Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 4229–4233, May 1994
Biochemistry

Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins

SUK PAUL OH*, YUSUKE KAMAGATA†, YASUTERU MURAGAKI‡, SHEILA TIMMONS*, AKIRA OOSHIMA‡, AND BJORN R. OLSEN*

*Department of Cell Biology, Harvard Medical School, 220 Longwood Avenue, Boston, MA 02115; †Department of Oral Bacteriology, Ohu University School of Dentistry, 31-1 Misumido, Tomita-machi, Koriyama-shi, Fukushima 963, Japan; and ‡Department of Pathology, Wakayama Medical College, 9 Bancho, Wakayama 640, Japan

Communicated by Elizabeth D. Hay, January 3, 1994 (received for review June 7, 1993)

ABSTRACT **We have isolated overlapping mouse cDNAs encoding a collagenous polypeptide that we have designated α1(XVIII) collagen. Nucleotide sequence analysis shows that α1(XVIII) collagen contains 10 triple-helical domains separated and flanked by non-triple-helical regions. Within the non-triple-helical regions, there are several Ser-Gly-containing sequences that conform to consensus sequences for glycosaminoglycan attachment sites in proteoglycan core proteins. Northern blots show that α1(XVIII) transcripts are present in multiple organs, with the highest levels in liver, lung, and kidney. We have also isolated overlapping cDNAs encoding human α1(XV) collagen, and their sequence extends a published partial α1(XV) sequence to the 3′ end. Comparison of the α1(XV) and α1(XVIII) sequences reveals a striking similarity in the lengths of the six most carboxyl-terminal triple-helical domains. In addition, within the carboxyl non-triple-helical domain NC1 of the two chains, a region of 177 amino acid residues shows about 60% identity at the amino acid level. We suggest, therefore, that α1(XV) and α1(XVIII) collagens are structurally related. Their structure is different from that of other known collagen types. We conclude that they belong to a subfamily of extracellular matrix proteins and we suggest the designation multiplexins (for protein with multiple triple-helix domains and interruptions) for members of this subfamily.**

The ability of collagenous proteins to form structures of high tensile strength is based on the rigid structure of collagen molecules. Collagen polypeptides contain one or more blocks of Gly-Xaa-Yaa repeats, in which Yaa frequently represents a proline or hydroxyproline residue. The presence of such sequence repeats allows groups of three polypeptides to fold into triple-helical domains which are rigid and inextensible. The use of such triple-helical domains was initially thought to be limited to molecules that make up collagen fibrils in tissues, but it is now known that such domains are present in a large number of proteins. Most of these proteins are found in the extracellular matrix and serve a structural role; thus they are considered members of the collagen superfamily of proteins. Within the superfamily of collagens, the fibrillar collagens represent a distinct family. Their triple-helical domains polymerize in a staggered fashion to form fibrils (1–3).

Members of other collagen families do not by themselves form cross-striated fibrils but may be associated with fibrils (fibril-associated collagens with interrupted triple helices, FACIT) (4) or form their own distinct polymers (5, 6). The lengths as well as the number of triple-helical domains within molecules of nonfibrillar collagens are frequently quite different from those of triple-helical domains in fibrillar collagens.

The non-triple-helical domains that separate triple-helical domains in some nonfibrillar collagens represent regions of flexibility. For example, in types IX, XII, and XIV collagens, non-triple-helical hinges allow triple-helical domains on either side to be oriented in a variety of directions (7–9). Whether this is the case also in other collagens with short triple-helical domains, such as α1(XVI) (10) and α1(XVII) (11), is not known, but it is conceivable that one function of non-triple-helical domains in such collagen types is to provide for flexibility between triple-helical regions.

Here we report on a collagenous polypeptide with multiple regions of such potential flexibility. This polypeptide, which we designate α1(XVIII) collagen, contains 10 triple-helical domains separated and flanked by non-triple-helical regions. Within the non-triple-helical regions are several Ser-Gly-containing sequences that conform to consensus sequences for glycosaminoglycan attachment sites in proteoglycan core proteins (12). The α1(XVIII) collagen gene is expressed in multiple organs with the highest levels of RNA in liver, lung, and kidney. Comparison of the α1(XVIII) sequence with recently published sequences of α1(XV) collagen (13), as well as with sequences of additional α1(XV) cDNAs reported in this paper,[§] reveals a striking similarity in the lengths of the six most carboxyl-terminal triple-helical domains. In addition, within NC1, the non-triple-helical carboxyl domain of the two collagen chains, the carboxyl-most region of 177 aa shows about 60% identity [for mouse α1(XVIII) versus human α1(XV)] at the amino acid level. We suggest, therefore, that α1(XV) and α1(XVIII) collagens belong to a subfamily of collagenous proteins with multiple short triple-helical domains. Members of this subfamily, for which we propose the designation multiplexin (for multiple triple-helix domains and interruptions), share a highly homologous non-triple-helical carboxyl domain. Rehn and Pihlajaniemi (14) have independently isolated cDNAs that also code for α1(XVIII) collagen, and differences between the 5′ regions of their sequences and our clones suggest that the α1(XVIII) gene is transcribed from alternative transcription start sites or gives rise to alternatively spliced transcripts.

MATERIALS AND METHODS

Screening of cDNA Libraries and Nucleotide Sequencing. For isolation of mouse α1(XVIII) collagen cDNAs, two cDNA libraries purchased from Clontech were screened. The first library contained cDNA synthesized with RNA isolated from 15.5-day mouse embryos and cloned into the *Eco*RI site of λgt10. For screening, a 1.3-kb *Hin*dIII–*Eco*RI fragment of the mouse type XII collagen cDNA mXIIc5 was used as

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: 5′ RACE, rapid amplification of cDNA 5′ ends.

[§]The nucleotide sequences reported in this paper have been deposited in the GenBank database (accession no. L22545).

probe (15). One positive clone, mc19, containing a 1.5-kb insert, was isolated and characterized.

To extend the sequence defined by mc19, a second library was screened with a 0.5-kb *Eco*RI–*Apa* I fragment from the 5′ region of mc19. The second library contained cDNA synthesized with RNA isolated from 17.5-day mouse embryos and cloned into the *Eco*RI site of λgt11. This led to the isolation of mcE4, a clone containing a 2.4-kb insert. This λgt11 library was also screened with a 2.0-kb *Pst* I fragment of the human cDNA hc1-1 encoding the 3′ portion of α1(XVIII) (16), leading to the isolation of a third mouse cDNA, mc3, with a 3.7-kb insert.

For isolation of cDNAs encoding α1(XV) collagen, a cDNA library in λgt11 from human placenta (Clontech) was used. One probe was a cDNA fragment encoding an unidentified collagenous protein (23). A second probe was a 550-bp *Eco*RI–*Apa* I fragment from the 5′ end of the insert of the α1(XVIII) cDNA mc19 (see above). This led to the isolation of the cDNAs YMh46, YMh4, YKh17-1, and YKh17-2.

Nucleotide sequence was analyzed by the dideoxynucleotide chain-termination method (17).

mRNA Preparation and Rapid Amplification of cDNA 5′ Ends (5′ RACE). Livers were dissected from 2-month-old C57BL/6J mice. Total liver RNA was prepared by the guanidinium isothiocyanate method (18), and poly(A)$^+$ RNA was enriched on an oligo(dT) column from the FastTrack 2.0 kit (Invitrogen). The 5′ RACE procedures were slightly modified from Frohman (19). Adaptor-$(dT)_{17}$, containing *Xho* I, *Sal* I, and *Cla* I restriction enzyme sites (19), and specific antisense primers PS1 (5′-GTGACAGGAGGTGGCTGA-3′), PS2 (5′-TGTGTGACTTGCTGCTTT-3′), PS3 (5′-TAGCTCC AGTC-CCTGCGA-3′), and PS4 (5′-CCGAGCAAATGGCACCCT-3′) were synthesized on a Cyclone Plus oligonucleotide synthesizer (MilliGen).

Second-strand cDNA was synthesized with an aliquot of oligo(dA)-tailed cDNA by using the adaptor-$(dT)_{17}$ primer and *Taq* DNA polymerase. Initial denaturation at 94°C for 3 min was followed by annealing at 55°C and 37°C for 5 min each and extension at 72°C for 40 min. The specific primer PS1 was added to the reaction mixture and 30 cycles of first-round PCR were performed; initial denaturation at 94°C for 3 min was followed by annealing at 42°C for 90 sec, extension at 72°C for 150 sec, and denaturation at 94°C for 45 sec. One-tenth of the first-round PCR products was used for second-round nested PCR using the adaptor primer and PS2. The PCR conditions were the same as for the first round except that the annealing temperature was 52°C instead of 42°C. The nested PCR products were cut with the restriction enzymes *Sma* I and *Sal* I and subcloned into pBluescript (Stratagene). The subclones were screened by Southern blotting with a 145-bp *Eco*RI–*Sma* I fragment of mc19 as probe (Fig. 1*A*). One positive clone, SS8, contained a sequence matching that of the 5′ region of mc19 and extending 70 bp further in the 5′ direction. The specific primers PS3 and PS4 were used for first-round and nested PCR with the same strategy as described above. A major PCR product was obtained, purified by agarose gel electrophoresis, and subcloned into the modified *Eco*RV site of pBluescript. One clone, TA5, contained the 70 bp of sequence of the 5′ region of SS8 and extended 530 bp further in the 5′ direction.

Southern and Northern Hybridization. Filters containing DNA or mRNA were hybridized with probes labeled with the random-primer labeling method (Boehringer Mannheim) at 42°C overnight in 50% formamide/6× standard saline citrate/5% dextran sulfate, 1 mM EDTA/0.5% SDS/1× Denhardt's solution containing salmon sperm DNA at 25 μg/ml.

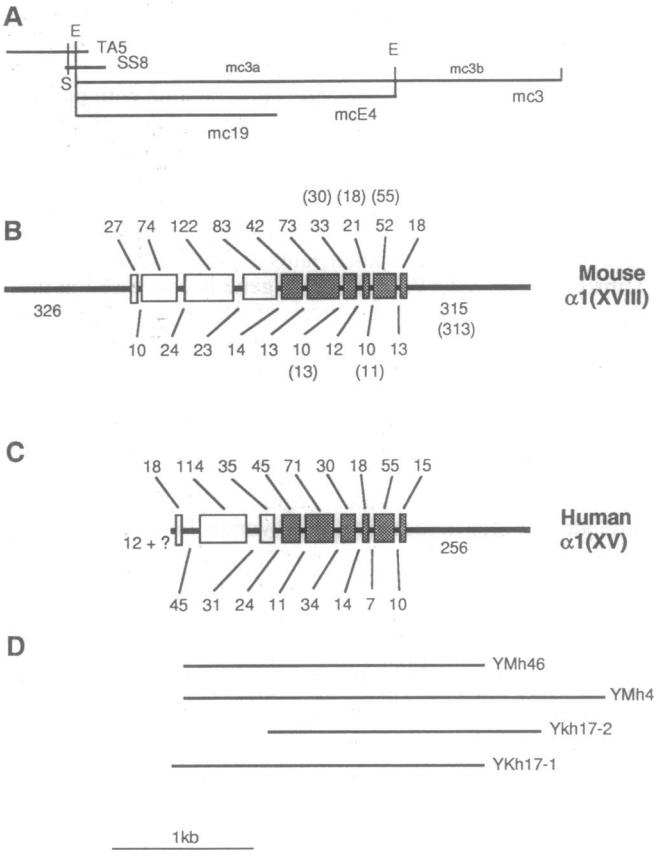

FIG. 1. Diagrams showing overlapping cDNAs that encode the complete mouse α1(XVIII) collagen chain, except the amino-terminal 27 aa (*A*); the domain structure of α1(XVIII) collagen (*B*); the domain structure of human α1(XV) (*C*); and overlapping cDNAs encoding human α1(XV) chain (*D*). Triple-helical domains are indicated by rectangular areas; non-triple-helical regions are indicated by heavy lines. Numbers indicate the length (in amino acid residues) of the domains. For the human α1(XVIII) chain only the carboxyl half of the sequences is known (16); within the human sequence the domains are slightly different in size from the mouse sequence. The human sizes (when different from the mouse) are indicated in parentheses above and below the numbers for the mouse. The six most carboxyl-terminal triple-helical domains in α1(XVIII) and α1(XV) are remarkably similar in size; these domains are shaded in the diagram. E, *Eco*RI; S, *Sma* I.

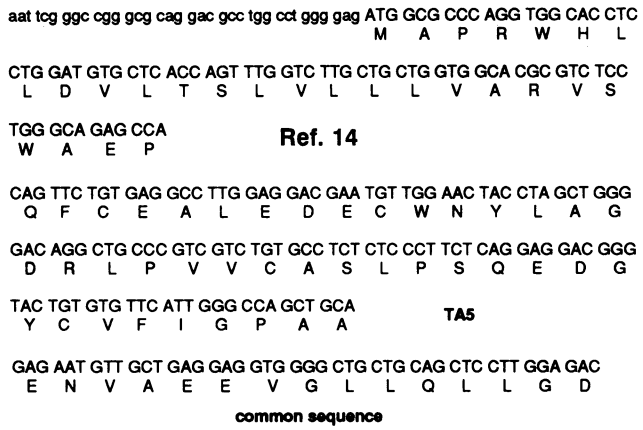

FIG. 2. For the 5′ region of the α1(XVIII) collagen mRNA we (*Middle*) and Rehn and Pihlajaniemi (14) (*Top*) have isolated cDNAs that differ in their 5′ region but have common 3′ sequences (*Bottom*, overlapping with the sequence shown in Fig. 3). The sequence of Rehn and Pihlajaniemi (14) contains a clear signal-peptide sequence, whereas our sequence (TA5) has a different reading frame that continues to the 5′ end.

RESULTS AND DISCUSSION

Primary Structure of α1(XVIII) Collagen. The first α1(XVIII) collagen cDNA, mc19, was isolated by cross-hybridization during screening of a cDNA library from 15.5-day whole mouse embryo with an α1(XII) collagen probe, mXIIc5 (15). Nucleotide sequencing of mc19 revealed that it encoded part of a polypeptide containing several triple-helical domains separated by non-triple-helical sequences (Fig. 1). Since mc19 contained an open reading frame without a trans-

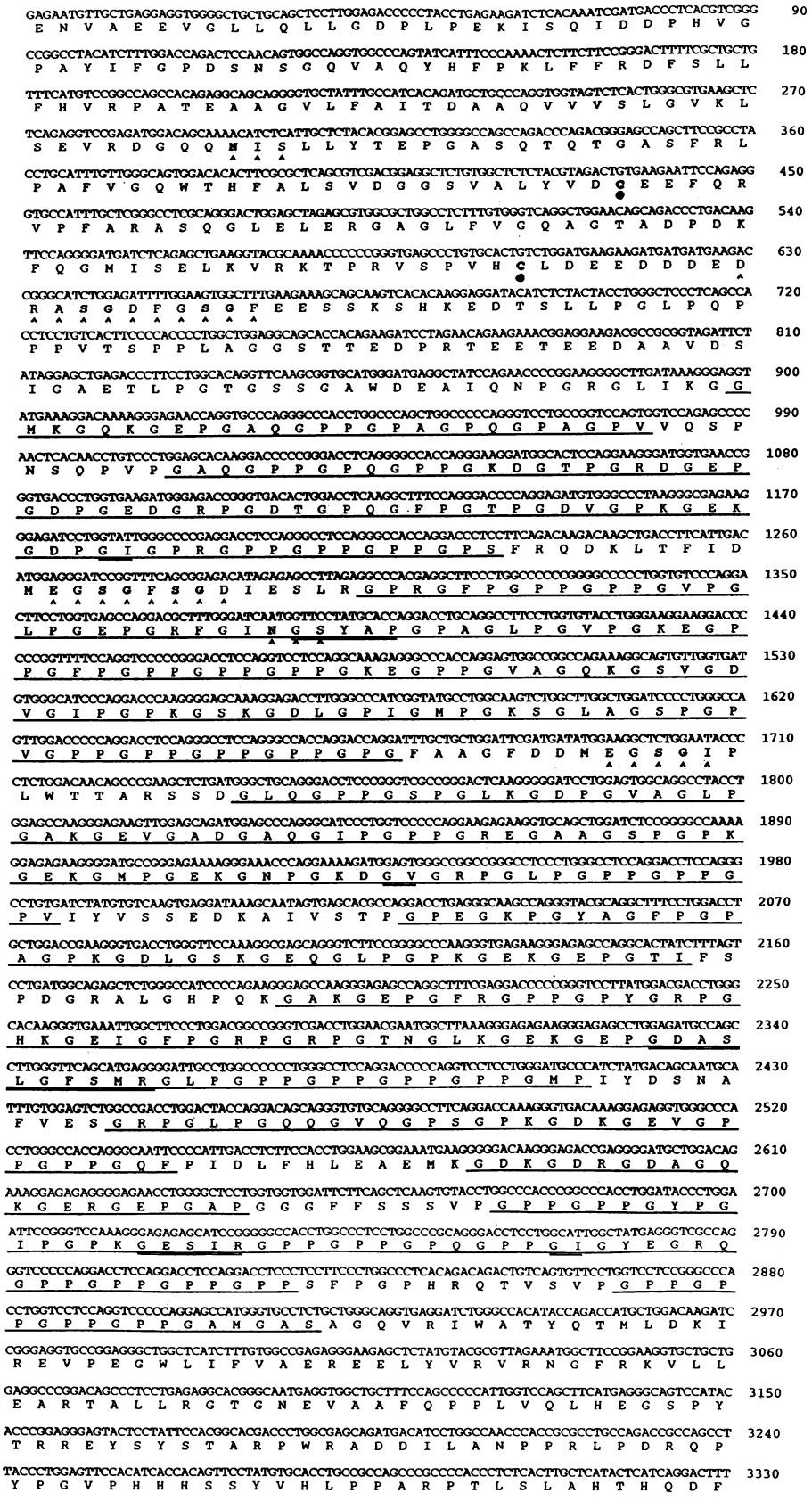

FIG. 3. *(Figure continues on the opposite page.)*

```
CAGCCAGTGCTCCACCTGGTGGCACTGAACACCCCCCTGTCTGGAGGCATGCGTGGTATCCGTGGAGCAGATTTCCAGTGCTTCCAGCAA  3420
 Q  P  V  L  H  L  V  A  L  N  T  P  L  S  G  G  M  R  G  I  R  G  A  D  F  Q  C  F  Q  Q
                                                                                 •

GCCCGAGCCGTGGGGCTGTCGGGCACCTTCCCGGGCTTTCCTGTCCTCTAGGCTGCAGGATCTCTATAGCATCGTGCGCCGTGCTGACCGG  3510
 A  R  A  V  G  L  S  G  T  F  R  A  F  L  S  S  R  L  Q  D  L  Y  S  I  V  R  R  A  D  R

GGGTCTGTGCCCATCGTCAACCTGAAGGACGAGGTGCTATCTCCCAGCTGGGACTCCCTGTTTTCTGGCTCCCAGGGTCAAGTGCAACCC  3600
 G  S  V  P  I  V  N  L  K  D  E  V  L  S  P  S  W  D  S  L  F  S  G  S  Q  G  Q  V  Q  P

GGGGCCCGCATCTTTTCTTTTGACGGCAGAGATGTCCTGAGACACCCAGCCTGGCCGCAGAAGAGCGTATGGCACGGCTCGGACCCCAGT  3690
 G  A  R  I  F  S  F  D  G  R  D  V  L  R  H  P  A  W  P  Q  K  S  V  W  H  G  S  D  P  S
                                                                        ^  ^  ^

GGGCGGAGGCTGATGGAGTTACTGTGAGACATGGCGAACTGAAACTACTGGGGCTACAGGTCAGGCCTCCTCCCTGCTGTCAGGCAGG  3780
 G  R  R  L  M  E  S  Y  C  E  T  W  R  T  E  T  T  G  A  T  G  Q  A  S  S  L  L  S  G  R
 ^  ^  ^           •

CTCCTGGAACAGAAAGCTGCGAGCTGCCACAACAGCTACATCGTCCTGTGCATTGAGAATAGCTTCATGACCTCTTTCTCCAAATAGGCC  3870
 L  L  E  Q  K  A  A  S  C  H  N  S  Y  I  V  L  C  I  E  N  S  F  M  T  S  F  S  K  *
                        •                    •

TCTGCCAGCTAGGGTGGCAGACAGAGGCCATGCAGAACTTTGACACAGCGCAGGGAGCATTCAGTCAGCACCCAGGGCTCTGGCTGGGAT  3960

ACAACTCCTGTATAGTTCCCATTTTTATGTAATCCTCAAGAAATAAAAGGAAGCCAAAGAGTAAAAAAAAA                    4031
```

FIG. 3. Complete nucleotide and corresponding amino acid sequence of mouse α1(XVIII) collagen, except for the amino-terminal 27 aa. Triple-helical regions are underlined. The imperfections in Gly-Xaa-Yaa repeats are doubly underlined. Cysteine residues are indicated by dots. Potential glycosaminoglycan attachment sites and N-linked glycosylation sites are indicated by a series of carets.

lation start or stop codon, we used it as a probe to isolate overlapping cDNAs that would extend the reading frame in both 5′ and 3′ directions. Screening of both mouse and human libraries led to the isolation of two additional overlapping mouse cDNAs (Fig. 1*A*). The mouse cDNAs, mc19, mcE4, and mc3, have a common 5′ end but vary in their lengths. The common 5′ end coincides with the 5′ *Eco*RI cloning site; this *Eco*RI site is an internal site in the α1(XVIII) sequence. The sequences of mc19 and mcE4 are contained within the sequence of mc3, except for an A-rich sequence of ≈70 nt at the 3′ end of mc19. We believe that this sequence represents a cloning artifact.

Preliminary sequence analysis showed that mc3 encodes the carboxyl end of the α1(XVIII) translation product. To extend the sequence in the 5′ direction, we used 5′ RACE with RNA isolated from mouse liver. This led to the isolation of the cDNAs SS8 and TA5. SS8 spans the *Eco*RI site at the 5′ end of mc19, mcE4, and mc3; TA5 extends the sequence further in the 5′ direction.

Except for the amino-terminal region the sequence of the overlapping cDNAs defines an open reading frame of 1288 aa. Comparison of this open reading frame with the sequence defined by independently derived cDNA clones characterized by Rehn and Pihlajaniemi (14) shows 100% identity at the nucleotide and amino acid levels within a region of 901 aa (the cDNA clones isolated by Rehn and Pihlajaniemi do not cover the carboxyl region of the polypeptide). Upstream of this open reading frame, however, the sequence defined by the cDNA TA5 differs from that described by Rehn and Pihlajaniemi (14). In fact, while the 3′ region of TA5 defines a sequence that is identical to that reported by Rehn and Pihlajaniemi, the 5′ region defines an open reading frame that is completely different (Fig. 2). We do not believe that the 5′-most region of TA5 represents a cloning artifact, since the same sequence is found in a cDNA isolated from a mouse lung library (data not shown). More likely is the possibility that the TA5 sequence represents an alternatively spliced transcript. Therefore, at least two forms of α1(XVIII) collagen may exist. One form contains a signal peptide of 25 aa and 2 additional residues, defined by the cDNA of Rehn and Pihlajaniemi (Fig. 2), attached to the open reading frame of 1288 aa (Fig. 3). A second form contains a larger amino-terminal non-triple-helical domain (the 5′ portion of this domain is defined by TA5) attached to the common open reading frame of 1288 aa (Figs. 2 and 3). Further studies are needed to establish whether additional variants may exist.

The amino acid sequence of α1(XVIII) collagen starting at residue 28 [as defined by Rehn and Pihlajaniemi (14)] and the corresponding nucleotide sequence are shown in Fig. 3. Also included are 155 nt of 3′ untranslated sequence and a short poly(A) tail. The sequence defines 10 domains of Gly-Xaa-Yaa repeats (COL domains), which are separated and flanked by non-triple-helical regions (NC domains). The COL domains, numbered from the carboxyl end of the polypeptide chain, vary in length from 18 aa (COL1) to 122 aa (COL8). Because of imperfections in the Gly-Xaa-Yaa triplet structure (Fig. 3), their lengths are not always an integral number of triplets.

Potential Glycosylation Sites in α1(XVIII) Collagen. Within the α1(XVIII) sequence there are two potential sites for N-linked glycosylation, one site in the amino-terminal NC11 domain and one site in COL8 (Fig. 3). In addition, six Ser-Gly sequences are potential attachment sites for glycosaminoglycan chains (Fig. 3), since they are located within consensus sequence contexts for glycosaminoglycan side chains in proteoglycan core proteins (12). If these sequences are indeed utilized for attachment of glycosaminoglycans, α1(XVIII) collagen would be an additional member of a growing group of known collagen–proteoglycans, including types IX and XII collagens (7, 9, 20–22). Also, α1(XV) collagen contains potential sites for N- and O-linked glycosylation (13).

α1(XVIII) Collagen mRNA Is Expressed in Several Internal Organs. Northern blot analysis (Clontech multiple tissue Northern blot) demonstrated that liver, lung, and kidney contained the highest levels of α1(XVIII) mRNA (Fig. 4). The mRNA migrated as two or three bands, depending on the tissue source. In testis, kidney, spleen, brain, and heart one transcript was about 4.5 kb; a second transcript was 5.5 kb.

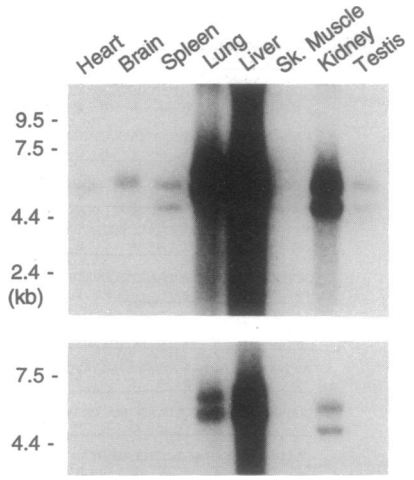

FIG. 4. Northern blot showing that α1(XVIII) transcripts are present in multiple organs. The probe (mc3b; Fig. 1) recognizes two bands (4.5 and 5.5 kb) in heart, brain, spleen, kidney, and testis. Note that the two bands in lung and liver have a lower mobility than the two bands in other tissues. The hybridized filter was exposed for a short period of time (1 hr) to demonstrate these differences more clearly (*Lower*). In brain RNA, a third band migrated slightly above the 5.5-kb band.

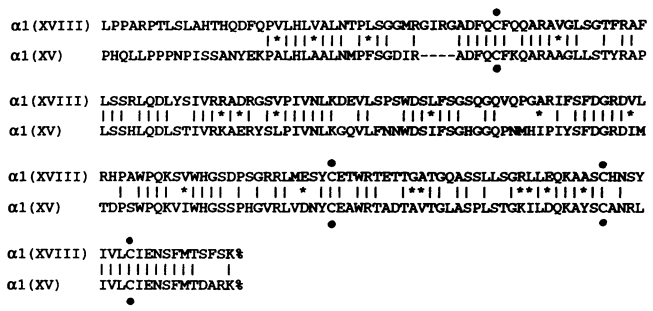

FIG. 5. Comparison of amino acid sequences within the carboxyl two-thirds of the carboxyl non-triple-helical (NC1) domain of mouse α1(XVIII) and human α1(XV) collagen chains. Identical residues are indicated by vertical lines, and similar residues are indicated by stars. To obtain the best alignment, a gap was introduced (dashes). Four cysteine residues are marked with dots. The termination of translation is indicated (%).

In brain a third band migrated above the 5.5-kb band. In humans, it is known that two different-size α1(XVIII) transcripts are generated due to utilization of alternative polyadenylylation signals (16). Since the pattern of two different-size transcripts in mouse tissues is similar to that in human tissues, it is possible that these two major bands are also produced by alternative polyadenylylation. Two major bands were also seen in liver and lung, but they had a slightly lower mobility than the two bands seen in the kidney.

α1(XVIII) and α1(XV) Collagens Belong to a Subfamily of Collagens. The domain organization of α1(XVIII) collagen is different from most other known collagen types. Comparison with a recently defined human collagen chain, α1(XV) (13), shows, however, that the lengths of the six most carboxyl-terminal triple-helical domains are almost identical in α1(XV) and α1(XVIII). In fact, they differ in size only by one amino acid triplet. In the 5′ direction, beyond the six domains, the two chains are quite different. Thus, the mouse α1(XVIII) chain contains four additional amino-terminal triple-helical domains of 27, 74, 122, and 83 aa, whereas human α1(XV) contains three domains of 18, 114, and 35 aa residues in the same region (Fig. 1).

The published α1(XV) sequence (13) does not extend to the translational stop codon, making a comparison between the NC1 domains of α1(XVIII) and α1(XV) not possible, based on the published sequence. Screening of a human placenta cDNA library, however, resulted in clones that extended into the 3′ untranslated region of α1(XV) (Fig. 1). A comparison of the amino acid sequences of α1(XV) and α1(XVIII) shows a remarkable similarity within the carboxyl-terminal non-triple-helical domain NC1. In this region, the most carboxyl-terminal 177 amino acid residues are about 60% identical [comparing mouse α1(XVIII) with human α1(XV)] at the amino acid level, with the locations of four cysteine residues conserved (Fig. 5).

Based on these similarities, we propose that α1(XV) and α1(XVIII) collagen chains are two members of a collagen subfamily, which we designate the multiplexin family (for collagens with multiple triple-helix domains and interruptions). A common and distinguishing feature of members of this family would be a highly conserved carboxyl-terminal non-triple-helical domain.

Note Added in Proof. Since the communication of this article, the complete sequence of the human α1(XV) collagen chain has been published (23).

We thank Ms. Masha Jakoulov for expert secretarial assistance. Work was supported by National Institutes of Health Grants AR36819, AR36820, and HL33014 (to B.R.O) and HL30648 (to S.T.).

1. Vuorio, E. & de Crombrugghe, B. (1990) *Annu. Rev. Biochem.* **59**, 837–872.
2. Jacenko, O., Olsen, B. R. & LuValle, P. (1991) *Crit. Rev. Eukaryotic Gene Exp.* **1**, 327–353.
3. van der Rest, M. & Garrone, R. (1991) *FASEB J.* **5**, 2814–2823.
4. Shaw, L. M. & Olsen, B. R. (1990) *Trends Biochem. Sci.* **16**, 191–194.
5. Yurchenko, P. D. & Schittny, J. C. (1990) *FASEB J.* **4**, 1577–1590.
6. Bachinger, H. P., Morris, N. P., Lunstrum, M. G., Keene, D. R., Rosenbaum, L. M., Compton, L. A. & Burgeson, R. E. (1990) *J. Biol. Chem.* **265**, 10095–10101.
7. Irwin, M. H. & Mayne, R. (1986) *J. Biol. Chem.* **261**, 16281–16283.
8. Dublet, B., Oh, S., Sugrue, S., Gordon, M. K., Gerecke, D. R., Olsen, B. R. & van der Rest, M. (1989) *J. Biol. Chem.* **264**, 13150–13156.
9. Watt, S. L., Lunstrum, G. P., McDonough, A. M., Keene, D. R., Burgeson, R. E. & Morris, N. P. (1992) *J. Biol. Chem.* **267**, 20093–20099.
10. Pan, T. C., Zhang, R. Z., Mattei, M.-G., Timpl, R. & Chu, M. L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6565–6569.
11. Li, K., Takai, K., Tan, E. M. L. & Uitto, J. (1993) *J. Biol. Chem.* **268**, 8825–8834.
12. Bourdon, M. A. (1990) in *Extracellular Matrix Genes*, eds. Sandell, L. J. & Boyd, C. E. (Academic, San Diego), pp. 157–174.
13. Myers, J. C., Kivirikko, S., Gordon, M. K., Dion, A. S. & Pihlajaniemi, T. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10144–10148.
14. Rehn, M. & Pihlajaniemi, T. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4234–4238.
15. Oh, S. P., Taylor, R. W., Gerecke, D. R., Rochelle, J. M., Seldin, M. & Olsen, B. R. (1992) *Genomics* **14**, 225–231.
16. Oh, S. P., Warman, M., Seldin, M., Cheng, S.-D., Knoll, J. H. M., Timmons, S. & Olsen, B. R. (1994) *Genomics* **19**, 494–499.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
19. Frohman, M. A. (1990) in *PCR Protocols: A Guide to Methods and Applications*, eds. Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J. (Academic, New York), pp. 28–38.
20. Bruckner, P., Vaughan, L. & Winterhalter, K. H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2608–2612.
21. McCormick, D., van der Rest, M., Goodship, J., Lozano, G., Ninomiya, Y. & Olsen, B. R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4044–4048.
22. Koch, M., Bernasconi, C. & Chiquet, M. (1992) *Eur. J. Biochem.* **207**, 847–856.
23. Muragaki, Y., Abe, N., Ninomiya, Y., Olsen, B. R. & Oshima, A. (1994) *J. Biol. Chem.* **269**, 4042–4046.