides containing a single polyadenylation site (AATAAA), 16 nucleotides upstream from the poly(A)$^+$ tail.

Chromosomal Location

For interphase ordering, two-color three-probe hybridizations were performed. The three probes were hybridized simultaneously and detected by red and green fluorescence. Representative hybridizations are shown in Fig. 2. From the combined results of the two hybridization experiments, the following probe order was deduced: αYt9E [α1(XIX) collagen]–8B211 [α1(IX) collagen]–RT5-1 [α1(XII) collagen]. The α1(XIX) and α1(IX) collagen genes appeared to be closer to each other on interphase chromatin than the α1(XII) collagen gene. This finding was confirmed on metaphase chromosomes from hybridization 1, in which chromosomes from 13 cells revealed that the green signal (RT5-1) was distinct from the red signal and more dis-

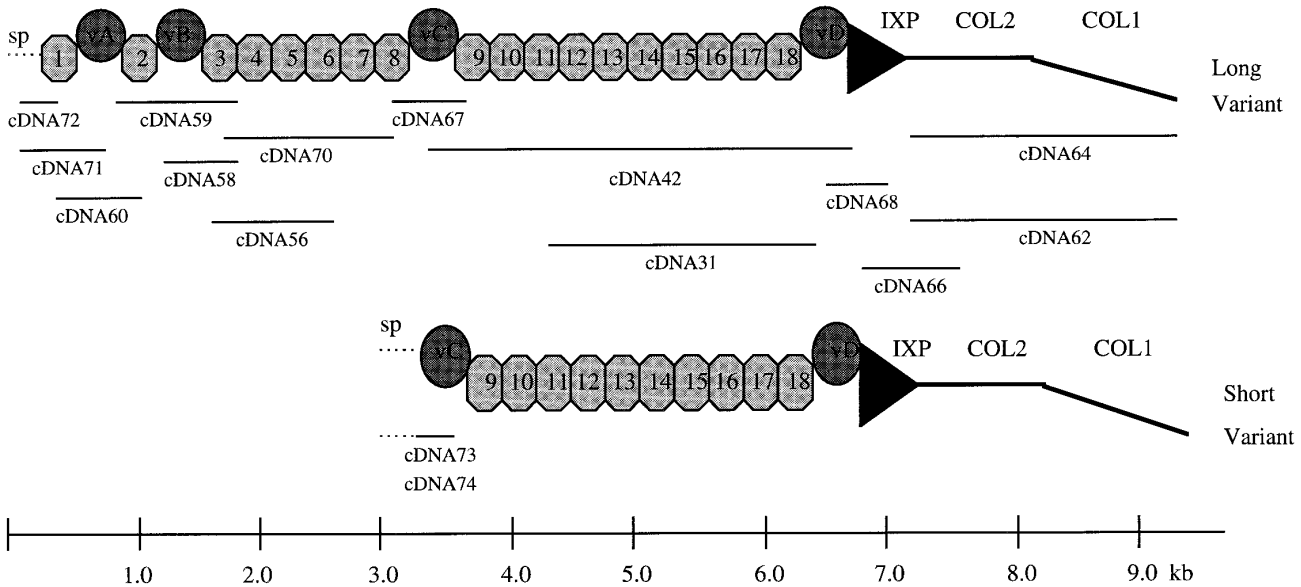

FIG. 1. Schematic diagram of the position of cDNA clones relative the domains of the α1(XII) collagen protein. Signal peptide is labeled sp; fibronectin type III repeats are numbered 1–18; von Willebrand factor A-like domains are labeled vA, vB, vC, and vD; collagen α1(IX)-like domain is labeled IXP; triple-helical collagen domains are labeled COL1 and COL2. Bar represent 1.0 kb of nucleotide per vertical mark.

tal to the centromere than the red signal (8B211 and αYt9E). A single red hybridization was observed on the metaphase chromosomes in this experiment, suggesting that 8B211 and αYt9E are less than 1000 kb apart. This suggested difference in distance between the probes could result in chromatin looping in interphase cells and would account for the 12 chromosomes in hybridization 2 that suggested that αYt9E was between RT5-1 and 8B211.

Expression of Long and Short Form Collagen XII Variants in Tissues and Cell Culture

To examine the potential differential expression of the two possible transcripts of collagen XII, RT-PCR was done as previously described on a variety of cDNA samples. The results of the experiments are presented in Fig. 3. Both variants appear in fibroblasts, keratinocytes, endothelial cells, amnion, chorion, skeletal muscle, and small intestine (lanes 2, 3, 4, 7, 8, 11, and 12). It is interesting to note that the relative amount of the long variant to the short variant appears to be higher in amnion (lane 7). Only the short variant is present in squamous cell carcinoma cells, placenta, lung, and kidney (lanes 5, 6, 9, and 10). Human liver cells were negative for both variants, but positive for the control housekeeping gene, glyceraldehyde-3-phosphate dehydrogenase (data not shown).

DISCUSSION

The nucleotide sequence described in this paper is the first complete human sequence for collagen XII. Table 1 compares the sequence to the other published collagen XII sequences and shows that the overall nucleotide identity is 88% (mouse), 76% (chicken), and 68% (newt, partial). The amino acid residue identity is 92 and 78%, respectively, for mouse and chicken. On the basis of this high identity and from the three peptide sequences obtained from collagen XII protein, we conclude that the sequence deposited with the EMBL/GenBank sequence databank is the human α1(XII) collagen polypeptide chain. There is very high amino acid identity (greater than 95%) in many of the subdomains across species. It is therefore likely that the molecule has functions that are conserved between the species.

There are two putative alternative splice variants at the 5′ end of human collagen XII. These variants have not been shown to be alternatively spliced products from a single gene, and it has never been shown whether they share a common promoter, however, on the cDNA level they share the same 5′ untranslated sequences, as well as identical signal peptides in mouse, chicken, and human. On the protein level these variants differ by a molecular weight of about 100 kDa in all three species. The long variant of collagen XII is a proteoglycan (data not shown), while the short variant is not, which is consistent with the other two vertebrate species. There are three potential glycosaminoglycan glycosylation sites in the large variant in subdomains FN4 (ser-gly-pro-gly), FN5 (ser-gly-ala-gly), and FN6 (ser-gly-glu-gly). These three sites have absolute sequence identity in all four species examined to date (human, mouse, chicken, and newt). A fourth site that was reported for mouse (ser-gly-glu-gly in FN5) is not conserved in any of the other species and in human is instead phe-gly-glu-gly. One may conclude that glycosylation will occur on at least one of these sites as has been reported for bovine collagen XII protein (Watt *et*

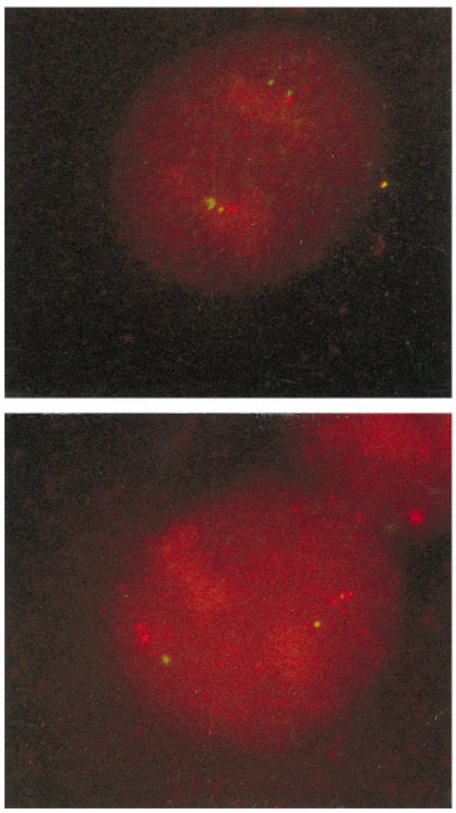

FIG. 2. Interphase ordering of Col19A1, Col9A1, and Col12A1 by fluorescence *in situ* hybridization. (**Left**) 8B211 (Col9A1) and αYt9E (Col19A1) were labeled with digoxigenin–dUTP and detected with rhodamine conjugated antibody (red) while RT5-1 (Col12A1) was labeled with biotin–dUTP and detected with avidin-FITC (green). The resulting order revealed Col9A1/Col19A1, Col12A1 (red/red/green). The resulting orientation of Col9A1 and Col19A1 could not be distinguished. Red signals appear closer to each other than to the green signal. (**Right**) In a subsequent hybridization Col9A1 was labelled with biotin–dUTP while labeling of the other two probes remained the same as above. The resulting order revealed Col9A1/Col12A1, Col19A1 (green/green/red). The red signal appears close to one green signal. Combining the results from both hybridization experiments, the order is Col19A1, Col9A1, Col12A1.

al., 1992). The long variant has two potential N-linked oligosaccharide sites that are missing in the short variant, but it is currently unclear whether any of these sites are utilized for carbohydrate attachment.

It has been reported that the prevalent form of collagen XII in chicken skin is the short variant (Koch *et al.*, 1995). Both variants are also present in mouse (Bohme *et al.*, 1995), bovine (Lunstrum *et al.*, 1991; Watt *et al.*, 1992), and human (Keene *et al.*, 1991) skin. Collagen XII has the potential to be expressed by dermal fibroblasts or keratinocytes in the skin. We examined these cell types by RT-PCR and found they each contain transcripts for both variants of collagen XII. Placenta had only the short variant transcript and while amnion appeared to have both variants, the relative amount of the long variant appeared higher. It would be interesting to quantitate the actual mRNA levels for both variants in the amnion and compare it to immunofluorescence and immunoprecipitation patterns of the tissue. Placenta will be a good source for the isolation of the individual short protein variant that will be required for future studies. Past antibody studies have indicated that collagen XII is generally not found in organs, but our results indicate that the short variant transcript is present in lung, kidney, muscle, and small intestine, but not in liver. One possibility is that the product seen actually comes from contaminating blood vessels, which are known to contain collagen XII. However, the liver is rich in blood vessels and did not demonstrate either transcript, so our conclusion is that collagen XII message is a true component of these tissue cDNAs. The more likely possibility is that the RNA is present in these organs, but not translated. The regulation of collagen XII needs to be investigated in future studies.

Collagen IX is also a member of the FACIT group of collagens (Shaw and Olsen, 1991). It is a heterotrimer, containing three polypeptide chains that are separate gene products, the largest of which, α1(IX), has domain similarities to collagen XII. The largest domain, NC4, on the amino-terminus of collagen IX, is represented almost in its entirety as a subdomain of collagen XII (IXP, Fig. 1) and the two polypeptides might have overlapping functions. The human chromosomal location for the α1(IX), α1(XII), and α1(XIX) FACIT collagens is 6q12–q13, leading us to speculate that the three polypeptides arose by ancestral gene duplication. Which of the genes originated first cannot be determined until the exon/intron structures have been determined and compared for all three polypeptides.

The function of collagen XII is currently unknown. Our *in vitro* data suggest that it may play a role in modulating the density of type I collagen-containing tissues (Nishiyama *et al.*, 1994). Gordon *et al.* (1996) have noted a correlation between the appearance of collagen XII morphogenetic changes in the chick cornea and suggest that collagen XII may affect matrix remodeling. Chiquet *et al.* (1996) also suggested a remodeling

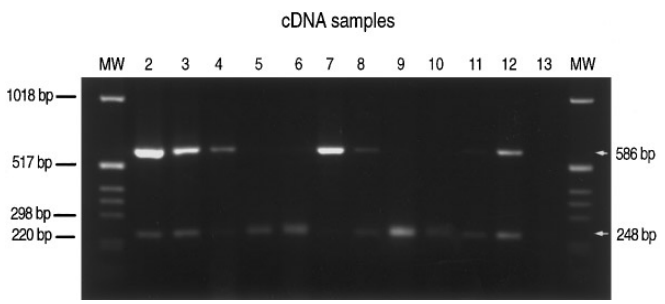

FIG. 3. One percent agarose gel stained with ethidium bromide showing the PCR products for the long (586 bp) and short (248 bp) variants of collagen XII in a variety of human cell cultures and tissues. Lanes **1** and **14**, molecular weight standards (MW) in nucleotide basepairs (bp). Lane **2**, foreskin fibroblast culture. Lane **3**, foreskin keratinocyte culture. Lane **4**, endothelial cell culture. Lane **5**, squamous cell carcinoma culture. Lane **6**, placenta; Lane **7**, amnion; Lane **8**, chorion; Lane **9**, lung; Lane **10**, kidney; Lane **11**, skeletal muscle; Lane **12**, small intestine; Lane **13**, liver.

TABLE 1

Comparison of the Conservation of Structural Domains between Human and Other α1(XII) Collagen Amino Acid Sequences

Domains	Human domain boundaries	Mouse		Chicken		Newt	
		% sim	% id	% sim	% id	% sim	% id
Signal	1–24	83.3	83.3	75.0	70.8		
fn1	25–115	96.7	92.3	89.0	76.9		
vwa	130–320	97.9	95.3	94.2	89.0		
fn2	332–420	91.0	77.5	83.0	65.9		
vwb	427–620	95.9	95.4	94.8	87.6	92.3	90.4
fn3	630–721	90.2	79.3	69.6	59.8	63.0	52.2
fn4	722–812	91.2	83.5	82.4	67.0	83.5	72.5
fn5	813–903	90.1	85.7	79.1	70.3	72.5	60.4
fn6	904–996	95.7	89.4	78.5	58.1	80.6	65.6
fn7	999–1084	95.3	88.4	65.1	47.7	84.7	69.4
fn8	1087–1178	98.9	96.7	93.5	91.3	84.8	73.9
vWC	1189–1410	98.2	96.4	86.0	77.0	83.8	71.2
FN9	1384–1470	98.9	95.4	81.6	72.4	80.5	67.8
FN10	1476–1564	87.6	77.5	73.0	56.2	73.7	52.6
FN11	1565–1648	86.9	83.3	79.8	69.0		
FN12	1655–1751	90.7	84.5	88.7	74.2		
FN13	1752–1842	96.7	92.3	91.2	81.3		
FN14	1843–1932	94.4	93.3	86.7	76.7		
FN15	1933–2023	91.1	86.7	85.6	75.6		
FN16	2024–2114	100.0	97.8	96.7	93.4		
FN17	2115–2202	97.7	96.6	95.4	92.0		
FN18	2203–2290	97.7	92.0	90.9	86.4		
vWD	2293–2504	98.6	96.7	94.8	88.7		
IXP	2505–2720	98.1	95.8	91.2	82.9		
COL2	2747–2898	98.0	96.6	90.1	82.9		
NC2	2899–2941	97.7	95.3	90.7	98.7		
COL1	2942–3044	97.1	97.1	94.2	92.2		
NC1	3045–3063	100.0	94.1	94.4	77.8		
Entire protein	aa 1–3063	95.5	91.6	87.3	78.2	79.9	68.1
Entire cDNA	nt 1–9750		87.5		76.0		68.1

Note. The boundaries of the structural domains, in amino acid residue number, are given for the human sequence. They were determined by homology with the published mouse α1(XII) collagen domain boundaries (Bohme *et al.*, 1995). The accession numbers of the sequences compared are: human (U73778, this paper), mouse (U25652, Bohme *et al.*, 1995), chicken (D00824, Yamagata *et al.*, 1991), and newt (U19494, Wei *et al.*, 1995). The domains that are present only in the long form variant are denoted by lowercase letters. The long/short form variant splice junctions are denoted by single underline. sim, amino acid residue similarity; id, amino acid residue exact identity; fn, fibronectin type III-like domains; vW, von Willebrand factor A-like domains; IXP, type IX collagen-like domain; COL, triple-helical domains; NC, nontriple-helical domains; aa, amino acid residues; nt, nucleotides.

role for collagen XII through a mechanical stimuli mechanism. If it is indeed involved in matrix remodeling/tissue density, then the differential expression of the two size variants of collagen XII, the differences in their glycosylation, and the potential for heparin binding all could affect the tissue by influencing collagen I fibril spacing or diameters. More studies are required to elucidate further the interactions of collagens I and XII.

ACKNOWLEDGMENTS

These studies were supported by a research fellowship from the Dermatology Foundation (D.R.G.); National Institute of Health [USPHS AR35689 (R.E.B.) and HD18568 (J.H.M.K.)]; Beth Israel Pathology Foundation, Inc. (J.H.M.K.); and Cutaneous Biology Research Center through the Massachusetts General Hospital/Shiseido Co., Ltd. Agreement (R.E.B.). The technical assistance of Bonnie Blenner, Kerry Rowland, and Joshua Mularella is greatly appreciated. We thank Dr. Ulla Wewer for the liver and kidney RNA.

REFERENCES

Bohme, K., Li, Y., Oh, P. S., and Olsen, B. (1995). Primary structure of the long and short splice variants of mouse collagen XII and their tissue-specific expression during embryonic development. *Dev. Dyn.* **204:** 432–445.

Chiquet, M., Matthisson, M., Koch, M., Tannheimer, M., and Chiquet-Ehrismann, R. (1996). Regulation of extracellular matrix synthesis by mechanical stress. *Biochem. Cell Biol.* **74:** 101–108.

Dublet, B., Oh, S., Sugrue, S. P., Gordon, M. K., Gerecke, D. R., Olsen, B. R., and van der Rest, M. (1989). The structure of avian type XII collagen. Alpha 1 (XII) chains contain 190-kDa non-triple helical amino-terminal domains and form homotrimeric molecules. *J. Biol. Chem.* **264:** 13150–13156. [Published erratum appears in *J. Biol. Chem.* (1990) Jan. 15; **265**(2):1231]

Gerecke, D. R., Wagman, W. D., Champliaud, M. F., and Burgeson,

R. E. (1994). The complete primary structure for a novel laminin chain, the laminin B1k chain. *J. Biol. Chem.* **102:** 11073–11080.

Gordon, M. K., Gerecke, D. R., and Olsen, B. R. (1987). Type XII collagen: Distinct extracellular matrix component discovered by cDNA cloning. *Proc. Natl. Acad. Sci. USA* **84:** 6040–6044.

Gordon, M. K., Foley, J. W., Linsenmayer, T. F., and Fitch, J. M. (1996). Temporal expression of types XII and XIV collagen mRNA and protein during avian corneal development. *Dev. Dyn.* **206:** 49–58.

Inoguchi, K., Yoshioka, H., Khaleduzzaman, M., and Ninomiya, Y. (1995). The mRNA for α1(XIX) collagen chain, a new member of FACITs, contains a long unusual 3′ untranslated region and displays many unique splicing variants. *J. Biochem.* **117:** 137–146.

Keene, D. R., Lunstrum, G. P., Morris, N. P., Stoddard, D. W., and Burgeson, R. E. (1991). Two type XII-like collagens localize to the surface of banded collagen fibrils. *J. Cell Biol.* **113:** 971–978.

Knoll, J. H. M., and Lichter, P. (1994). In situ hybridization to metaphase chromosomes and interphase nuclei. *In* "Current protocols in Human Genetics" (N. C. Dracopoli, J. L. Haines, B. R. Korf, D. T. Moir, C. C. Morton, C. E. Seidman, J. G. Seidman, and D. R. Smith, Eds.), Vol. 1, Green–Wiley, New York.

Koch, M., Bohrmann, B., Matthison, M., Hagios, C., Trueb, B., and Chiquet, M. (1995). Large and small splice variants of collagen XII: Differential expression and ligand binding. *J. Cell Biol.* **130:** 1005–1014.

Kozak, M. (1989). The scanning model for translation: An update. *J. Cell Biol.* **108:** 229–241.

Lunstrum, G. P., Morris, N. P., McDonough, A. M., Keene, D. R., and Burgeson, R. E. (1991). Identification and partial characterization of two type XII-like collagen molecules. *J. Cell Biol.* **113:** 963–969.

Nishiyama, T., McDonough, A. M., Bruns, R. R., and Burgeson, R. E. (1994). Type XII and XIV collagens mediate interactions between banded collagen fibers in vitro and may modulate extracellular matrix deformability. *J. Biol. Chem.* **269:** 28193–28199.

Oh, S. P., Taylor, R. W., Gerecke, D. R., Rochelle, J. M., Seldin, M. F., and Olsen, B. R. (1992). The mouse alpha 1(XII) and human alpha 1(XII)-like collagen genes are localized on mouse chromosome 9 and human chromosome 6. *Genomics* **14:** 225–231.

Oh, S. P., Griffith, C. M., Hay, E. D., and Olsen, B. R. (1993). Tissue-specific expression of type XII collagen during mouse embryonic development. *Dev. Dyn.* **196:** 37–46.

Olsen, B. R., Winterhalter, K. H., and Gordon, M. K. (1995). FACIT collagens and their biological roles. *Trends Glycosci. Glycotech.* **7,** 115–127.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:** 5463–5467.

Shaw, L. M., and Olsen, B. R. (1991). FACIT collagens: Diverse molecular bridges in extracellular matrices. *Trends Biochem. Sci.* **16:** 191–194. [Review]

Sugrue, S. P., Gordon, M. K., Seyer, J., Dublet, B., van der Rest, M., and Olsen, B. R. (1989). Immunoidentification of type XII collagen in embryonic tissues. *J. Cell Biol.* **109:** 939–945.

Treub, J., and Treub, B. (1992). The two splice variants of collagen XII share a common 5′ end. *Biochim. Biophys. Acta* **1171,** 97–98.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 4683–4689.

Walchli, C., Koch, M., Chiquet, M., Odermatt, B. F., and Treub, B. (1994). Tissue-specific expression of the fibril-associated collagens XII and XIV. *J. Cell Sci.* **107:** 669–681.

Warman, M. L., Tiller, G. E., Polumbo, P. A., Seldin, M. F., Rochelle, J. M., Knoll, J. H. M., Cheng, S.-D., and Olsen, B. R. (1993). Physical and linkage mapping of the human and murine genes for the a1 chain of type IX collagen (COL9A1). *Genomics* **17:** 694–698.

Watt, S. L., Lunstrum, G. P., McDonough, A. M., Keene, D. R., Burgeson, R. E., and Morris, N. P. (1992). Characterization of collagen types XII and XIV from fetal bovine cartilage. *J. Biol. Chem.* **267:** 20093–20099.

Wei, Y., Yang, E.V., Klatt, K.P., and Tassava, R.A. (1995). Monoclonal antibody MT2 identifies the urodele α1 chain of type XII collagen, a developmentally regulated extracellular matrix protein in regenerating newt limbs. *Dev. Biol.* **168:** 503–513.

Yamagata, M., Yamada, K. M., Yamada, S. S., Shinomura, T., Tanaka, H., Nishida, Y., Obara, M., and Kimata, K. (1991). The complete primary structure of type XII collagen shows a chimeric molecule with reiterated fibronectin type III motifs, von Willebrand factor A motifs, a domain homologous to a noncollagenous region of type IX collagen, and short collagenous domains with an arg-gly-asp site. *J. Cell Biol.* **115:** 209–221.

Yamauchi, M., Kuboki, Y., Sasaki, S., and Mechanic, G. L. (1986). New pepsin-solubilized low molecular weight collagenous component possibly unique to periodontal ligament. *Biochem.* **25:** 1997–2002.