J. Biochem. **123**, 767-776 (1998)

Organization and Expression of Basement Membrane Collagen IV Genes and Their Roles in Human Disorders[1]

Yoshikazu Sado,* Megumi Kagawa,* Ichiro Naito,† Yasuyoshi Ueki,‡ Tsugio Seki,‡ Ryusuke Momota,‡ Toshitaka Oohashi,‡ and Yoshifumi Ninomiya‡,[2]

*Divisions of *Immunology and †Ultrastructural Biology, Shigei Medical Research Institute, 2117 Yamada, Okayama 701-02; and ‡Department of Molecular Biology and Biochemistry, Okayama University Medical School, 2-5-1 Shikata-cho, Okayama 700-8558*

Received for publication, January 22, 1998

Six distinct genes have been identified as belonging to the type IV collagen gene family. They can be organized into three sets, *i.e.*, COL4A1/COL4A2, COL4A3/COL4A4, and COL4A5/COL4A6, which are localized on three different chromosomes in humans, 13, 2, and X, respectively. Within each set the genes are aligned head-to-head and their expression is regulated by bidirectional promoters between the genes. Transcriptional regulation of the COL4A1/COL4A2 set has been well characterized. The transcription of COL4A6 seems to be controlled by two alternative promoters. While collagen IV molecules composed of α1 and α2 chains are broadly distributed, molecules comprising combinations of the other four chains, α3-α6, are important components of specialized basement membranes. The precise chain composition of triple-helical molecules assembled from the α3-α6 chains is not entirely clear, but it is hypothesized that α3-α5 chains and α5 and α6 chains form heterotrimeric molecules. Several pieces of evidence indicate that α3/α4/α5 molecules and α5/α6 molecules are components of the basement membrane network. This helps explain the observation that the kidney and skin basement membranes from patients with Alport syndrome caused by mutations in the α5 coding gene, COL4A5, are defective in the α3, α4, and α6 chains together with the α5 chain. Large deletions involving the COL4A5 and COL4A6 genes have been found in rare cases of diffuse leiomyomatosis associated with Alport syndrome.

Key words: Alport syndrome, basement membrane, bidirectional promoter, collagen IV, diffuse leiomyomatosis associated with Alport syndrome.

Basement membrane (BM)s are thin and amorphous specialized extracellular matrices that play roles in diverse biological events, including embryonic development, maintenance of tissue architecture, protection of tissues and organs from mechanical stress and exogenous factors, tissue remodeling during development and wound healing, and filtration of blood and air. The BM located underneath epithelial cells, *i.e.* the epidermal BM, usually faces a cell membrane on one surface and the extracellular matrix on the other, however, there are several types of cells in which the BMs face cell layers on both surfaces, *e.g.*, smooth muscle, skeletal muscle, alveoli, capillaries in the brain, and neuromuscular junctions. BMs are composed of collagen IV, laminin, nidogen, heparan sulphate proteoglycan, and other glycoproteins (*1*). Recent progress in research on basement membrane biology has revealed new BM components including laminins, collagens XV, XVIII, and XIX, fibulins, and dystroglycan (*1, 2*). Interactions between cells and BM molecules, and between BM molecules and BM molecules are being investigated, but the biological functions of BMs in general have yet to be determined. Furthermore, investigations involving gene targeting technology and fine analyses of genetic disorders, including Alport syndrome and others, have allowed us to take a comprehensive approach for clarifying the biological roles of BMs. Because of the huge number of papers published on the molecular, biological and pathological aspects of BMs, our present review will be selective and focused on the following: the organization and expression of collagen IV genes, collagen IV molecules and chain compositions, the evolution of collagen IV genes, and what we have learned from mutation analysis of cases of Alport syndrome (AS) and diffuse leiomyomatosis (DL) associated with AS.

Chromosomal location of collagen IV genes

Since the first two human collagen IV genes found, COL4A1 and COL4A2, were assigned to the same chromosomal locus, 13q34 (*3, 4*), it was hypothesized that addi-

[1] This work was supported by Grants-in-Aid for Scientific Research (09671309 to T.O. and 08457154 to Y.N.) from the Ministry of Education, Science, Sports and Culture of Japan and by research grants from the Terumo Life Science Foundation, the Ryobi Teien Foundation, and the Yamanouchi Foundation for Research on Metabolic Disorders.

[2] To whom correspondence should be addressed. Phone: +81-86-235-7127, Fax: +81-86-222-7768, E-mail: yoshinin@cc.okayama-u.ac.jp

Abbreviations: BM, basement membrane; AS, Alport syndrome; DS, diffuse leiomyomatosis.

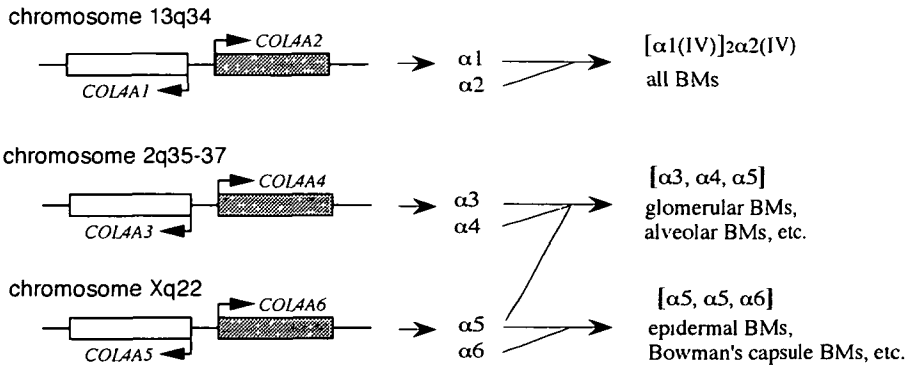

Fig. 1. **Organization of collagen IV genes.** The six human collagen IV genes are located in pairs in a head-to-head manner on three different chromosomes. These six genes can be classified into two groups, Group A (depicted by white boxes): COL4A1, COL4A3, and COL4A5 encoding the $\alpha 1$, $\alpha 3$, and $\alpha 5$ chains, respectively, and group B (shaded boxed): COL4A2, COL4A4, and COL4A6 encoding the $\alpha 2$, $\alpha 4$, and $\alpha 6$ chains, respectively. These are based on the gene structure, amino acid sequence homology, and domain structure of the translation products. Three molecules consisting of $\alpha 1/\alpha 2$ chains, $\alpha 3/\alpha 4/\alpha 5$ chains, and $\alpha 5/\alpha 6$ chains are predicted to be present in different BMs.

tional genes belonging to the same subclass would also be located in the same region, like in other multigene families of globin and histone genes. But the other four genes turned out to be on different chromosomes (*5-8*). COL4A3 and COL4A4 were located on 2q35-q37 (*5, 6*) and COL4A5 and COL4A6 on Xq22 (*7-9*). Namely, the three paired genes were found to reside at the three different chromosomal regions, as shown in Fig. 1. It is intriguing that two sets of the paired genes, COL4A1/COL4A2 and COL4A5/COL4A6, are located head-to-head on chromosome 13 and X, respectively, sharing common bidirectional promoters (*8, 10-14*). The structural relationship between the COL4A3 and COL4A4 genes is less characterized, although they are likely to be head-to-head on chromosome 2 as in the case of the other two pairs. Three more collagen genes, COL6A3, COL3A1, and COL5A2, are known to be located on chromosome 2 (*15, 16*). It is not known yet how close the two collagen IV genes are located to these three genes.

The mouse and dog (*17*) collagen IV genes are known to be arranged like the human collagen IV genes. It is interesting that in the nematode, *Caenorhabditis elegans*, there are two collagen IV genes; one is most similar to the mammalian $\alpha 2$(IV) chain and the other is rather similar to the $\alpha 1$(IV) chain. However, the genes are located on separate chromosomes, X and III (*18*), suggesting that the primordial collagen IV gene became duplicated prior to the separation of vertebrates and invertebrates. However, the two collagen IV genes in *Drosophila* appear to be located on the same chromosome (*19*).

Differential expression of collagen IV genes

The complete exon/intron structure of a collagen IV gene was first characterized for the COL4A1 gene, which contains 52 exons (*20*). The last five exons at the 3′ end encode the NC1 domain, which is located at the carboxyl end of the $\alpha 1$(IV) chain. The central 39 exons encode the COL domain, which contains the typical collagenous Gly-X-Y repeated structure with many breaks. The first intron of COL4A1 is large (>18 kb). Repetitive sequences are present at numerous sites in the intervening sequences, some of which could regulate the expression of the gene. The structural organization of the human COL4A2 gene is similar to that of COL4A1; but instead of the long first intron, it has a long intron 3 (*21*). Thus, the promoter and about 100 bp of the coding sequence in both genes are flanked by the two long intron sequences of >20 kb.

Although the COL4A3 and COL4A4 genes have not been characterized completely (*22, 23*), the two genes, COL4A5 and COL4A6, on chromosome X have been characterized extensively (*24, 25*). One of the reasons for this is that mutations in COL4A5 cause Alport syndrome. The gene structure is very similar in COL4A1 and COL4A5. In both genes, the first intron is very long and the second exon starts with the last residue of the signal peptide (*24*). The COL4A6 gene was found in a head-to-head arrangement within 500 bp from the COL4A5 gene on chromosome X (*14*). An interesting finding is that there are two alternative promoters in the COL4A6 gene and that they are used in a tissue-specific manner.

In order to determine whether the translation products of the six genes are co-localized in various tissues, we raised $\alpha 1$-$\alpha 6$ chain-specific rat monoclonal antibodies against synthetic peptides reflecting sequences near the carboxy terminus of each NC1 domain (*26, 27*). In the kidney, all BMs in glomeruli and tubules were stained positively by the antibodies for the $\alpha 1$ and $\alpha 2$ chains, whereas the $\alpha 3$ and $\alpha 4$ chains were confined to the glomerular BMs and some parts of the tubules. Thus, the $\alpha 3$ and $\alpha 4$ staining pattern is rather limited as compared to that of the $\alpha 1$ and $\alpha 2$ chains, but their common location suggests that they are synthesized by the same cells. In stark contrast, the expression pattern of the $\alpha 5$ and $\alpha 6$ chains is different in the glomerulus: the $\alpha 6$ chain is never detected in the glomerular BMs, but is positive in the BMs of the Bowman's capsules and some distal tubules in humans, whereas the $\alpha 5$ chain is densely positive in the glomerular BMs and its staining pattern is similar to that obtained with the $\alpha 3$ and $\alpha 4$ antibodies. These results suggested that there are two kinds of collagen IV molecules, one with $\alpha 3$, $\alpha 4$, and $\alpha 5$ chains, and the other with $\alpha 5$ and $\alpha 6$ chains (*27*).

In skin tissues, the staining pattern for the $\alpha 1$ and $\alpha 2$ chains was almost the same: positive in all epidermal BMs and capillary BMs. However, little if any reaction was seen in basement membranes with the $\alpha 3$ and $\alpha 4$ antibodies. The $\alpha 5$ and $\alpha 6$ chains were co-localized only in the epidermal BMs. The temporal and spatial expression of the $\alpha 1$, $\alpha 2$, $\alpha 5$, and $\alpha 6$ chains was demonstrated during the formation of the basal lamina in an "*in vitro*" skin model (*28*). Loss of $\alpha 5$ and $\alpha 6$ chains has been demonstrated in basal cell carcinomas (*29*), invasive prostate carcinomas (*30*), and DL associated with AS (*31, 32*). Thus, the COL4A3/COL4A4 and COL4A5/COL4A6 genes are expressed in a tissue-specific manner, probably due to the unique function of the bidirectional promoters of these genes.

Regulation of collagen IV gene expression

The transcription start sites for COL4A1 and COL4A2 are separated only by 130 bp (*10-13*). The genes are transcribed from the opposite DNA strands starting from the short intergenic space that harbors the common bidirectional promoters. The incomplete palindromic symmetry with the Sp1 binding site in the center and several inverted repeats have been recognized in the region. Several novel cis elements including a CTC box within the short space bind to some transcription factors (*33, 34*). Distal regulatory elements including enhancers located in the first intron or in close vicinity to it, and silencers have been identified (*35, 36*).

The COL4A5 and COL4A6 genes are aligned head-to-head on chromosome X in a similar manner to as found for the COL4A1 and COL4A2 genes (*8, 14*). Interestingly, the COL4A6 gene is transcribed from two alternate promoters in a tissue-specific manner (*14*). One transcription start site is 442 bp away from the transcription start site of COL4A5, whereas the alternative transcription start site is located 1,050 bp apart from the first one and drives the expression of a second transcript that encodes an α6(IV) chain with a different signal peptide. The presence of a third cryptic promoter was hypothesized to exist in the second intron of COL4A6 (*37*), which was estimated to be as long as 180 kb (*38*).

Collagen IV molecules and their α chain composition

It is well known that a molecule with two α1 chains and one α2 chain exists in various BMs. A huge amount of biochemical, immunohistological and molecular biological data have been accumulated supporting the presence of this molecular form: (i) Collagen IV molecules purified from placenta were composed of α1 and α2 chains (*39*). (ii) Immunohistochemical analysis involving α-chain specific monoclonal antibodies demonstrated that the staining pattern for α1 was exactly the same as that for α2 in serial sections of kidney, skin, and placenta (*14, 40, 41*). (iii) Transfection experiments with α1 and α2 cDNAs introduced into CHO cells showed that the secreted collagen molecule was composed of α1 and α2 chains in a ratio of 2:1 (*42*).

As mentioned above, collagen IV molecules consisting of α3, α4, and α5 chains are distributed in glomerular BMs. This was also found in alveolar BMs and some parts of tubular BMs. In these BMs, the three α chains were co-distributed (*14, 40, 41*). In male cases of X-linked Alport syndrome, these three α chains could not be detected in kidney BMs or were greatly decreased in amount (*43, 44*). Although a mosaic pattern was observed in female cases of X-linked Alport, in BMs negative for the α5 chain, the staining for both α3 and α4 chains was also negative, suggesting that the three α chains are co-distributed. In some autosomal recessive-type Alport cases, staining of all three chains was negative (*45*). Furthermore, two autosomal Alport mouse models lacking α3(IV) were recently established, and immunohistochemical analysis showed the absence of α4 and α5 chains as well as of α3 chains in their glomerular BM (*46, 47*).

The co-distribution of α5 and α6 chains was detected in BMs of the Bowman's capsule and some tubules in kidney, and in epidermal BMs (*14, 40, 41*). The presence of a heterotrimeric molecule composed of α5, α6 and an unknown third chain was also supported by the following evidence: In typical male Alport cases in which mutations in COL4A5 cause a lack of the α5 chain, the α6 chain is also missing in dermal BMs. However, epidermal BMs of female Alport cases showed a mosaic pattern; *i.e.* in BMs in which α5 was present, α6 existed, whereas in BMs in which α5 was missing, α6 was missing as well (*40*), which strongly supports the presence of the molecule. The third chain cannot be α3 or α4, since neither chain is distributed in epidermal BMs. Furthermore, in autosomal recessive Alport cases there are α5 and α6 chains in dermal BMs just like in normal skin (Naito and Sado, unpublished). Another piece of evidence is that the distribution of α5 and α6 chains seems to be quite normal in COL4A3-knockout

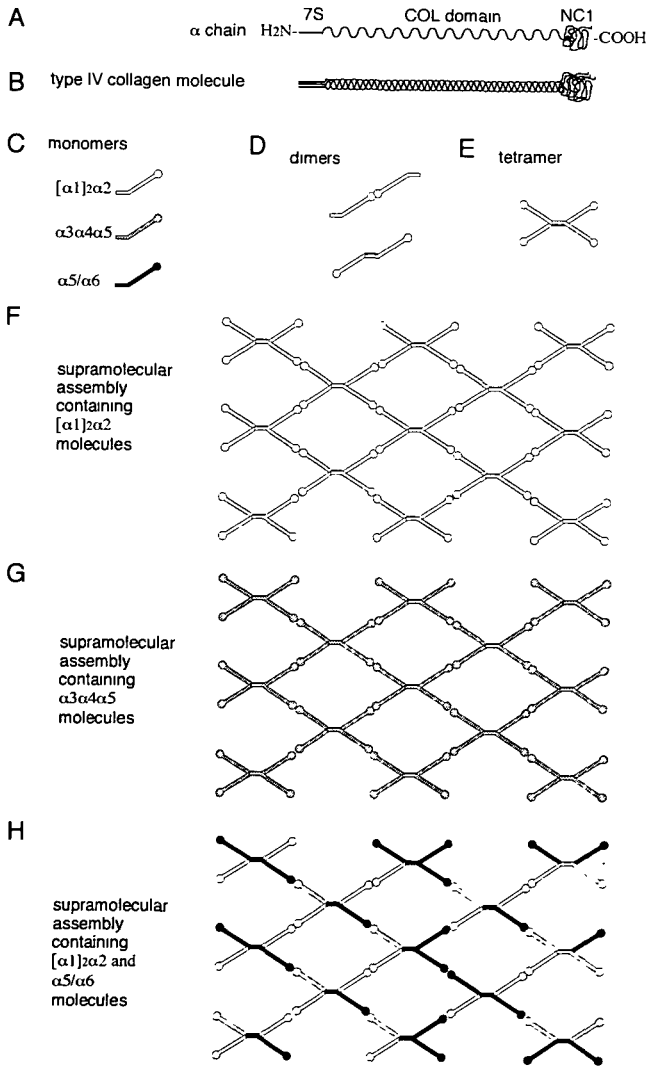

Fig. 2. **Schematic representation of the collagen IV α chain (A) and a collagen IV molecule (B) consisting of three α chains.** Three monomers, white, gray, and black (in C), consisting of α1/α2 chains, α3/α4/α5 chains, and α5/α6 chains, respectively, form dimers (D) through interactions between the amino terminal 7S domains or the carboxy termini, and tetramers (E) through the combination of the two dimers at the 7S domains. Thus, the supramolecular network (F, G, and H) is formed through the assembly of dimers and further aggregates. Some aggregates consist of only α1/α2 (F) or α3/α4/α5 molecules (G), but others of α1/α2 and α5/α6 together in the same network (H).

```
                                      A                    B      C                                 D
α1: KGGCAGSGCGKCDCH          (120)LGHVPGMLLK   (51)FQ(21)QVQEKGDFATK                     (15)G
α2: KFDVPCGGRDCSGGCQCYPEK    (114)YALPKEERDRYR(51)FY(18)SDTLHPIIAPTGVTFHPDQYK(15)L
α3: KGCVCKDKGQCFCD           (117)AKEEDIELDAK (51)VI(21)IVTLTGPDNRTDLK                  (24)Y
α4: SGKKYIGPCGGRDCSVCHCVPEK(114)VFILGAVK     (51)GV(21)LLVEPPDFCLYK                    (24)G
α5: AACYGCSPGSKCDCS          (120)IMSSLP       (54)FQ(21)SEQKRPIDVEFQK                   (24)G
α6: SYGKPCGGQDCSGSCQCFPEK    (114)VLAPGSFK     (54)FQ(18)PSTGELEFMGFPK                   (24)G

           E                 F             G      H                        I               J
α1: V    (60)IGT          (39)VP        (24)LP  (24)NGIVECQP          (30)CLICDID     (60)Y
α2: K    (57)AYSPHPSLAK  (33)IGDGDQRR(18)ALYG(27)F                  (33)RCTEGDEAIK(57)R
α3: GSEK(57)EYYDTYQEK   (33)RP        (24)AM  (30)K                 (30)CTQCPYIP    (60)L
α4: I    (63)VTPPLPLK    (33)CA        (27)DSAP(24)VIYCSV            (30)ACEP        (63)V
α5:      (63)IPRPGTGITI  (36)VM        (24)IP  (24)IPPSDEICEP        (27)CFNCIGT     (63)L
α6: L    (66)VFIDIDGAVIS(33)ALS       (24)TI  (21)SPEFETETLHNKES(27)ACDGGVPNT  (57)I

            K    K1          K2     L                                  M   M1         M2      M3
α1: FDLRLK[66](66)               GY(33)VPLP                       [92](42)GI      (48)
α2: TITTK [65](36)GD      (27)GL(36)DCDTDVKRAVGGDRQEAIQPGCIA[93](93)
α3: QPE   [69](39)ALS     (27)GY(30)SVSTPVP                    [89](42)GI      (45)
α4: VSRVK [65](39)EDATP(21)GL(33)ISCNVTYP                   [89](42)EIP     (21)PV(21)
α5: TFP   [69](42)TFK     (24)GF(33)ITQP                       [92](42)GI      (48)
α6: LSTIQ [65](36)GD      (27)GN(33)LPCIIP                     [93](45)HLPELP(42)

            N      N1        N2       N3                     O     O1    O2           P        Q
α1: LP    [199]( 60)GI ( 36)MP (99)DKVDMGSMK      [108](108)      GI    (117)FP    [54]
α2: LGAQP[198](198)                   PVILPGMKDIK [106]( 42)K(63)INLP (114)IH    [55]
α3: AVAMP[196]( 60)IE ( 36)RS (96)EISHVI         [108](108)      EGTRP(111)RAPP[54]
α4: AF    [201]( 63)CER(135)          AIISQK         [106]( 42)R(63)ASHF (114)DV    [53]
α5: LP    [199]( 60)GI ( 36)VP (99)DPNLL          [111](111)      SI    (117)NP    [54]
α6: FGAEN[198]( 99)VKK( 30)ALS(63)SPRRPMSNLWLK[106]( 42)K(63)VEIS (114)IT    [55]

     Q1          Q2               R       R1          R2                S    S1          S2
α1: (54)             TE         [120](120)             DIIK          [84](84)                 SVD(NC1)
α2: (18)NTLP(33)PPSNIS [117]( 27)ALP  (87)QKIAIQP      [84](45)AVP(36)      VSI(NC1)
α3: (54)             IP         [116]( 27)VI   (87)KIISLP       [81](81)                 ATW(NC1)
α4: (24)PP     (27)PKDIPDP[112]( 27)DLLR(81)ADVDDCPRIP[80](24)GP  (54)PGYLGG(NC1)
α5: (54)             LE         [120](120)             IIIK          [84](84)                 SVA(NC1)
α6: (21)I      (33)LPSLIA [114]( 30)SNQ  (81)PTAEAVQVPP[84](84)                 MRV(NC1)
```

Fig. 3. **Comparison of the lengths of G-X-Y repeats and amino acid sequences of the interruptions of the six human α(IV) chains.** The G-X-Y repeat structure is interrupted by many non-G-X-Y structures (interruptions). G-X-Y repeats (represented only by the numbers of amino acid residues in parentheses) are named A to S from the amino terminus. When the numbers of residues did not correspond with each other among the six α chains in some G-X-Y repeats, they are presented in brackets as the numbers including those in interruptions. For instance, [65] in α2-K represents the total number of residues of 36 (in K1) + 2(G and D) + 27 (in K2). Note that the residue numbers in the G-X-Y repeats in α1 are pretty similar to those in the α2 chain, but quite different from those in α3–α6. One good example is D; thus, the numbers of 15 and 15 of the α1 and α2 chains, respectively, correspond exactly. Similarly, those of 24, 24, 24, and 24 of α3–α6 chains, respectively, correspond as well, however, the number of Gly-X-Y repeats is different from that for the α1 and α2 chains.

mouse skin (*46, 47*). A supramolecular assembly containing α5/α6 molecules and α1/α2 molecules form the basement membrane architecture, as shown in Fig. 2. This is still hypothetical, and more extensive investigations involving pseudolysine will provide some insight into the detailed assembly of the BMs, which comprises complex networks composed of α1–α6 chains and α3–α6 chains in addition to the classical network of α1 and α2 chains (*48*).

What we learn from sequence comparison among the six α(IV) chains

The entire amino acid sequences of the α1(IV) and α2(IV) chains were determined for man (*49–52*) and mouse (*53, 54*). The total length of the α1 polypeptide for man and mouse is the same, and the G-X-Y repeated structure, and positions and lengths of interruptions of G-X-Y are identical. The mouse α2 chain is 5 amino acid residues shorter than the human α2 one, which could be due to 2 residues between C and D regions and 3 residues between I and J. The positions and lengths of interruptions of G-X-Y are identical as well.

When the locations and lengths of G-X-Y repeats and interruptions of the human α1 and α2 chains are compared, as shown in Fig. 3, the difference in G-X-Y repeats was noted to be within 6 residues. At three places, A, F, and G, where 6 residues are different, the length might be adjusted

by means of the neighbouring interruptions. Namely,

(i) In A, the $\alpha 1$ chain is longer than $\alpha 2$ by 6 residues. Since A continues to the NC domain at the amino terminus, the difference should be considered as the length to the first cysteinyl residue. There is only one residue between this cysteine and A in $\alpha 1$ but 4 residues in $\alpha 2$.

(ii) The $\alpha 1$ chain is longer than $\alpha 2$ by 6 residues in F. Here, if one includes interruptions before and after, the length for $\alpha 2$ would be longer than that for $\alpha 1$ by 13 residues.

(iii) The $\alpha 1$ chain is longer than $\alpha 2$ by 6 residues in G, too; but if one includes neighbouring interruptions, the length for $\alpha 2$ would be longer than that for $\alpha 1$ by 8 residues.

The interruption following L in the $\alpha 2$ chain is 20 residues longer than that in $\alpha 1$. Since it contains 2 cysteinyl residues, we postulate that it would stick out and form a loop. Furthermore, there is a large difference in the interruption following C. It might be unimportant for triple-helix formation because it is different in the mouse.

The amino acid sequences of $\alpha 3$ and $\alpha 4$ were determined last for human collagen IV (*55, 56*). The $\alpha 5$ gene was first discovered as a gene responsible for X-linked Alport syndrome, and its entire sequence is known only for man (*57*). The partial sequence covering the entire NC1 domain and 532 residues of the COL domain was reported for the dog $\alpha 5$ (*58*). When the G-X-Y repeat structure and interruptions were compared between man and dog within this region, they were found to be identical.

The locations and lengths of the G-X-Y repeat structure and interruptions were compared among the $\alpha 3$, $\alpha 4$, and $\alpha 5$ chains, as shown in Fig. 3. It should be noted that when the lengths of the individual G-X-Y repeats are compared, the maximum difference among the three chains is 6–8 residues, as found on comparison of the $\alpha 1$ and $\alpha 2$ chains which form the collagen IV molecule. On more precise analysis,

(i) In A, $\alpha 5$ is longer than $\alpha 4$ by 6 residues. If the length is taken to include the residues to the first cysteinyl residue in the amino terminal NC domain, the difference would be smaller.

(ii) The $\alpha 4$ and $\alpha 5$ chains are 6 residues longer than $\alpha 3$ in E. This difference would be adjusted if one includes residues in the interruption in front of E.

(iii) In H, $\alpha 3$ is longer than $\alpha 4$ and $\alpha 5$ by 6 residues. The total lengths of the interruptions before and after H for $\alpha 4$ and $\alpha 5$ are longer than those for $\alpha 3$ by 7 and 9 residues, respectively.

(iv) In K2, $\alpha 3$ is longer than $\alpha 4$. The interruption in front of K2 in $\alpha 4$ is 2 residues longer than that in front of $\alpha 3$.

(v) In P, $\alpha 5$ is longer than $\alpha 3$ by 6 residues. Neighbouring interruptions in $\alpha 3$ represent 5 more residues than in $\alpha 5$.

(vi) In R, $\alpha 5$ is longer than $\alpha 4$ by 8 residues; however, the interruption in front of $\alpha 4$ is 5 residues longer than that in front of $\alpha 5$.

(vii) In R2, $\alpha 3$ is longer than $\alpha 4$, but neighbouring interruptions adjust the difference.

Vast differences in the corresponding interruptions among the three α chains were not observed, but small differences were noted in those after H, *i.e.* those of $\alpha 4$ and $\alpha 5$ were shorter than those of $\alpha 3$. Since there are cysteinyl residues in the interruptions in $\alpha 4$ and $\alpha 5$, a disulfide bridge might be formed between the two chains.

Can $\alpha 5$(IV) and $\alpha 6$(IV) form a molecule?

When the amino acid sequence of the human $\alpha 6$ chain (*9*) is compared with that of the $\alpha 5$ chain (*57*), striking similarity is noted, as shown in Fig. 3. The maximum difference in the lengths of G-X-Y repeats between $\alpha 5$ and $\alpha 6$ was found to be 6 residues, which is the same as that between $\alpha 1$ and $\alpha 2$. The following 5 G-X-Y repeats showed the maximum difference, which could be adjusted by interruptions to form a molecule:

(i) In A, the $\alpha 5$ chain is longer than $\alpha 6$ by 6 residues. Since this G-X-Y repeat is located closest to the amino terminus, the G-X-Y repeat should include residues to cysteines that are located in the amino terminal NC domains. Thus, the difference would be less.

(ii) In J, the $\alpha 5$ chain is again longer than $\alpha 6$ by 6 residues, but if one includes the interruptions before and after J, the $\alpha 6$ chain here would be longer than the $\alpha 5$ chain by 4 residues.

(iii) In K1, the $\alpha 5$ chain is again longer than $\alpha 6$ by 6 residues. The interruption in front of K1 of $\alpha 6$ is longer than that of $\alpha 5$ by 2 residues. There is thus only a 4-residue difference in the total residues in K here.

(iv) Although the $\alpha 5$ chain is longer than $\alpha 6$ by 6 residues in M2, neighboring interruptions in $\alpha 6$ add up to 11 residues, which is 7 residues longer than in $\alpha 5$. Therefore, the difference would be only one residue.

(v) In R, the $\alpha 5$ chain is longer than $\alpha 6$ by 6 residues, since the $\alpha 6$-interruption in front of R is 4 residues longer than that in front of $\alpha 5$.

Most of the corresponding interruptions in the $\alpha 5$ and $\alpha 6$ chains are strikingly similar to each other. Thus, as judged on comparison of the amino acid sequences, the lengths of G-X-Y repeats and interruptions of both the $\alpha 5$ and $\alpha 6$ chains are favorable enough to allow the formation of a triple-helical molecule between the two chains.

The next question would be if the third chain is different from the $\alpha 5$ or $\alpha 6$ chain. No experimental evidence is available to answer this question at present. It is clear that the $\alpha 3$ and $\alpha 4$ chains cannot be candidates since the $\alpha 5$ and $\alpha 6$ chains are distributed in epidermal BMs, where the $\alpha 3$ and $\alpha 4$ chains are not be found at all. As discussed below, the six mammalian collagen IV genes are assumed to have evolved from a common ancestral gene that was duplicated and inverted so that the genes became arranged head-to-head. Later the genomic region containing the two genes was then duplicated again, which resulted in the three pairs of collagen IV genes that we are able to recognize today in vertebrates. The first duplication and inversion event created the two groups of genes: groups A (odd numbers) and B (even numbers). Thus, group A duplicated into the COL4A1, COL4A3, and COL4A5 genes, and group B into the COL4A2, COL4A4, and COL4A6 genes. Since evidence for the two groups of genes has been obtained in *Caenorhabditis elegans* (*18*), and, furthermore, since the two chains form a trimeric collagen molecule, the first duplication event would have happened before the vertebrates evolved. Furthermore, it is reasonable to think that the rule that the collagen IV triple-helix is formed by the two groups of proteins, groups A and B, was kept during evolution. The

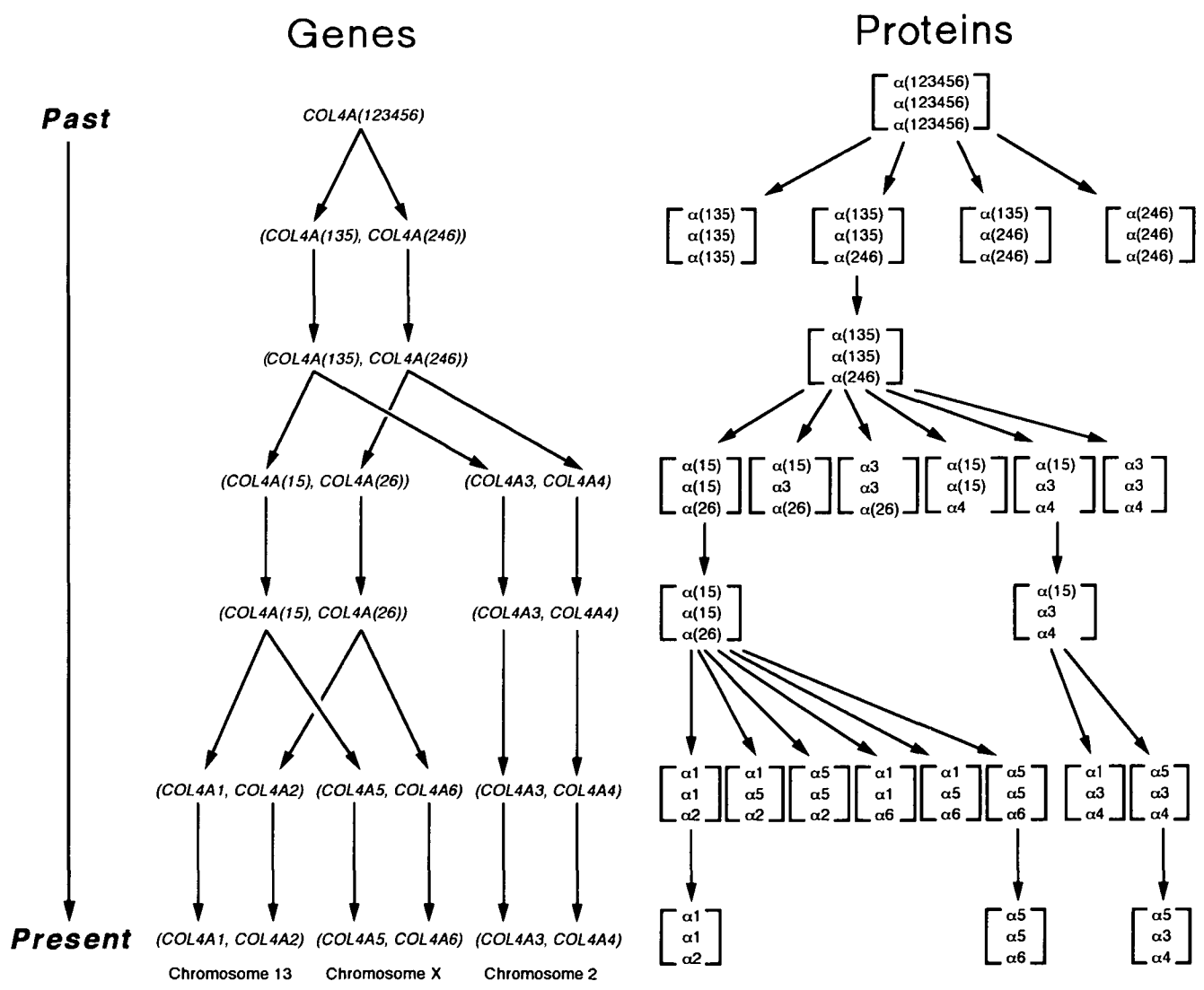

Fig. 4. **Transition of collagen IV chains during gene duplications.** COL4A(1,2,3,4,5,6) and α(1,2,3,4,5,6) represent a common ancestral collagen IV gene for the α1-α6 chains and its translation product, respectively. Similarly, COL4A(1,3,5) and α(1,3,5) represent a common ancestral collagen IV gene for the α1, α3, and α5 chains, and its translation product, respectively. The combinations of α chains of triple-helical molecules are shown in brackets. As gene duplication occurred, new collagen molecules of different α chain combinations appeared. As time elapsed a few molecules with certain α chain combinations became predominant and survived; thus we speculate that collagen molecules are composed of two odd numbered-chains and one even numbered-chain. Three different molecules, [α1, α1,α2], [α3,α4,α5], and [αX,α5,α6], are likely to be present in tissues, where we presume that X is 5 based upon several lines of evidence including that obtained on immunohistochemistry using α chain-specific antibodies, and amino acid sequence comparison between G-X-Y repeats and interruptions among α chains.

presence of the two molecular forms of [α1,α1,α2] and [α3,α4,α5] supports this idea. Thus, the two group A proteins and one group B protein form a collagen IV molecule. If this is true for the α5/α6-containing molecule, the third chain would be restricted to the α1 or α5 chain.

Supposing that the third chain is α1, the maximum difference in the length of G-X-Y repeats would be 9 residues in D (Fig. 3) between the α1 chain and the α5 or α6 chain. If we consider the results of the sequence comparison that we discussed above, this figure is too large for the α chains to form a collagen molecule. Furthermore, even if we consider the length of the neighboring interruptions flanking D, it cannot be adjusted to align α1 with α5 and α6 to form a molecule since the difference would be larger. Therefore, we do not think that α1 participates in the formation of the α5/α6-containing molecule. However, the α5 chain can easily find a position in the molecule. We need further experimental evidence to support this idea, since this conclusion is based on hypothetical evaluation.

Evolution of collagen IV genes

Since the cDNAs for the six α(IV) chains have been reported, it is possible to predict from the nucleotide and derived amino acid sequences how the six genes have evolved. Considering the α chain composition of individual collagen IV molecules together with the prediction, we are able to follow the route by which collagen polypeptides have originated, developed, and been improved.

Ideally the ancestral collagen α(IV) chain was able to form a homotrimeric collagen molecule (Fig. 4). The

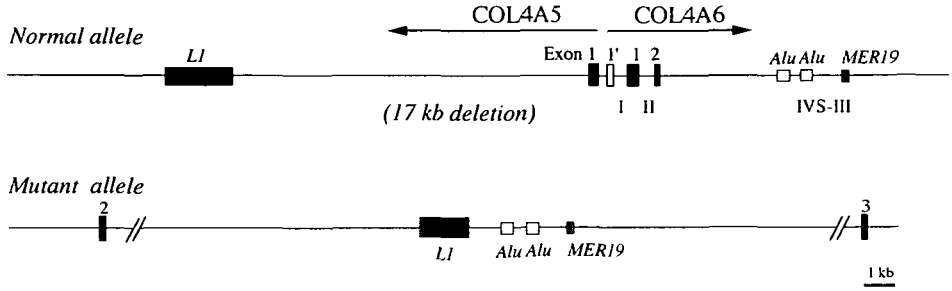

Fig. 5. **A large deletion spanning the 5′ region of the COL4A5/COL4A6 cluster recognized in a case of diffuse leiomyomatosis associated with Alport syndrome.** As indicated, the two genes, *COL4A5* and *COL4A6*, are arranged head-to-head. Exons 1 and 1′ of the *COL4A6* gene are used alternatively. The location of the 17-kb deletion is indicated on both the normal and mutant alleles. For the sake of simplicity, only *COL4A6* introns are indicated, as intervening sequences I, II, and III. Black and white boxes indicate exons with the numbers shown above. Several characteristic elements, *i.e.*, *L1*, *Alu*, and *MER19*, are shown by various boxes (*32*).

common primordial gene coding for the α(IV) chain was duplicated, and the duplicates were thus in neighboring positions. As time passed, the nucleotide sequences of these paired genes gradually became altered; one (group A) gene [COL4A(1,3,5)], coding for odd-numbered α chains, and another (group B) gene [COL4A(2,4,6)], coding for even-numbered α chains, becoming established. Collagen IV triple-helix molecules were thus formed from the two α chains [α(1,3,5)] encoded by the group A genes and the one α chain [α(2,4,6)] encoded by the group B gene. The rule would be strictly kept even after the gene duplication event that occurred later.

These paired genes duplicated again and diverged into the common ancestral gene [COL4A(1,5)] for the α1 and α5 chains, the gene (COL4A3) for the α3 chain, the common ancestral gene [COL4A(2,6)] for the α2 and α6 chains and the gene (COL4A4) for the α4 chain. Although at this time the possible combinations of the α chains was 6, only the two combinations of [α(1,5),α(1,5),α(2,6)] and [α(1,5),α3,α4] successfully survived to form triple-helices. At this time ideally it was possible for the combination of [α3,α3,α4] to survive but practically it did not. Similarly, the common ancestral COL4A(1,5) and COL4A(2,6) genes duplicated one more time. As time elapsed, the two ancestral genes diverged into the COL4A1 and COL4A5 genes, and COL4A2 and COL4A6 genes, respectively. At this time the possible combinations of [α(1,5), α(1,5),α(2,6)] was 6, but only two combinations, [α1,α1, α2] and [α5,α5,α6], remained. On the other hand, from the [α(1,5),α3,α4] combination there were two possibilities, [α1,α3,α4] and [α5,α3,α4]; but only the combination of [α3,α4,α5] was selected. Consequently, it is now possible to understand how the collagen IV genes and collagen triple-helices evolved. We believe that the combination of [α3, α4, and α5] was not an accidental but an inevitable product during evolution.

It is reasonable to consider that the combination of [α3, α4, and α5] is an indispensable molecule for the glomerular BMs in kidney since progressive glomerulonephritis occurs when it disappears from glomerular BMs. Supramolecular aggregates containing [α1,α1,α2] molecules maintain their function as basement membrane collagen in the basal laminae which are widespread throughout the animal kingdom; whereas the biological role of [α5,α5,α6] molecules is not known yet.

Mutations in the COL4A3, COL4A4, and COL4A5 genes cause Alport syndrome

Ultrastructural changes in glomerular BMs started being recognized in cases of hereditary Alport syndrome in the late 60s (*59*). These observations suggested that this disease might be caused by a mutation in some gene(s) coding for protein(s) functioning in glomerular BMs. Soon thereafter in 1971, Kefalides found a new collagen type IV in glomerular BMs (*60*), and Spear proposed that Alport syndrome might be caused by mutations in this gene (*61*). Cloning of cDNAs for the most abundant human α chains, α1 and α2, and chromosomal assignment of these genes (*62, 63*) did not reveal the pathogenesis of the disease. Obviously, the genes were on chromosome 13 instead of chromosome X. Attention was then focused on new collagen chains α3 and α4, the genes of which were found to be on chromosome 2 several years later. In 1988, linkage analysis assigned the Alport syndrome locus to Xq22-26 (*64*). Hostikka *et al.* reported the isolation of the cDNA for a novel α5(IV) chain and assignment of the gene on Xq22 in 1990 (*7*). Soon after that report, mutations in Alport cases were identified in the COL4A5 gene (*65*). To date, more than 200 mutations have been reported in the COL4A5 gene in X-linked Alport patients. These mutations include large and small rearrangements such as deletions, insertions, inversions and duplications, as well as single base changes. Strikingly, the mutations are highly dispersed throughout this huge gene (~250 kb), which contains 51 exons. The same mutations have not been found in unrelated families, except in a few cases. The highly dispersed Alport mutations in the COL4A5 gene remind us of the similar situation in the case of mutations causing osteogenesis imperfecta, a brittle bone disease characterized by a variety of mutations spread out in two type I collagen genes, COL1A1 and COL1A2 (*66*).

Autosomal forms of Alport syndrome have been postulated based on the results of linkage analyses. It was reasonable to search for mutations in the COL4A3 and COL4A4 genes since the translation products of both genes were found in glomerular BMs. Homozygous and heterozygous mutations were found in the COL4A3 and COL4A4 genes in four different autosomal Alport kindreds (*67, 68*). A splice-mediated insertion of an Alu sequence in the COL4A3 mRNA that might cause autosomal recessive Alport syndrome has been found as well (*69*).

At this time no information is available suggesting that mutations in the COL4A6 gene alone can cause Alport syndrome; however, deletions at the 5′ end of COL4A6 extending to the neighbouring COL4A5 gene (Fig. 5) have been reported to cause diffuse leiomyomatosis associated with Alport syndrome (*8, 37*), which is discussed below.

Diffuse Leiomyomatosis associated with Alport syndrome

One particularly interesting Alport phenotype includes diffuse leiomyomatosis (DL) in addition to the regular symptoms of Alport syndrome (AS). COL4A5 is tightly linked in a head-to-head manner to the functionally related and coordinately regulated COL4A6 gene, as mentioned above (see Figs. 1 and 5). However, there have been no X-linked AS cases due to COL4A6 mutations. On the other hand, DL/AS cases are always associated with deletions spanning the 5′ regions of the COL4A5/COL4A6 cluster (*31, 37, 70*). Unlike the COL4A5 breakpoints, the COL4A6 breakpoints have been found clustered within intron 2 of the gene. Recently, we identified a DL/AS deletion and characterized the breakpoint sequences, it being demonstrated that a 17 kb deletion eliminates the first coding exon of COL4A5 and the first two of COL4A6 (*32*). Intriguingly, the information also showed that the breakpoints share the same sequence, which is in turn closely homologous to the consensus sequences of topoisomerases I and II. These findings suggest that topoisomerase I and II activities might be involved in the mechanisms of DNA breakage and reunion. Additional DNA evidence indicated that the male patient is a somatic mosaic for the mutation.

What causes DL in the cases of deletion of the upstream region of both the *COL4A5* and *COL4A6* genes? Since one case having a large deletion including the entire COL4A6 exhibited only the AS and not the DL phenotype (*37*), it is not plausible that the entire loss of COL4A6 expression causes DL. There might be a third gene and/or its regulatory elements within the long intervening sequence III that regulates smooth muscle cell proliferation.

Concluding remarks

The collagen IV subfamily comprises six genes whose products constitute the major scaffold components of the BMs. Although we do not yet know the precise functions of all these components, it is already evident that they play important roles in development and morphogenesis. At least we predict the two types of collagen IV molecules exist in a tissue-specific manner; one might function as a filtering unit, which is observed in renal glomerular BMs, and the other might be needed in tissues, such as the aorta walls, where there are certain pressure loads against blood pressure. We hypothesized how the six genes evolved. Genetic analysis of the human and dog mutations in collagen IV genes provided significant information on the molecular forms of proteins and the organization of collagen IV genes. Experimental mutations in mice and gene targeting will supply key knowledge as to the functions of collagen IV genes.

REFERENCES

1. Timpl, R. and Brown, J.C. (1996) Supramolecular assembly of basement membranes. *Bioessays* **18**, 123-132
2. Olsen, B.R. and Ninomiya, Y. (1998) Collagens in *Guide Book to the Extracellular Matrix and Adhesion Proteins* (Kreis, T. and Vale, R., eds.) Boca Raton, FL, CRC Press in press
3. Emanuel, B.S., Sellinger, B.T., Gudas, L.J., and Myers, J.C. (1986) Localization of the human procollagen α1(IV) gene to chromosome 13q34 by in situ hybridization. *Am. J. Hum. Genet.* **38**, 38-44
4. Boyd, C.D., Toth-Fejel, S., Gadi, I.K., Litt, M., Condon, M.R., Kolbe, M., Hagen, I.K., Kurkinen, M., Mackenzie, J.W., and Magenis, E. (1988) The genes coding for human proα1(IV) and pro α2(IV) collagen are both located at the end of the long arm of chromosome 13. *Am. J. Hum. Genet.* **42**, 309-314
5. Morrison, K.E., Mariyama, M., Yang-Feng, T.L., and Reeders, S.T. (1991) Sequence and localization of a partial cDNA encoding the human α3 chain of type IV collagen. *Am. J. Hum. Genet.* **49**, 545-554
6. Kamagata, Y., Mattei, M.-G., and Ninomiya, Y. (1992) Isolation and sequencing of cDNAs and genomic DNAs encoding the α4 chain of basement membrane collagen type IV and assignment of the gene to the distal long arm of human chromosome 2. *J. Biol. Chem.* **267**, 23753-23758
7. Hostikka, S.L., Eddy, R.L., Byers, M.G., Hoyhtya, M., Shows, T.B., and Tryggvason, K. (1990) Identification of a distinct type IV collagen α chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome. *Proc. Natl. Acad. Sci. USA* **87**, 1606-1610
8. Zhou, J., Mochizuki, T., Smeets, H., Antignac C., Laurila, P., Paepe, A., Tryggvason, K., and Reeders, S.T. (1993) Deletion of the paired α5(IV) and α6(IV) collagen genes in inherited smooth muscle tumors. *Science* **261**, 1167-1169
9. Oohashi, T., Sugimoto, M., Mattei, M.-G., and Ninomiya, Y. (1994) Identification of a new collagen IV chain, α6(IV), by cDNA isolation and assignment of the gene to chromosome Xq22, which is the same locus for COL4A5. *J. Biol. Chem.* **269**, 7520-7526
10. Poschl, E., Pollner, R., and Kuhn, K. (1988) The gene for the α1(IV) and α2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure. *EMBO J.* **7**, 2687-2695
11. Soininen, R., Houtari, M., Hostikka, S.L., Prockop, D.J., and Tryggvason, K. (1988) The structural genes for α1 and α2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region. *J. Biol. Chem.* **263**, 17217-17220
12. Burbelo, P.D., Martin, G.R., and Yamada, Y. (1988) α1(IV) and α2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer. *Proc. Natl. Acad. Sci. USA* **85**, 9679-9682
13. Kaytes, P., Wood, L., Theriault, N., Kurkinen, M., and Vogeli, G. (1988) Head-to-head arrangement of murine type IV collagen genes. *J. Biol. Chem.* **263**, 19274-19277
14. Sugimoto, M., Oohashi, T., and Ninomiya, Y. (1994) The genes COL4A5 and COL4A6, coding for basement membrane collagen chains α5(IV) and α6(IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. *Proc. Natl. Acad. Sci. USA* **91**, 11679-11683
15. Weil, D., Mattei, M.-G. Passage, E., Vang Cong, N., Pribula-Conway, D., Mann, K., Deutzmann, R., Timpl, R., and Chu, M.-L. (1988) Cloning and chromosomal location of human genes encoding the three chains of type VI collagen. *Am. J. Hum. Genet.* **42**, 435-445
16. Emanuel, B.S., Cannizaro, L.A., Seyer, J.M., and Myers, J.C. (1985) Human α1(III) and α2(V) procollagen genes are located on the long arm of chromosome 2. *Proc. Natl. Acad. Sci. USA* **82**, 3385-3389
17. Thorner, P.S., Zheng, K., Kalluri, R., Jacobs, R., and Hudson, B.G. (1996) Coordinate gene expression of the alpha3, alpha4, and alpha5 chains of collagen type IV. Evidence from a canine model of X-linked nephritis with a COL4A5 gene mutation. *J. Biol. Chem.* **271**, 13821-13828
18. Guo, X. and Kramer, J.M. (1989) The two *Caenorhabditis elegans* basememt membrane (type IV) collagen genes are located on separate chromosomes. *J. Biol. Chem.* **264**, 17574-17582
19. Yasothornsrikul, S., Davis, W.J., Cramer, G., Kimbrell, D.A., and Dearolf, C.R. (1997) Viking: identification and characterization of a second type IV collagen in *Drosophila*. *Gene* **198**, 17-25
20. Soininen, R., Houtari, M., Ganguly, A., Prockop, D.J., and Tryggvason, K. (1989) Structural organization of the gene for the α1(IV) chain of human type IV collagen. *J. Biol. Chem.* **264**,

13565-13571
21. Buttice, G., Kaytes, P., D'Armiento, J., Vogeli, G., and Kurkinen, M. (1990) Evolution of collagen IV genes from a 54-base pair exon: A role for introns in gene evolution. *J. Mol. Evol.* **30**, 479-488
22. Quinones, S., Bernard, D., Carcia-Sogo, M., Elena, S.F., and Saus, J. (1992) Exon/intron structure of the human α3(IV) collagen gene encompassing the Goodpasture antigen (α3(IV) NC1). *J. Biol. Chem.* **267**, 19780-19784
23. Sugimoto, M., Oohashi, T., Yoshioka, H., Matsuo, N., and Ninomiya, Y. (1993) cDNA isolation and partial gene structure of the human α4(IV) collagen chain. *FEBS Lett.* **330**, 122-128
24. Zhou, J., Leinonen, A., and Tryggvason, K. (1994) Structure of the human type IV collagen *COL4A5* gene. *J. Biol. Chem.* **269**, 6608-6614
25. Oohashi, T., Ueki, Y., Sugimoto, M., and Ninomiya, Y. (1995) Isolation and structure of the COL4A6 gene encoding the human α6(IV) collagen chain and comparison with other type IV collagen genes. *J. Biol. Chem.* **270**, 26863-26867
26. Kishiro, Y., Kagawa, M., Naito, I., and Sado., Y. (1995) A novel method of preparing rat-monoclonal antibody-producing hybridomas by using rat medial iliac lymph node cells. *Cell Struct. Funct.* **20**, 151-156
27. Ninomiya, Y., Kagawa, M., Iyama, K.-I., Naito, I., Kishiro, Y., Seyer, J.M., Sugimoto, M., Oohashi, T., and Sado, Y. (1995) Differential expression of two basement membrane collagen genes, COL4A6 and COL4A5, demonstrated by immunofluorescence staining using peptide-specific monoclonal antibodies. *J. Cell Biol.* **130**, 1219-1229
28. Fleischmajer, R., Kuhn, K., Sato, Y., MacDonald, E.D., Perlish, J.S., Pan, T.C., Chu, M.L., Kishiro, Y., Oohashi, T., Bernier, S.M., Yamada, Y., and Ninomiya, Y. (1997) There is temporal and spatial expression of α1(IV), α2(IV), α5(IV), α6(IV) collagen chains and β integrins during the development of the basal lamina in an "in vitro" skin model. *J. Invest. Dermatol.* **109**, 527-533
29. Tanaka, K., Iyama, K., Kitaoka, K., Ninomiya, Y., Oohashi, T., Sado, Y., and Ono, T. (1997) Differential expression of α1(IV), α2(IV), α5(IV), α6(IV) collagen chains in the basement membrane of basal cell carcinoma. *Histochem. J.* **29**, 563-570
30. Dehan, P., Waltregny, D., Beschin, A., Noel, A., Castronovo, V., Tryggvason, K., De Leval, J., and Fiodart, J.M. (1997) Loss of type IV collagen α5 and α6 chains in human invasive prostate carcinoma. *Am. J. Pathol.* **151**, 1097-1104
31. Heidet, L., Cai, Y., Sado, Y., Ninomiya, Y., Thorner, P., Guicharnaud, L., Boye, E., Chauvet, V., Solal, L.C., Beziau, A., Torres, R.G., Antignac, C., and Gubler, M.-C. (1997) Diffuse leiomyomatosis associated with X-linked Alport syndrome: Extracellular matrix study using immunohistochemistry and in situ hybridization. *Lab. Invest.* **76**, 233-243
32. Ueki, Y., Naito, I., Oohashi, T., Sugimoto, M., Seki, T., Yoshioka, H., Sado, Y., Sato, H., Sawai, T., Sasaki, F., Matsuoka, M., Fukuda, S., and Ninomiya, Y. (1998) Topoisomerase I and II consensus sequences in a 17-kb deletion junction of the *COL4A5* and *COL4A6* genes and immunohistochemical analysis of esophageal leiomyomatosis associated with Alport syndrome. *Am. J. Hum. Genet.* **62**, 253-261
33. Genersch, E., Eckerskorn, C., Lottspeich, F., Herzog, C., Kuhn, K., and Poschl, E. (1995) Purification of the sequence-specific transcription factor CTCBF, involved in the control of human collagen IV genes: Subunits with homology to Ku antigen. *EMBO J.* **14**, 791-800
34. Burbelo, P.D., Utani, A., Pan, Z.-Q., and Yamada, Y. (1993) Cloning of the large subunit of activator 1 (replication factor C) reveals homology with bacterial DNA ligases. *Proc. Natl. Acad. Sci. USA* **90**, 11543-11547
35. Tanaka, S., Kaytes, P., and Kurkinen, M. (1993) An enhancer for transcription of collagen IV genes is activated by F9 cell differentiation. *J. Biol. Chem.* **268**, 8862-8870
36. Yamada, Y. and Kuhn, K. (1993) Genes and regulation of basement membrane collagen and laminin synthesis in *Molecular and Cellular Aspects of Basement Membranes* (Rohrbach, D.H. and Timpl, R., eds.) pp. 121-146, San Diego, Academic Press
37. Heidet, L., Dahan, K., Zhou, J., Xu, Z., Cochat, P., Gould, J.D.M., Leppig, K.A., *et al.* (1995) Deletions of both α5(IV) and α6(IV) collagen genes in Alport syndrome and in Alport syndrome associated with smooth muscle tumors. *Hum. Mol. Genet.* **4**, 99-108
38. Kendall, E., Evans, W., Jin, H., Holland, J., and Vetrie, D. (1997) A complete YAC contig and cosmid interval map covering the entirety of human Xq21.33 to Xq22.3 from DXS3 to DXS287. *Genomics* **43**, 171-182
39. Timpl, R. and Martin, G.R. (1981) Components of basement membranes in *Immunohistochemistry of the Extracellular Matrix*, Vol. II (Furthmayer, H., ed.) pp. 120-150, CRC Press, Boca Raton, Florida
40. Sado, Y., Kagawa, M., Kishiro, Y., Sugihara, K., Naito, I., Seyer, J.M., Sugimoto, M., Oohashi, T., and Ninomiya, Y. (1995) Establishment of epitope-defined monoclonal antibodies recognizing the six different α chains of human type IV collagen by the rat lymph node method. *Histochem. Cell Biol.* **104**, 267-275
41. Naito, I., Kawai, S., Nomura, S., Sado, Y., and Osawa, G. (1996) Relationship between COL4A5 gene mutations and distribution of type IV collagen in male X-linked Alport syndrome. *Kidney Int.* **50**, 304-311
42. Fukuda, K., Hori, H., Utani, A., Burbelo, P.D., and Yamada, Y. (1997) Formation of recombinant triple-helical $[\alpha1(IV)]_2\alpha2(IV)$ collagen molecules in CHO cells. *Biochem. Biophys. Res. Commun.* **231**, 178-182
43. Nakanishi, K., Yoshikawa, N., Iijima, K., Kitagawa, K., Nakamura, H., Ito, H., Yoshioka, K., Kagawa, M., and Sado, Y. (1994) Immunohistochemical study of α1-5 chains of type IV collagen in hereditary nephritis. *Kidney Int.* **46**, 1413-1421
44. Hino, S., Takemura, T., Sado, Y., Kagawa, M., Oohashi, T., Ninomiya, Y., and Yoshioka, K. (1996) Absence of α6(IV) collagen in kidney and skin of X-linked Alport syndrome. *Pediat. Nephrol.* **10**, 742-744
45. Gubler, M.C., Knebelmann, B., Beziau, A., Bryer, M., Pirson, Y., Haddoum, F., Kleppel, M.M., and Antignac, C. (1995) Autosomal recessive Alport syndrome: Immunohistochemical study of type IV collagen chain distribution. *Kidney Int.* **47**, 1142-1147
46. Miner, J.H. and Sanes, J.R. (1996) Molecular and functional defects in kidneys of mice lacking collagen α3(IV): Implications for Alport syndrome. *J. Cell Biol.* **135**, 1403-1413
47. Cosgrove, D., Meeham, D.T., Grunkemeyer, J.A., Kornak, J.M., Sayers, R., Hunter, W. J., and Samuelson, G.C. (1996) Collagen COL4A3 knockout: a mouse model for autosomal Alport syndrome. *Genes Dev.* **10**, 2981-2992
48. Kahsai, T.Z., Enders, G.C., Gunwar, S., Brunmark, C., Wieslander, J., Kalluri, R., Zhou, J., Noelken, M.E., and Hudson, B.G. (1997) Seminiferous tubule basement membrane. Composition and organization of type IV collagen chains, and the linkage of α3(IV) and α5(IV) chains. *J. Biol. Chem.* **272**, 17023-17032
49. Brazel, D., Oberbaumer, I., Dieringer, H., Babel, W., Glanville, R.W., Deutzmann, R., and Kuhn, K. (1987) Completion of amino acid sequence of the α1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain. *Eur. J. Biochem.* **168**, 529-536
50. Brazel, D., Pollner, R., Oberbaumer, I., and Kuhn, K. (1988) Human basement membrane collagen (type IV). The amino acid sequence of the α2(IV) chain and its comparison with α1(IV) chain reveals deletions in the α1(IV) chain. *Eur. J. Biochem.* **172**, 35-42
51. Soininen, R., Haka, R.T., Prockop, D.J., and Tryggvason, K. (1987) Complete primary structure of the α1-chain of human basement membrane (type IV) collagen. *FEBS Lett.* **225**, 188-194
52. Hostikka, S.L. and Tryggvason, K. (1988) The complete primary structure of the α2 chain of human type IV collagen and comparison with the α1(IV) chain. *J. Biol. Chem.* **263**, 19488-19493
53. Saus, J., Quinones, S., MacKrell, A., Blumberg, B., Muthukumaran, G., Pihlajaniemi, T., and Kurkinen, M. (1989) The complete primary structure of mouse α2(IV) collagen. Alignment with mouse α1(IV) collagen. *J. Biol. Chem.* **264**, 6318-6324

54. Muthukumaren, H., Blumberg, B., and Kurkinen, M. (1989) The complete primary structure of the α1-chain of mouse collagen IV. Differential evolution of collagen IV domains. *J. Biol. Chem.* **264**, 6310-6317
55. Mariyama, M., Leinonen, A., Mochizuki, T., Tryggvason, K., and Reeders, S. (1994) Complete primary structure of the human α3(IV) collagen chain: Coexpression of the α3(IV) and α4(IV) collagen chains in human tissues. *J. Biol. Chem.* **269**, 23013-23017
56. Leinonen, A., Mariyama, M., Mochizuki, T., Tryggvason, K., and Reeders, S. (1994) Complete primary structure of the human type IV collagen α4(IV) chain: Comparison with structure and expression of the other α(IV) chains. *J. Biol. Chem.* **269**, 26172-26177
57. Zhou, J., Hertz, J.M., Leinonen, A., and Tryggvason, K. (1992) Complete amino acid sequence of the human α5(IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient. *J. Biol. Chem.* **267**, 12475-12481
58. Zheng, K., Thorner, P.S., Marrano, P., Baumal, R., and McInnes, R.R. (1994) Canine X chromosome linked hereditary nephritis: A genetic model for human X-linked hereditary nephritis resulting from a single base mutation in the gene encoding the α5 chain of collagen type IV. *Proc. Natl. Acad. Sci. USA* **91**, 3989-3993
59. Kinoshita, Y., Osawa, G., Morita, T., Kobayashi, N., Wada, J., Ebe, T., Watanabe, M., Muronashi, K., and Murayama, M. (1969) Hereditary chronic nephritis (Alport) complicated by nephrotic syndrome. Light, fluorescent and electronmicroscopic studies of renal biopsy specimens. *Acta Med. Biol. Niigata* **17**, 101-117
60. Kefalides, N.A. (1971) Isolation of a collagen from basement membranes containing three identical α-chains. *Biochem. Biophys. Res. Commun.* **45**, 226-234
61. Spear, G.S. (1973) Alport's syndrome: A consideration of pathogenesis. *Clin. Nephrol.* **1**, 336-337
62. Pihlajaniemi, T., Tryggvason, K., Myers, J., Kurkinen, M., Lebo, R., Cheung, M., Prockop, D.J., and Boyd, C.D. (1985) cDNA clones for the pro-α1(IV) chain of human type IV procollagen reveal an unusual homology of amino acid sequences in two halves of the carboxyl-terminal domain. *J. Biol. Chem.* **260**, 7681-7687
63. Hostikka, S.L., Kurkinen, M., and Tryggvason, K. (1987) Nucleotide sequence coding for the human type IV collagen α2 chain cDNA reveals extensive homology with the NC-1 domain of α1(IV), but not with the collagenous domain or untranslated region. *FEBS Lett.* **216**, 281-286
64. Atkin, C.L., Hasstedt, S.J., Menlove, L., Cannon, L., Kirchner, N., Schwartz, N., Nguyen, K., and Skolnick, M. (1988) Mapping of Alport syndrome to the long arm of the X-chromosome. *Am. J. Hum. Genet.* **42**, 249-255
65. Barker, D., Hostikka, S.L., Zhou, J., Chow, L.T., Oliphant, A.R., Gerken, S.C., Gregory, M.C., Skolnick, M.H., Atkin, C.L., and Tryggvason, K. (1990) Identification of mutations in the COL4A5 collagen gene in Alport syndrome. *Science* **248**, 1224-1227
66. Prockop, D. and Kivirikko, K.I. (1997) Collagens: molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* **64**, 403-434
67. Mochizuki, T., Lemmink, H.H., Mariyama, M., Antignac, C., Gubler, M.-C., Pirson, Y., Verellen-Dumoulin, C., Chan, B., Schroder, C.H., Smeets, H.J., and Reeders, T. (1994) Identification of mutations in the α3(IV) and α4(IV) collagen genes in autosomal recessive Alport syndrome. *Nature Genet.* **8**, 77-82
68. Lemmink, H.H., Mochizuki, T., van den Heuvel, L.P.W.J., Schroder, C.H., Barrientos, A., Monnens, L.A.H., van Oost, B.A., Brunner, H.G., Reeders, S.T., and Smeets, H.J. (1994) Mutations in the type IV collagen α3 (COL4A3) gene in autosomal recessive Alport syndrome. *Hum. Mol. Genet.* **3**, 1269-1273
69. Knebelmann, B., Forestier, R., Drouot, L., Quinones, S., Chuet, C., Benessy, F., Saus, J., and Antignac, C. (1995) Splice-mediated insertion of an *Alu* sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. *Hum. Mol. Genet.* **4**, 675-679
70. Dahan, K., Heidet, L., Zhou, J., Mettler, G., Leppig, K.A., Proesmans, W., David, A., *et al.* (1995) Smooth muscle tumors associated with X-linked Alport syndrome: Carrier detection in females. *Kidney Int* **48**, 1900-1906