The human gene for the α1 chain of type XV collagen (COL15A1) is about 145 kilobases in size and contains 42 exons. The promoter is characterized by the lack of a TATAA motif and the presence of several Sp1 binding sites, some of which appeared to be functional in transfected HeLa cells. Comparison with Col18a1, which encodes the α1(XVIII) collagen chain homologous with α1(XV), indicates marked structural homology spread throughout the two genes. The mouse Col18a1 contains one exon more than COL15A1, due to the fact that COL15A1 lacks sequences corresponding to exon 3 of Col18a1, which encodes a cysteine-rich sequence motif. Twenty-five of the exons of the two genes are almost identical in size, six of them contain conserved split codons, and the locations of the respective exon-intron junctions are identical or almost identical in the two genes. The homologous exons include the closely adjacent first pair of exons and the exons encoding a thrombospondin-1 homology found in the N-terminal noncollagenous domain 1, which are followed by the most variable part of the two genes, covering the C-terminal half of their noncollagenous domain 1 and the beginning of the collagenous portion, after which most of the exons are homologous. The lengths of the introns are not similar in these genes, with two exceptions, namely the first intron, which is very short, less than 100 base pairs, and the second intron, which is very large, about 50 kilobases, in both genes. It can be concluded that COL15A1 and Col18a1 are derived from a common ancestor.

The family of collagens is large, and the number of known collagenous proteins is increasing. Nineteen genetically distinct vertebrate collagen types and more than 30 genes that encode their constitutive α chains have been identified to date (1–4). The criteria for classification as collagen are that such proteins have at least one triple-helical domain consisting of polypeptide chains with a repeated Gly-X-Y sequence and are structural components of the extracellular matrix. The collagen types have been named with Roman numerals in the order of their discovery. The fibril-forming collagens, types I, II, III, V, and XI, have a single, uninterrupted triple-helical domain that is available for fibril formation. The genes encoding these types are highly homologous (1, 5–7), and those for the three major ones, COL1A1, COL1A2, COL2A1, and COL3A1, are characterized by 51–52 exons. Their triple-helical domain is coded for by 41–42 exons, most of which are 54 bp1 in size or multiples thereof, and each exon begins with a complete codon for a glycine. The class of nonfibril-forming collagens includes types IV, VI–X, and XII–XIX, which all have one or more interruptions in the collagenous sequence. The genes coding for this heterogeneous group are more divergent in structure, and their numbers and sizes of exons can vary considerably (1, 5, 6).

The complete primary structure of the human α1(XV) chain consists of 1388 residues, with the following domains: a 25-residue putative signal peptide, a 530-residue N-terminal noncollagenous domain, a 577-residue collagenous sequence, and a 256-residue C-terminal noncollagenous domain (8). The collagenous sequence consists of nine collagenous domains, which are separated by eight noncollagenous domains. Collagen types XV and XVIII have been found to be homologous (8–13), and it has been suggested that they should be called multplexins (multiple triple helix domains and interruptions) (10). The N-terminal noncollagenous domains of both collagen chains contain sequence homology to thrombospondin, and seven of their collagenous domains are homologous, as are the C-terminal noncollagenous domains.

The exon-intron organization of the mouse type XVIII collagen gene has recently been determined (14), and a partial structure corresponding to the seven extreme 3’ exons has been described for the gene encoding human type XV collagen (8). The genes encoding the homologous collagens are located on separate chromosomes, the human gene for the α1(XV) collagen chain having been mapped to chromosome 9 (15) and its mouse counterpart to chromosome 4 (16), whereas the α1(XVIII) collagen gene is located on human chromosome 21 and mouse chromosome 10 (11).

We report here on the isolation of genomic clones for the human type XV collagen and characterization of the exon-intron organization of the entire gene. Comparison of the type XV collagen gene with that encoding type XVIII collagen reveals marked conservation in exon-intron organization, thus indicating that the two genes derive from a common ancestor. Analyses of the 5’-flanking sequence of the COL15A1 gene using a computer search for promoter elements and deletion constructs transfected into HeLa cells suggested a “housekeeping promoter” characterized by the lack of a TATAA motif and the presence of apparently functional Sp1 binding sites.

1 The abbreviations used are: bp, base pair(s); kb, kilobase(s); PCR, polymerase chain reaction; Pipes, piperazine-N,N’-bis[2-ethanesulfonic acid]; del, deletion.

‡ To whom correspondence should be addressed. Tel.: 358-8-5375800; Fax: 358-8-5375810; E-mail: taina.pihlajaniemi@oulu.fi.

Received for publication, January 21, 1998, and in revised form, April 13, 1998

© 1998 by The American Society for Biochemistry and Molecular Biology, Inc.

This paper is available online at http://www.jbc.org

Printed in U.S.A.
noncharacterized sequences. The scale is marked with a bar. 50-mer DNA-

was used as a template in a 10-

PCR containing polymerase buffer (Promega; 50 mM KCl, 1 mM MgCl2, 0.1% Triton X-100, 10 mM Tris, pH 9.0), 0.2 mM of each dNTP, 2.5 U of Taq DNA polymerase. Two-

mer primers synthesized in an Applied Biosystems DNA synthesizer (Promega Pharmacia Biotech). Next, 2.5 l of the diluted ligation mixture was used as the template in a 10-

PCR. The PCR mixtures contained

according to the manufacturer's protocol. Briefly, the various luciferase deletion constructs (5 µg) were transfected with 1 µg of pCMV-β-galactosidase plasmid (CLONTECH) to normalize for transfection efficiencies. For cotransfection experiments, 5 µg of luciferase plasmids were cotransfected with either 1 µg of the human Sp1 expression vector (pEVR2/Spl) or 1 µg of the control expression vector (pEVR2/0 plasmid). Cells were harvested 24 h after transfection, and luciferase activity was determined from cell extracts using the luciferase assay system (Promega). The β-galactosidase activity was measured using the β-galactosidase enzyme assay system (Promega). To normalize transfection efficiency for the cotransfection experiments, total DNA was extracted from each sample, and Dot-blot was performed. The nitrocellulose membrane was hybridized with a probe corresponding to a fragment of the luciferase reporter gene. Densitometry scanning of the

FIG. 1. Organization of the gene and isolated genomic clones encoding the human α1(XV) collagen chain. The locations of the 42 exons are indicated by vertical bars and the intervening sequences by horizontal lines. The genomic clones shown underneath are aligned according to the positions of the restriction sites EcoRI (E) and SacI (S). The dashed lines indicate that the corresponding clones encode additional noncharacterized sequences. The scale is marked with a bar indicating a length of 10 kb.

EXPERIMENTAL PROCEDURES

Isolation and Characterization of Genomic Clones—Radioactively labeled human cDNA clones for type XV collagen were used as probes for screening human genomic libraries: a human lung fibroblast genomic library in the A F IX vector (944201; Stratagene), a human leukocyte genomic library in the vector EMBL-3 (HL1006d; CLONTECH), a human lymphocyte cosmid library in pGEM15 (951203; Stratagene), and a human genomic library in the cosmid vector PJB8 (a gift from Dr. Leena Ala-Kokko, University of Oulu, Finland). The screenings were performed under stringent conditions (17): hybridizations were carried out at 41 °C in 50% (v/v) formamide in 5× SSC (1× SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 6.8), 1% (w/v) bovine serum albumin, 1% Ficoll (w/v), 1% polyvinylpyrrolidone (w/v), 0.25 mg of denatured salmon sperm DNA/ml, and 0.1% (w/v) SDS. The final washes for the filters were carried out in 0.5× SSC, 0.1% SDS at 65 °C. The positive clones picked from the libraries were analyzed by restriction enzyme mapping and Southern blotting, and suitable restriction fragments were subcloned into the plasmid pBluescript SK (Stratagene).

Gaps in the genomic sequences not covered by the subclones were filled by the PCR method designed for the isolation of end fragments from yeast artificial chromosome clones (18). 100-ng aliquots of the isolated λ or cosmID DNA were blunt-end-digested separately with AluI, EcoRV, HaeIII, HincII, PvuII, RsaI, SmaI, and StuI in a 20-µl reaction containing the appropriate buffer as suggested by the supplier of the enzymes (Amersham Pharmacia Biotech). Next, 2.5 µl of 10× ligation buffer (1× ligation buffer = 50 mM Tris, pH 7.8, 10 mM MgCl2, 1 mM dithiothreitol, 1 mM ATP, and 25 µg/ml bovine serum albumin), 0.5 µl of 5 µl linker solution (18), 1 µl of water, and 1 µl of T4 DNA ligase (New England Biolabs Inc.) were added to the individual reactions, which were then incubated overnight at 12 °C. The incubation was stopped by adding 75 µl of water and heating for 10 min at 95 °C. Two µl of the diluted ligation mixture was used as the template in a 10-µl PCR containing Taq polymerase buffer (Promega; 50 mM KCl, 1.5 mM MgCl2, 0.1% Triton X-100, and 10 mM Tris-Cl, pH 9.0), 0.2 mM of each deoxynucleotide, 10 pmol of gene-specific 23-mer primer, 1 pmol of 5-mer primer (nucleotides 469 to +112 in Fig. 3) was 5'-end-labeled with T4 polynucleotide kinase and [γ-32P]ATP (3000 Ci/mmol, Amersham Pharmacia Biotech). The double-stranded probe (3× 108 cpm) was hybridized to 20 µg of total RNA from human skin fibroblasts in the presence of 80% formamide, 40 mM Pipes, pH 6.4, 400 mM NaCl, and 1 mM EDTA at 67 °C for 15 h. After hybridization, 300 µl of buffer (280 mM NaCl, 50 mM sodium acetate, pH 4.5, 4.5 mM ZnSO4) was added, and the mixture was digested with 800 units of S1 nuclease (Boehringer Mannheim) at room temperature for 20 min. The protected fragments were analyzed on a 6% polyacrylamide sequencing gel. 20 µg of yeast tRNA was used as a negative control. The exact sizes of the protected fragments were determined by comparison with adjacent deoxyribonucleotide sequencing reactions (21).

Nucleotide Sequencing and Sequence Analysis—The nucleotide sequences were determined by the Sanger deoxyribonucleotide chain termination method (21) either manually or using an automated DNA sequencing (Applied Biosystems). Vector-specific or sequence-specific 17-mer primers synthesized in an Applied Biosystems DNA synthesizer (Department of Biochemistry, University of Oulu) were used, and the nucleotide sequence data were analyzed by DNASIS (Amersham Pharmacia Biotech). Consensus sites for the binding of transcription factors were searched for in the Transcription Factor Data Base using the Sequence Analysis software package, Version 8.0 (Genetics Computer Group, Inc.).

Human Type XV Collagen Gene

Northern Blot Analysis—Human adult multisource Northern blots (7750-1 and 7759-1; CLONTECH) were hybridized under stringent conditions with 32P-labeled probes covering 33 kb of the intron 2 in the COL15A1 gene in the manner suggested in the manufacturer's protocol.

Deletion Constructs for Promoter Analysis—Five deletion constructs consisting of different lengths of 5'-flanking sequences of the human type XV collagen gene were made. All of the fragments were restriction enzyme-digested from a HindIII subclone derived from a cosmid clone HG-23 (Fig. 1) and subcloned into the pGL2-β-Vector (Promega) upstream of the luciferase gene. An Espl restriction site at the position +27 was utilized as a common 3'-end for all of the constructs. A linker primer containing restriction sites EspI-SalI-HindIII was attached to the 3'-end of all the constructs, and a HindIII site from a pGL2-β-Vector (Promega) was used in subcloning. As the 5'-ends of the constructs, different restriction sites were used (HindIII for del 1, NheI for del 2, XhoI for del 3, XbaI for del 4, and XbaI for del 5). Accordingly, the 5'-subcloning position in the vector depended on the construct, so that del 1 was subcloned as a HindIII fragment, del 2 as a SmaI-HindIII fragment, del 3 as a XhoI-HindIII fragment, del 4 as a NheI-HindIII fragment, and del 5 as a SacI-HindIII fragment. Deletion constructs used in promoter analysis consisted of the following fragments: del 1, bp –3598 to +27; del 2, bp –2615 to +27; del 3, bp –1850 to +27; del 4, bp –1117 to +27; and del 5, bp –474 to +27.

Cell Culture and Transfection Assays—HeLa cells were routinely maintained at 37 °C in Dulbecco's modified Eagle's medium (Imperial) supplemented with 10% fetal calf serum, 50 µg of ascorbate per ml, 2 mM glutamine, 100 units/ml of penicillin, and 50 µg/ml of streptomycin. HeLa cells were transiently transfected with a liposome-based method (DOTAP liposomal transfection reagent kit, Boehringer Mannheim), according to the manufacturer's protocol. Briefly, the various luciferase deletion constructs (5 µg) were transfected with 1 µg of pCMV-β-galactosidase plasmid (CLONTECH) to normalize for transfection efficiencies. For cotransfection experiments, 5 µg of luciferase plasmids were cotransfected with either 1 µg of the human Sp1 expression vector (pEVR2/Spl plasmid) or 1 µg of the control expression vector (pEVR2/0 plasmid). Cells were harvested 24 h after transfection, and luciferase activity was determined from cell extracts using the luciferase assay system (Promega). The β-galactosidase activity was measured using the β-galactosidase enzyme assay system (Promega). To normalize transfection efficiency for the cotransfection experiments, total DNA was extracted from each sample, and Dot-blot was performed. The nitrocellulose membrane was hybridized with a probe corresponding to a fragment of the luciferase reporter gene. Densitometry scanning of the
RESULTS AND DISCUSSION

Isolation and Characterization of Genomic Clones—The isolation and characterization of the seven extreme 3' exons of the genomic lambda clone HLF-15 (Fig. 1) encoding part of the human α1(XV) chain gene has been described previously (8). In order to isolate additional clones, the same lambda library that yielded clone HLF-15 was screened four times using different fragments of the human type XV collagen cDNA (8) as probes. These screenings resulted in the isolation of five new clones, HLF-3, HLF-5, HLF-13, HLF-17, and HLF-18 (Fig. 1).

To find genomic sequences covering the gap between clones HLF-13 and HLF-15 (Fig. 1), two additional human genomic libraries, a human leukocyte library, and a cosmid library in the vector PJIB8 (see “Experimental Procedures”), were screened with the cDNA clone PF-19 (8). This resulted in the isolation of one clone from each library, a 12-kb lambda clone HL-1-1 and a 30-kb cosmid clone C-1-8-1. These contained sequences that overlapped with each other and with the clones HLF-13 and HLF-15.

To characterize the extreme 5'-end of the gene, a human lymphocyte cosmid library was screened with the cDNA clone F-10 encoding the extreme 5’ sequences of the human type XV collagen cDNAs (8). This screening gave two positive clones, of which HG-23, with an insert of 38 kb, was characterized further and was found to code for the missing N-terminal sequences and also several kb of 5'-flanking sequences.

Identification of the Transcription Initiation Site and Sequences of the 5'-Flanking Region of the Gene—The transcription initiation site of the gene was determined by S1 nuclease protection analysis (Fig. 2). A double-stranded SacI-BanI DNA fragment corresponding to the sequence −469 to +112 in Fig. 3 was isolated and 5'-end-labeled with γ-32P. This probe was then hybridized to total RNA isolated from cultured human skin fibroblasts. When the hybrids were subjected to nuclease S1 digestion, three major protected fragments of sizes 120, 164, and 165 nucleotides and nine minor bands were detected (Fig. 2). Comparison of the sizes of the protected fragments with adjacent dideoxynucleotide sequencing reactions indicated that a major transcription initiation site is located at an adenosine (A) nucleotide and another at two thymidines (T) 44–45 nucleotides upstream of this. Because the former showed the stronger band and was accompanied by several other initiation sites, it is marked +1 in Fig. 3.

About 3.6 kb of the 5'-flanking region of the gene was sequenced and further studied by computer analysis using the Transcription Factor Data Base program. The program predicted a promoter area between −398 and −142, which corresponds well to the transcription initiation sites predicted by S1 nuclease protection assay. There is no TATAA box in the vicinity of the transcription start sites, but the 5'-flanking region of the gene contains a TATAA sequence located between −404 and −400 relative to the predicted major transcription start site (Fig. 3). This motif may not be functional, however, in view of results obtained by S1 nuclease protection assay (and the overall structure of the 5'-flanking region). If it were functional, transcription initiation would occur about 30 nucleotides downstream. Furthermore, in the presence of a functional TATAA box, transcription initiation should start from a very
precise area, so that the lack of a functional TATAA motif agrees with the presence of multiple transcription initiation sites. The sequence covering nucleotides from -474 to the ATG was found to be rich in G+C (68.4%) and to contain consensus motifs for several transcription factors, some of which are shown in Fig. 3. This region contains four potential Sp1 binding sites, for example, and there is also one Sp1 binding site in the first intron. Recently a new protein binding sequence known as multiple start site element downstream (MED-1) was detected in many TATAA-less promoters with multiple start sites (22). This consensus sequence GCTCC(C/G) is found downstream of the mapped transcription initiation sites in the human COL15A1 gene (Fig. 3).

Deletion Analysis of the COL15A1 Promoter—To investigate the functional properties of the human COL15A1 promoter, we performed reporter gene analysis using various deletions con-
Human Type XV Collagen Gene

Table I

Exon-intron organization of the gene coding for the human α1(XV) collagen

| No. | Location of 5′-end of exon | Location of 3′-end of exon | Exon size (bp) | Intron size (bp) | Features |
|-----|-----------------------------|-----------------------------|---------------|-----------------|----------|
| 1   | 5′-UTR*a                     | NC1                         | 120           | 89              | Junction exon |
| 2   | NC1                          | NC1                         | 89            | ~55,000         |          |
| 3   | NC1                          | NC1                         | 548           | 1172            | tsp-1 homology |
| 4   | NC1                          | NC1                         | 75            | 1800*e           | tsp-1 homology |
| 5   | NC1                          | NC1                         | 81            | 7500*e           |          |
| 6   | NC1                          | NC1                         | 148           | 3900*e           |          |
| 7   | NC1                          | NC1                         | 113           | 2500*e           |          |
| 8   | NC1                          | NC1                         | 135           | 1400*e           |          |
| 9   | NC1                          | NC1                         | 153           | 13,600*e         |          |
| 10  | NC1                          | NC1                         | 150           | 408*e            |          |
| 11  | NC1                          | NC1                         | 144           | 430*e            |          |
| 12  | NC1                          | COL1                        | 54            | 4780*e           |          |
| 13  | COL1                         | NC2                         | 60            | 1250*e           |          |
| 14  | NC2                          | NC2/COL2                    | 93            | 1424*e           |          |
| 15  | NC2/COL2                     | COL2                        | 126           | 903*e            |          |
| 16  | COL2                         | COL2                        | 63            | 9070*e           |          |
| 17  | COL2                         | COL2                        | 36            | 432*e            |          |
| 18  | COL2                         | NC3                         | 141           | 195*e            |          |
| 19  | NC3                          | NC3/COL3                    | 69            | 751*e            |          |
| 20  | NC3/COL3                     | COL3                        | 36            | 107*e            |          |
| 21  | COL3                         | NC4                         | 84            | 2250*e           |          |
| 22  | NC4                          | COL4                        | 66            | 1800*e           |          |
| 23  | COL4                         | COL4                        | 36            | 1400*e           |          |
| 24  | COL4                         | COL4                        | 63            | 2500*e           |          |
| 25  | COL4                         | COL5                        | 69            | 98*e             |          |
| 26  | COL5                         | COL5                        | 36            | 3000*e           |          |
| 27  | COL5                         | COL5                        | 54            | 101*e            |          |
| 28  | COL5                         | NC6                         | 69            | 1701*e           |          |
| 29  | NC6                          | NC6                         | 42            | 2500*e           |          |
| 30  | NC6                          | NC6                         | 46            | 2200*e           |          |
| 31  | NC6                          | NC6/COL6                    | 81            | 444*e            |          |
| 32  | NC6/COL6                     | COL6                        | 50            | 108*e            |          |
| 33  | COL6                         | NC7/COL7                    | 111           | 850*e            |          |
| 34  | NC7/COL7                     | COL8                        | 145           | 4300*e           |          |
| 35  | COL8                         | NC9                         | 65            | 1500*e           |          |
| 36  | NC9                          | NC10                        | 129           | 146*e            |          |
| 37  | NC10                         | NC10                        | 60            | 710*e            |          |
| 38  | NC10                         | NC10                        | 60            | 3000*e           |          |
| 39  | NC10                         | NC10                        | 186           | 1500*e           |          |
| 40  | NC10                         | NC10                        | 116           | 1080*e           |          |
| 41  | NC10                         | NC10                        | –             | 1119*e           |          |
| 42  | NC10                         | NC10                        | –             | 1119*e           |          |

*a UTR, untranslated region.

b Size of the exon calculated from the major transcription start site as determined by S1 nuclease assay (see “Results”).

c Size of an intron determined by sequencing.

d Size of an intron determined by restriction mapping.

e Size of an intron determined by PCR.

f 211-nucleotide coding sequences followed by 3′-UTR sequences.

structs. A series of 5′ deletions from bp −3598 to +27 were constructed from the human promoter and linked to the luciferase reporter gene. Transient transfection experiments were carried out in HeLa cells, which express collagen type XV, in five independent experiments, each run in duplicate (Fig. 4, A and B). After normalization on β-galactosidase activity, all the promoter constructs exhibited similar luciferase activity. Consistent with these data, the shortest promoter fragment, from bp −474 to +27, is sufficient to give the entire promoter activity for the COL15A1 gene in HeLa cells.

**Cotransfections with Sp1 Expression Vector**—Because the sequence from bp −474 to the transcription start site was found to be rich in G+C and to contain four potential Sp1 binding sites, we investigated whether Sp1 has a potential role in the regulation of the COL15A1 gene. The different deletion constructs were cotransfected in HeLa cells with a human Sp1 expression vector or with the corresponding vector without the Sp1 cDNA (control). Results are expressed for each deletion construct as a ratio of the relative luciferase activity obtained with the Sp1 expression vector to that obtained with the control (Fig. 4C). Although basal luciferase activity obtained with the negative control pGL2-Basic vector was not changed, cotransfection with the Sp1 expression vector induced the promoter activity of all constructs from 5.5-fold for the longest construct to 10.3-fold for the shortest one. These results suggest that Sp1 is involved in the regulation of the human type XV collagen gene.

**Exon-Intron Organization of the Human Gene for the α1(XV) Collagen Chain**—DNA sequencing of the genomic clones indicated that the human type XV collagen gene consists of 42 exons and 41 introns (Fig. 1). Sequences were determined for all the exons, their intron junctions, most of the intronic sequences of reasonable size and about 3.6 kb of the 5′-flanking region of the gene. Exons 1–41 vary in size from 36 to 548 bp, whereas the extreme 3′ exon is 1119 bp in length, containing 908 bp of 3′-untranslated sequences (Table I). The introns vary in length between 89 bp and about 55 kb (Table I). The various overlapping genomic clones covered the entire gene with the exception of introns 2 and 9, the sizes of which were obtained by Southern blotting of genomic DNA. The exon-intron boundaries (Table II) agree with published consensus sequences for splice donor and acceptor sites (23). The donor site following exon 6 is unusual in that the normally invariant GT dinucleotide is replaced by GC. Fewer than 30 examples of GC donors have
been observed among the thousands of donor sites catalogued thus far (23–25). Two other examples of GC splice donors in collagen genes, **COL4A1** and **COL7A1**, have been reported (26, 27).

The data indicate that the human type XV collagen gene is quite large, its transcribed region being about 145 kb. The coding information is unevenly distributed in the gene, because about 90 kb of 5' genomic sequences contains only the first 11 exons encoding the N-terminal noncollagenous domain, whereas the rest of the gene, exons 12–42, lies within a 55-kb genomic area. The longest intron, intron 2, is about 55 kb in size, and in order to find out whether this contains coding sequences, we hybridized Northern blots containing poly(A)⁺ RNA isolated from human heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon, and peripheral blood leukocytes with probes covering those of its regions included in clones HG-23 and HLF-17, but no signal was detected.

Exons 12–36 cover collagenous sequences with multiple interruptions, exons 12 and 36 themselves being junction exons encoding both noncollagenous terminal domains and collagenous sequences (Fig. 1 and Table I). Because type XV collagen contains several interruptions in the collagenous sequence, many of the exons encoding this region contain both collagenous and noncollagenous sequences. There are eight exons encoding solely collagenous sequences (exons 16, 17, 20, 23, 24, 26, 27, and 32) (Table I), of which four are 36 bp in length, one is 54 bp, two are 63 bp, and one is 81 bp. With the exception of...
more than type XV collagen, and consequently, the homologous domains differ by one in their numbering. Exons that begin with split codons are marked with asterisks. Type XVIII collagen has one collagenous domain and one noncollagenous domain noncollagenous N and C-terminal sequences. The exons encoding the central collagenous sequences are, on average, shorter than those encoding flanking tryptophan. The exons encoding the central collagenous domain contain sequences coding for the central collagenous domain contain sequences showing homology between the C-terminal noncollagenous domain, and exon 42 begins with a split codon for a glutamic acid. This feature is also conserved in the human and mouse thrombospondin genes (28, 29), demonstrating marked genomic conservation of this sequence motif. Five additional split codons are conserved in the type XV and type XVIII collagen genes (Table III). The lengths of the exons coding for the collagenous sequences showing homology between the type XV and type XVIII chains (Table III) and the locations of the respective exon-intron junctions are identical or almost identical in the two genes. As described previously, the marked homology between the C-terminal noncollagenous domains of the type XV and type XVIII chains extends to the exon-intron organization of the corresponding regions of their genes (13). The last three exons encoding the most conserved portion of the polypeptides share a 55–61% homology at the nucleotide level (Table III).

**Exon 32**, all of them begin with a complete codon for glycine, which is characteristic of collagen genes. The 54-bp exon is typical of the fibril-forming collagen genes, whereas 36- and 63-bp exons are found in several genes encoding nonfibril-forming collagens (1, 5, 6).

Altogether, seven exons in the type XV collagen gene begin with a split codon (Table II); three of them are located in the genomic region encoding the N-terminal noncollagenous domain. The large 548-bp exon homologous to thrombospondin-1 begins with a split codon for a glutamic acid. The genomic sequences coding for the central collagenous domain contain three split codons, two of which are located in consecutive exons in the collagenous sequence and one of which is located in the junction exon between the last collagenous domain and the C terminus. Exons 37–42 encode the C-terminal noncollagenous domain, and exon 42 begins with a split codon for a tryptophan. The exons encoding the central collagenous sequences are, on average, shorter than those encoding flanking noncollagenous N and C-terminal sequences.

Comparison of the Human a(XV) and Mouse a(XVIII) Collagen Genes—The human type XV and mouse type XVIII collagen genes are of somewhat different sizes, the former being about 145 kb in size and the latter about 102 kb, and they have 42 and 43 exons, respectively. They are highly similar in their exon-intron organization, but the introns in the type XV gene are in most cases longer than those in the type XVIII gene.

The first intron is less than 100 bp in both genes, whereas the second is conspicuously longer, 55 and 50 kb in the type XV and XVIII genes, respectively. The 548-bp exon 3 in the type XV gene and the 551-bp exon 4 in the type XVIII gene (14) are both homologous to thrombospondin-1, and both begin with a split codon for a glutamic acid. This feature is also conserved in the human and mouse thrombospondin genes (28, 29), demonstrating marked genomic conservation of this sequence motif. Five additional split codons are conserved in the type XV and type XVIII collagen genes (Table III). The lengths of the exons coding for the collagenous sequences showing homology between the a(XV) and a(XVIII) chains (Table III) and the locations of the respective exon-intron junctions are identical or almost identical in the two genes. As described previously, the marked homology between the C-terminal noncollagenous domains of the a(XV) and a(XVIII) chains extends to the exon-intron organization of the corresponding regions of their genes (13). The last three exons encoding the most conserved portion of the polypeptides share a 55–61% homology at the nucleotide level (Table III).

**Conclusions**—The genes encoding the fibril-forming collagens range in size from 18 to 53 kb and consist of over 50 exons (1, 5, 6), whereas those encoding the nonfibril-forming collagens show more extensive heterogeneity in their genomic organization: they vary in size from 5 kb for COL10A1 (30) to 750 kb for COL5A1 (31) and in number of exons from 3 for COL10A1 (30) to 118 for COL7A1 (26). The present characterization of the complete exon-intron structure of the COL15A1 gene, showing it to be about 145 kb in size and to contain 42 exons, makes it one of the largest collagen genes, with a typically high number of exons.

The exon pattern of the COL15A1 gene differs markedly from that of the fibril-forming collagen genes, in which the triple-helix is encoded predominantly by exons of 54 bp or multiples of this (1, 5, 6). Only one of the exons in the COL15A1 gene that code for purely collagenous sequences is 54 bp in size.
In fact, none of the nonfibril-forming collagen genes characterized so far displays the 54-bp exon pattern observed in the fibril-forming collagen genes, whereas many of them, including COL15A1, typically contain 36- and 63-bp exons, in addition to exons of more variable length, encoding the interrupted collagenous sequences (1, 5, 6).

The 5'-flanking region of COL15A1 is characterized by the lack of a TATAA motif and the presence of several GC motifs. This renders the 5'-flanking region of COL15A1 similar to promoters of the "housekeeping genes," which are transcribed widely but at low RNA levels in many tissues. Several other collagen genes also contain multiple GC boxes as their main promoter elements, including the COL5A1, COL7A1, COL11A1, and COL11A2 genes (26, 32–34). In addition, the downstream promoter of COL6A2 (35) and the upstream promoter of Col18a1 (14), two collagen genes with alternate promoters, are also of this kind. Transient transfection experiments, which were performed on HeLa cells with 5' deletion constructs ranging from bp −3598 to +27, indicated that the shortest promoter fragment, from bp −474 to +27 had the same promoter strength as all the longer constructs. Furthermore, cotransfections with the Sp1 expression vector directly demonstrated that this transcription factor could regulate the expression of the human type XV collagen gene in HeLa cells through binding to one or more of the four Sp1 sites within this fragment.

The COL15A1 and Col18a1 genes show marked structural similarity. The mouse Col18a1 gene contains one exon more than the human COL15A1 gene, but it is about 40 kb smaller in size, thus presenting a nonconserved picture of the sizes of the introns. This may also be due to species differences. Their homology covers 25 exons that are nearly identical in size, 6 of which contain conserved split codons (Table III). The homologous exons are spread throughout the entire gene, including the closely adjacent first pair of exons and the exons manifesting the thrombospondin-1 homology, which are followed by the most variable region of the two genes, covering part of the noncollagenous domain 1 and the beginning of the collagenous portion, after which most of the exons are homologous. The homology is most pronounced in the region encoded by the last three large exons (Table III). The second exon is large in both genes, over 50 kb. It is typical of collagen genes that they possess regulatory elements in intron 1, e.g., in the COL1A1 and COL2A1 genes (5). Intron 1 of the COL15A1 gene is only 89 bp in length, and that of Col18a1 is only 71 bp, but it is highly possible that the large intron 2 in both genes may contain such elements. One notable difference likely to occur between the two genes is that the COL15A1 gene lacks sequences corresponding to exon 3 of the Col18a1 gene (14), which encodes a cysteine-rich region of the mouse α1(XVIII) collagen chain non-collagenous domain 1 homologous to rat and Drosophila frizzled proteins (36). This cysteine-rich sequence has not been found in any of the human type XV collagen cDNA clones characterized so far (8, 9), or in any of the mouse type XV collagen cDNAs (16), suggesting that this exon is indeed absent from COL15A1. Furthermore, our Northern blotting experiments with probes covering 33 kb of intron 2 did not reveal any mRNA signals. All in all, the comparison of the two genes clearly points to a common ancestor.

Collagen types XV and XVIII have both similarities and differences in tissue distribution. Both are prominently synthesized by endothelial cells in practically all tissues studied (37, 38). The most obvious differences are the strong type XV collagen expression in muscle tissue, where type XVIII is present in much lower amounts, whereas the opposite situation prevails in liver tissue (12, 38). There are no data available on the mechanisms that regulate the expression of these genes, but the tissue distribution data suggest that these mechanisms are not identical despite their otherwise extensive homology.

Acknowledgments—We gratefully thank Ritva Savilaakso and Jaana Vaisinén for expert technical assistance. We gratefully acknowledge Dr. G. Suske (Institut für Molekularbiologie und Tumorforschung, Marburg, Germany) for the pEVR2/Sp1 expression vector.

REFERENCES
1. Vuorio, E., and de Crombrugghe, B. (1990) Annu. Rev. Biochem. 59, 181–198
2. Kivirikko, K. I. (1993) Annu. Rev. Genet. 27, 113–136
3. Pihlajaniemi, T., and Rehn, M. (1995) Prog. Nucleic Acid Res. Mol. Biol. 50, 217–224
4. Chu, M.-L., and Prockop, D. (1993) in Connective Tissue and Its Heritable Disorders. Molecular, Genetics, and Medical Aspects. (Royce, P. M., and Steinmann, B., eds) pp. 149–165, Wiley-Liss, Inc., New York
5. Takemura, S., Lamande, S. R., and Ramshaw, J. A. M. (1996) in Extracellular Matrix (Comper, W. D., ed) Vol. 2, pp. 22–67, Harwood Academic Publishers, Amsterdam
6. Kivirikko, S., Heinämäki, P., Rehn, M., Honkanen, N., Myers, J. C., and Pihlajaniemi, T. (1984) J. Biol. Chem. 259, 4775–4779
7. Muragaki, Y., Abe, N., Ninomiya, Y., Olsen, B. R., and Ooshima, A. (1994) J. Biol. Chem. 269, 4042–4046
8. Oh, S. P., Kamagata, Y., Muragaki, Y., Timmons, S., Oshimaa, A., and Olsen, B. R. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 4229–4233
9. Oh, S. P., Warman, M. L., Seldin, M. F., Cheng, S.-D., Knoll, J. H. M., Timmons, S., and Olsen, B. R. (1994) Genomics 19, 494–499
10. Rehn, M., and Pihlajaniemi, T. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 4234–4238
11. Rehn, M., Hintikka, E., and Pihlajaniemi, T. (1994) J. Biol. Chem. 269, 13929–13935
12. Rehn, M., Hintikka, E., and Pihlajaniemi, T. (1996) Genomics 22, 436–446
13. Huebner, K., Cannizzaro, L. A., Jabs, E. W., Kivirikko, S., Manzone, H., Pihlajaniemi, T., and Myers, J. C. (1992) Genomics 14, 220–224
14. Hagg, P. M., Hofreiter-Kuitunen, N., Ekland, L., Palote, A., and Pihlajaniemi, T. (1997) Genomics 35, 31–41
15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., pp. 7.58–7.65 and 9.1–9.62, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
16. Rehn, M., Pihlajaniemi, T., and Seldin, M. F. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 19273–19278
17. Soininen, R., Huotari, M., Ganguly, A., Prockop, D. J., and Tryggvason, K. (1995) J. Biol. Chem. 270, 13556–13571
18. Wolf, P. W., Eddy, R. L., Shows, T. B., and Dixit, V. M. (1990) Genomics 6, 685–691
19. Lawler, J., Duquette, M., Ferro, P., Copeland, N. G., Gilbert, D. J., and Jenkins, N. A. (1991) Genomics 11, 587–609
20. Apte, S. S., Seldin, M. F., Hayashi, M., and Olsen, B. R. (1994) Eur. J. Biochem. 260, 217–224
21. Takahara, K., Hoffma, G. G. & Greenspan, D. S. (1995) Genomics 29, 588–597
22. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156–159
23. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
24. Ince, T. A., and Scotto, K. W. (1995) J. Biol. Chem. 270, 30249–30252
25. Shapiro, M. B., and Senapathy, P. (1987) Nucleic Acids Res. 15, 7155–7174
26. Senapathy, P., Shapiro, M. B., and Harris, N. L. (1990) Methods Enzymol. 183, 252–278
27. Jackson, I. J. (1991) Nucleic Acids Res. 19, 3795–3798
28. Christiano, A. M., Hoffman, G. G., Chung-Honet, L. C., Lee, S. W., Cittus, H., and Greenspan, D. S. (1994) Genomics 21, 169–179
29. Sohinen, R., Huotari, M., Ganguly, A., Prockop, D. J., and Tryggvason, K. (1995) J. Biol. Chem. 270, 13565–13571
30. Wolf, P. W., Eddy, R. L., Shows, T. B., and Dixit, V. M. (1990) Genomics 6, 685–691