

Two genes, *COL4A3* and *COL4A4* coding for the human $\alpha 3(\text{IV})$ and $\alpha 4(\text{IV})$ collagen chains are arranged head-to-head on chromosome 2q36

Ryusuke Momota^a, Manabu Sugimoto^a, Toshitaka Oohashi^a, Kazuteru Kigasawa^b,
Hidekatsu Yoshioka^a, Yoshifumi Ninomiya^{a,*}

^aDepartment of Molecular Biology and Biochemistry, Okayama University Medical School, 2-5-1, Shikata-cho, Okayama 700, Japan

^bDepartment of Ophthalmology, Tokai University School of Medicine, Bohseidai, Isehara, Kanagawa 259-11, Japan

Received 27 October 1997; revised version received 27 January 1998

Abstract We first isolated and characterized genomic DNA fragments that cover the 5' flanking sequences of *COL4A3* and *COL4A4* encoding the human basement membrane $\alpha 3(\text{IV})$ and $\alpha 4(\text{IV})$ collagen chains, respectively. Nucleotide sequence analysis indicated that the two genes are arranged head-to-head. To determine transcription start site for *COL4A4* gene, we performed RACE and RNase protection assays, indicating that there are two alternative transcripts presumably derived from two different promoters. Interestingly, one transcription start site (from exon 1') of *COL4A4* is only 5 bp away from the reported transcription start site of *COL4A3*, whereas the other transcript (from exon 1) starts 373 nucleotides downstream from the first one, generating the two kinds of transcripts that differ in the 5' UTR regions. Expression of these two transcripts appears tissue-specific; exon 1 transcript was expressed predominantly in epithelial cells, while exon 1' transcript showed rather ubiquitous and low expression. The nucleotide sequence of the promoter region is composed of dense CpG dinucleotides, GC boxes, CTC boxes and a CCAAT box but no TATA box. These results provide information to delineate the promoter activity for the tissue-specific expression of the six type IV collagen genes and basement membrane assembly in different tissues and organs.

© 1998 Federation of European Biochemical Societies.

Key words: Alternative transcription; Type IV collagen; *COL4A4*; *COL4A3*; Synteny; Basement membrane

1. Introduction

The collagen superfamily now contains 19 molecules formally defined as collagens and at least 33 genetically different collagen α chains [1]. The superfamily can be divided into several classes on the bases of polymeric structural features, expression manner, and gene structure: fibrillar, FACIT, basement membrane, short chain, multiplexins, and other collagens [2]. One of them, basement membrane type IV collagen, is composed of six genetically distinct α chains; $\alpha 1(\text{IV})$ and $\alpha 2(\text{IV})$, the two major chains, and $\alpha 3(\text{IV})$, $\alpha 4(\text{IV})$, $\alpha 5(\text{IV})$, and $\alpha 6(\text{IV})$, the four minor ones [3]. The six chains differ with respect to their tissue distribution; i.e. $\alpha 1/\alpha 2(\text{IV})$

*Corresponding author. Fax: +81 (86) 222-7768.
E-mail: yoshinin@ccews2.cc.okayama-u.ac.jp

Abbreviations: nt, nucleotide(s); COL, collagenous; NC, non-collagenous; RPE, retinal pigmental epithelium; RACE, rapid amplification of cDNA ends

The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank Nucleotide sequence Databases under accession number AB008495 and AB008496.

are ubiquitous, whereas the other minor chains appear to have a more restricted distribution [4]. The genes for the human $\alpha 1(\text{IV})$ and $\alpha 2(\text{IV})$ chains, *COL4A1* and *COL4A2*, are known to reside closely together at locus q22 on chromosome 13 and to form a transcription unit in which they are arranged head-to-head and fused together by a common promoter region [5,6]. A similar promoter arrangement for the genes of the human $\alpha 5(\text{IV})$ and $\alpha 6(\text{IV})$ chains, *COL4A5* and *COL4A6*, is seen in region q22 of chromosome X; however, there are two alternative promoters that drive the formation of two alternative transcripts from the *COL4A6* gene [7]. This pair shares some characteristic features in their 5' flanking region: the lack of a TATA motif and the presence of CCAAT, GC box and CTC box. In this report we describe the promoter region of the human genes for $\alpha 3(\text{IV})$ and $\alpha 4(\text{IV})$, i.e. *COL4A3* and *COL4A4*, which are known to exist on chromosome 2 [8,9]. Nucleotide sequence analysis revealed the head-to-head arrangement and the regulatory elements of the two genes. Further, we demonstrate the presence of the two alternative transcripts for the *COL4A4* gene, which were derived from the presumptive alternative promoters.

2. Materials and methods

2.1. Cell culture

TCO-1 (thyroid carcinoma cell, [10]), MRC-9 (fetal lung fibroblast), TIG-1 (fetal lung fibroblast, [11]), and IMR-90 (fetal lung fibroblast) cell lines were obtained through the Human Science Research Resources Bank (Osaka, Japan). Human retinal pigment epithelial cell (K-1034) was described previously [12]. Cultured cells were grown in 75 cm² plastic dishes in Dulbecco's modified Eagle's medium with 10% fetal bovine serum at 37°C in 5% CO₂ incubator.

2.2. cDNA cloning and screening of a genomic library

A λ gt11 cDNA library (Clontech, HL1123b) from human whole kidney was screened by use of the previously described cDNA (MS2) encoding the human $\alpha 4(\text{IV})$ chain as a probe [13]. Several overlapping clones, shown in Fig. 1, were obtained and characterized by nucleotide sequencing. The seven clones represented the entire coding region of the human $\alpha 4(\text{IV})$ chain; and the amino-terminal clone, MT107, contained a putative signal peptide and a part of the 5' untranslated region. A total genomic library in EMBL3 (Clontech, HL1006d) was screened, and the isolated clones were confirmed by sequential Southern-blot analysis using a *Sau*96I fragment of MT107 (MT107Sau96I, Figs. 1 and 2) and an oligomer, A3L2, 5'-ATGG-TGGGAGAGAGCTCAGA-3', complementary to nucleotides 163–144 of $\alpha 3(\text{IV})$ cDNA counting from the transcription start site [14], on the assumption that the genes for $\alpha 3(\text{IV})$ and $\alpha 4(\text{IV})$ would be closely together as observed for other $\alpha(\text{IV})$ gene pairs [5–7]. DNA labeling for probes and screening for clones were carried out by standard methods [15].

2.3. Nucleotide sequence analysis

Enzymatically fragmented phage clones were subcloned in double-

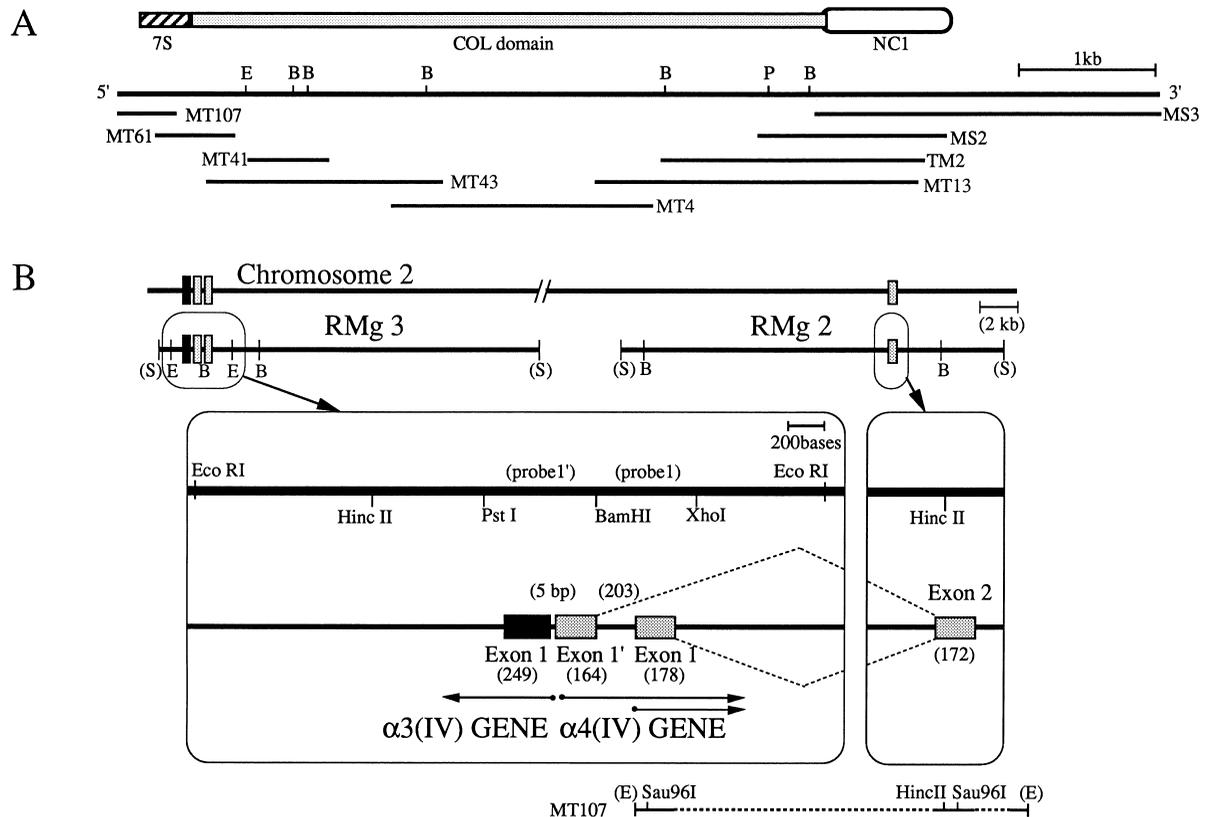


Fig. 1. A: Schematic representation of the human $\alpha 4(\text{IV})$ chain and restriction enzyme map of the overlapping cDNA clones. Top: the relative domain structure of the $\alpha 4(\text{IV})$ chain. The amino-terminal 7S domain, the central triple-helical domain, and the carboxyl non-collagenous domain are indicated by 7S, COL, and NCI, respectively. Bottom: restriction map with the consecutive cDNA sequence and corresponding cDNA clones of MT107 (insert 0.5 kb in size), MT61 (0.8 kb), MT41 (0.6 kb), MT43 (1.7 kb), MT4 (1.8 kb), MT13 (2.2 kb), TM2 (1.9 kb), MS2 (1.2 kb), and MS3 (2.5 kb). Restriction enzyme sites *Bam*HI (B), *Eco*RI (E), and *Pst*I (P) are shown. B: Schematic representation of *COL4A4* and *COL4A3* on chromosome 2. Relative location of exon 1 for *COL4A3*, of exons 1 and 1' for *COL4A4* in region 235 of chromosome 2 (top), and of the two isolated genomic clones, RMg2 and RMg3 (middle), is shown in a same scale. Exon-containing regions are magnified in the two boxes to indicate how the two transcripts for *COL4A4* are spliced. Restriction enzyme sites for *Bam*HI and *Eco*RI are indicated by B and E, respectively, below the two genomic clones. *Sal*I sites (S) at both ends of the clones represent the cloning sites. Directions of transcription are indicated by arrows. Note the two transcripts for *COL4A4* starting from exons 1' and 1. Below the two boxes the MT107 cDNA is shown.

stranded pBluescript II vectors and subjected to sequence analysis. The fluorescence labeled dye-terminator method was used with an Applied Biosystems model 373A automatic sequencer (Central Research Laboratory, Okayama University Medical School). Sequence was checked from both strands and was analyzed by the computer program MacVector (Oxford Molecular Group PLC, UK).

2.4. Rapid amplification of cDNA 5' ends (RACE)

To determine the 5' end of the transcript and to see expression in human tissues or cultured cells, we carried out 5' RACE [16]. Total RNA was extracted from cultured retinal pigment epithelial (RPE) cells by the guanidinium thiocyanate method [12], and 5 μg of total RNA was used as a template for synthesis of the first-strand cDNA. We synthesized two primers, MT107R2: 5'-TTCTGCCAGT-CTTCTCTTCC-3' (complementary to nucleotides 72–53 of exon 2 of *COL4A4*) and MT107R4: 5'-ACGTGAGGTTCTGTGTTCTG-3' (complementary to nucleotides 35–16 of exon 2 of *COL4A4*). The first-strand cDNA was synthesized by Moloney murine leukemia virus reverse transcriptase (United States Biochemical) from the MT107R2 primer and was tailed by a stretch of dA by terminal deoxytransferase (GIBCO). Using a small aliquot of the reaction as a template we performed PCR: 94°C, 40 s; 58°C, 1 min; 70°C, 3 min for 40 cycles for amplification, and 70°C, 15 min for elongation. The products were electrophoresed in 1.5% agarose gels, blotted onto Hybond-N⁺ (Amersham) membranes, and hybridized with the two specific probes: RMg3-1*Pst*I *Bam*HI fragment (833 nt) for exon 1' (probe 1', see Fig. 1) and RMg3-1*Bam*HI *Xho*I fragment (603 nt) for exon 1 (probe 1). Southern blotting was carried out by standard methods. Cloning of

the PCR products was performed by use of the TA cloning system (Invitrogen).

2.5. RNase protection assay

Two primers (A3L5: 5'-CTCTCAGGCCACCGCGTCTC-3' (nt 560–579 in *COL4A3*, Fig. 3), RMgsp1: 5'-CGCGCCTCCCG-CTTAATCTG-3' (nt 1023–1004, Fig. 3) were designed for obtaining the fragments encompassing each portion of the upstream exons of both *COL4A3* and *COL4A4* genes by PCR. The generated fragments were subcloned into pBluescript after some modifications, thus obtaining the templates of cRNA probes (pBIBamL5). For $\alpha 3$ cRNA probe production, pBIBamL5 was linearized with *Bam*HI and transcribed with T3 RNA polymerase. For $\alpha 4$ exon 1' cRNA probe, it was linearized with *Eco*RI and transcribed with T3 RNA polymerase. RMg3-1*Bam*HI*Xho*I fragment was digested with *Not*I and was subcloned into pCRII vector. This plasmid was linearized with *Bam*HI and transcribed with T7 RNA polymerase. These riboprobes were purified with 7 M urea-5% polyacrylamide gel. Preparation of [α -³²P]UTP labeled cRNA probes and RNase mapping of the transcripts were performed as described [15].

Ten micrograms of total RNAs were hybridized with each probe. The procedure was followed by manufacturer's manual (Amersham) except that hybridization was performed at 45°C. In each experiment, 10 μg of yeast tRNAs were hybridized with each probe, as negative controls. Samples were resolved in 7 M urea-5% polyacrylamide gels. After electrophoresis the gels were dried and autoradiographed. All experiments using radioisotope were done in Radioisotope Center, Okayama University.

2.6. Quantitative competitive PCR

To estimate promoter utilization rate of the two transcripts of COL4A4, we performed quantitative competitive PCR analysis according to the method by Fend et al. [17]. Two primer sets were designed for specific amplification of exon 1' and exon 1 containing transcripts. The first set for exon 1' containing transcript was au2: 5'-CCAACTTACCAGGCAGTGTC-3' (nt 751 to 770) and MT107R4 and the other set for exon 1 transcript was u2: 5'-GGGTG-GCAAGGCGGCAGCT-3' (nt 1191 to 1209) and MT107R4. The competitor for both exon 1' and exon 1 transcripts was generated by PCR using au2u2: 5'-CCAACTTACCAGGCAGTGTCGG-GTGGCAAGGCGGCAGCT-3' and glR4: 5'-AAGTGAGGTTCT-GTGTCTGACACAGCACCCAGTTAAAGC-3' (last half is complementary to 1424–1405). This fragment was subcloned into plasmid, cut out with appropriate restriction enzyme, purified with Prep-A-gene (BIO-RAD). The concentration of competitor was determined by spectrometry. One hundredth of reverse transcribed cDNA was added into PCR reaction mixture with 10-fold serial dilutions of competitor, 10 pmol of primers au2 or u2 and MT107R4 in 20 μ l of reaction volume. The 30 cycles consisted of each reaction; 94°C for 1 min, 62°C for 1 min and 72°C for 2 min. After amplification half of reaction was resolved by 1.7% agarose gel containing 10 mg/ml EtBr. This gel was photographed, scanned and analyzed using NIH image. The logarithm of the ratio of transcripts to competitor was graphed and used for estimation of the amount of transcripts.

3. Results and discussion

3.1. Isolation of genomic clones that contain the 5' upstream region

We previously isolated and reported cDNAs encoding the carboxyl part of the human α 4(IV) chain [13]. Using one of the cDNAs, MS2, as a probe, we started isolating overlapping cDNAs covering the region further upstream. Seven consecutive overlapping clones, TM2, MT4, MT13, MT41, MT43, MT61, and MT107 (Fig. 1A), were isolated and characterized by nucleotide sequencing and Southern-blot analysis. We identified that they encoded the human α 4(IV) chain.

MT107 was found as the most upstream cDNA clone, encoding a part of the 7S domain, a putative signal peptide and some 5' untranslated region.

Isolation of genomic clones containing the 5' upstream region of COL4A4 was performed along with the cDNA cloning. Two genomic clones of 20 kb in size, RMg2 and RMg3, were isolated from a genomic library screened with the cDNA probe MT107Sau96I (*Sau*96I-*Sau*96I fragment, 327 bp, Fig. 1B, bottom). Although both clones were isolated by use of the MT107Sau96I probe, when a 3' fragment (*Hinc*II-*Eco*RI, 272 bp, Fig. 1B) of MT107 was used as the probe, RMg2 showed hybridization; whereas RMg3 did not (data not shown). This result indicated that RMg3 contained genomic DNA upstream from RMg2. The MT107Sau96I cDNA probe hybridized to the RMg3 3 kb *Eco*RI-*Eco*RI fragment (Fig. 1B, left box) of the genomic DNA. Further, in Southern-blot analysis using the α 3(IV) gene-specific oligomer A3L2 (see Section 2) as a probe, A3L2 showed hybridization to the whole RMg3 and to the *Eco*RI-*Eco*RI fragment of it as well. These results suggested that the 3 kb *Eco*RI-*Eco*RI fragment contained the upstream regions of both COL4A4 and COL4A3 genes.

To make sure that this clone was not an artifact, we digested total human genomic DNA with *Eco*RI and hybridized the digest with the RMg3PstI/BamHI fragment (0.6 kb, Fig. 1B). This Southern-blot analysis (data not shown) resulted in the detection of the 3 kb *Eco*RI fragment as expected and confirmed the map shown in Fig. 1B. Nucleotide sequence analysis revealed the presence of exon 1 and exon 2 of COL4A4 in RMg3 and RMg2, respectively, and showed that the first exons of both COL4A3 and COL4A4 were within the 3 kb *Eco*RI-*Eco*RI fragment on opposite strands and separated by 372 bp, indicating that the genes were transcribed in opposite directions and shared the same promoter region, as observed for other α (IV) genes [5–7].

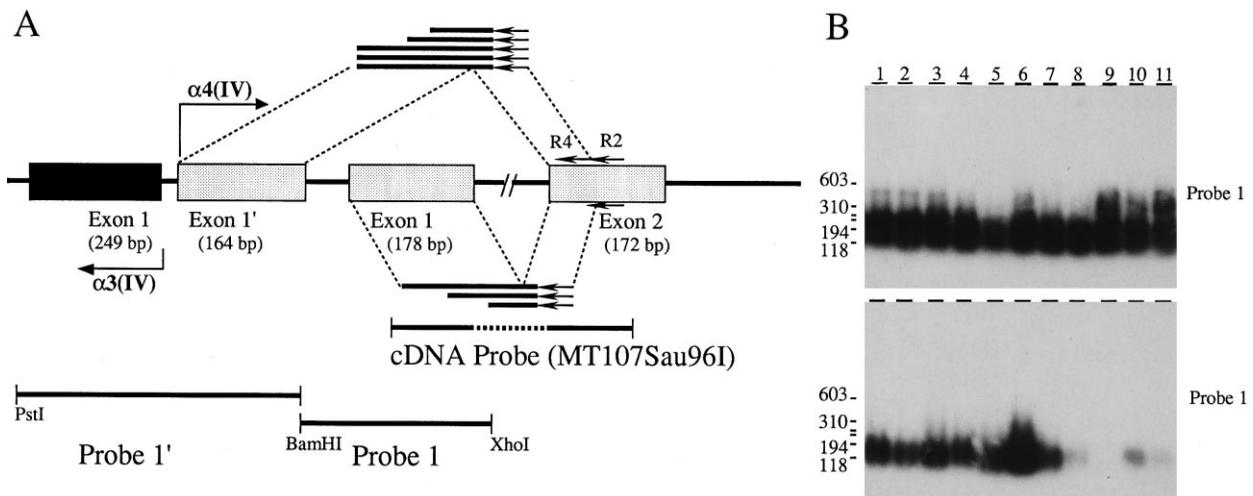


Fig. 2. A: Representation of the 5' RACE products and relative locations of probes. The first strand was synthesized from the R2 primer along the RNA of RPE cells. 5' RACE was performed between hybrid primer and R4. The 5' RACE products thus synthesized were hybridized by use of the two different probes. One group of the subclones that was hybridized with probe 1, the *Bam*HI-*Xho*I fragment (603 bp) from RMg3, contained exon 1 and a part of exon 2, which was the same as cDNA MT107. For the other group, the probe 1', the *Pst*I-*Bam*HI fragment (837 bp) of the RMg3, was used for subcloning. As expected the subclones contained the nucleotide sequence from exon 1' and the 5' part of exon 2. The two different transcripts were thus recognized for the COL4A4 gene expression. B: Expression of the two transcripts. RNAs from human lung (lane 1), kidney (lane 2), HeLa (human) cells (lane 3), retinal pigmented epithelial cells (4), thyroid carcinoma cells (5), keratinocytes (6), arterial endothelial cells (7), different lines of fetal lung fibroblasts (8–10), and adult skin fibroblasts (11) were used as templates for 5' RACE. The 5' RACE products were hybridized with the two different probes prepared from RMg3: probe 1 and probe 1' (see panel A). Note the very low expression of exon 1-containing transcript in lung (lanes 8–10) and skin (lane 11) fibroblasts.

A

CTGCAGGTGTCTCCTGCCAGCCCTCTGTGCGCTGGGCCAAGCTGCTTGATTTGATTACCCAGCGTGGACACCGTGCATCTCCACATCCCG 90
 GACGTCCACAGAGGACGGTTCGGGAGACACGCGACCCCGTTTCAGCAACTAACTAAGTGGTCGCACCTGTGGCACGTAGAGGTGTAGGGC

GTGTTCTCACGTATCATCCCTGGCGATTATCTATGTGTCTATTTTTTAATTTACTGACCTCTGGAGGCTGTCCCCGCCATCCCTGACTC 180
 CACAAGAGTGCATAGTAGGGACCCCTAATAGATACACAGATAAAAAATTAATGACTGGAGACCTCCGGACAGGGGCGGTAGGGACTGAG

CTAGGCACTGGACAAACCCCTCCTCGCCCTTTCCCCAGCCCCGCGCATTCCACCTGTTCGGGGGCATCCCCGCTTGCTCCCTTCCGCT 270
 GATCCGTGACCTGTTTCGGGGGAGGAGCGGAGAAGGGGTCGGGGCGCGTAAGGTGGACAAGGCCCCCTAGGGGGCAACGAGGGAAGGGCA

CGCTGCCTGTCCCGGGACCTGCGTGAAGCCCTCAGCCTCGCCACTAGCCCGGCAACCGCTCCACTACCGCACCGCTGCGGGTGGGGCCAG 360
 GCGACGGACAGGGCCCTGGACGCACTTCGGGAGTCGGAGCGGTGATCGGGCCGTGGCGAGGTGATGGCGTGGCGACGCCACCCCGTTC

CGCGTCCCCGGACGCGTGGAGGAGGGATGGAAAGTGCGGGGGTGGGGTTCGCGCAGCCCCCACTCACCTTGCTGGTGCGGGCGCCGCGC 450
 GCGCAGGGGGCCTGCGCACCTCTCCCTACCTTACGCCCCACCCCCAGCGCTCGGGGTGAGTGAACGACCGACGCCCGCGGGCGG

V T T A T T T d T T T A O d R d V L R V S W
 CCAGGAGCACCAGCAGGAGCGGCAGCAGGAGCACCTGCGGCCTGGGGCGGTCCGGGCGCTCATGGTGGGCGCGCGCTCAGAGCCTGGGC 540
 GTTCTCGTGGTCTCTCGCGTCTGTTGGAACGCGGACCCCGCCAGGCCCGGAGTACCACCGCGCGCGAGTCTCGGACCCG

GAGTCCGGGACCCCTCAGGCTCTCAGGCCACCGCGTCTCTGCGCAGCGCGCTCTGGAGCCCGGTGGAAAGCTCGCCTCGCCACTCG 630
 CTCAGGCCCTGGGAGTCCGAGAGTCCGGTGGCGCAGAGGACGCGTTCGCGCGAGACCTTCGGGCCGACCTTTCAGAGCGGAGCGGTGAGC

COL4A3 exon 1 ←
 GCGCGGGCGGGTTTTGTAGGGCTCGGTGCGCGGTTCTGCTCCCTCCCAACCCAGGCTTTTTGGTTGTGGCTTCAAGGGAGGGAAGACGTGG 720
 CGGCCCGGCCAAAACATCCCGAGCCAGCGCGCAAGCAGGAGGGTGGGTCCGAAAACCAACACCGAAGTTCCTCCCTTCTGACCC

COL4A4 exon 1' →
 CTCAGTCCCGCCCTCATAACCCCTCCCTCCCAACTTACCAGGCAGTGTCCCTTCTACCCGGGCATCGTGCCAGGAGCGGAGAAAGGGC 810
 GAGTCCAGGGCGGATGATGGGGGAGGGAGGGTTGAATGGTCCGTCACAGGGAAGATGGGCCGTAGCACGGTCTCTCCGCTTTCCCG

au2
 TGGCGCCCTGAGGGGAGGTGGGCGCGGATCCAGGCTAAGGGTTAAGAGGTGATGCTCCCAAGTCCGGGGGCCAGGCGGGGAAGTG 900
 ACCCGCGGACTCCCGTCCACCCGCGCCTAGGTCCATTCCCAATTCTCCACTACGAGGGTTCAGGCCCCCGGGTCCGCCCTTCC

GGGTGCGGTTCGGAGGTGGGGCGTTTGGGGTGTGCGGTGGGAGGTGGCTGCGCGGTGGGGAGGTGGCTGCGCAGGGTGGAGGGAGTGGC 990
 CCCACGCCAGCCTCCACCCCGCAAACCCACACGCGCACCCCTCCAGCGAGCGGCCACCCCTCCACGACGCGTCCCACCTCCCTCACC

CAGCGGCCCTGCCAGATTAAGCGGGAGGCGCGCGGGCGTCTGCACTCGGGAGGCCCGCTTGTTCGCCGCGCCCGGGCGGTCCGGGT 1080
 GTCGCCGGACGGGTCTAATTCGCCCTCCGCGCGGCCCGCAGACGTGAGCCCTCCGGGGCGAAACAAGGGGCGCGGGCGCCGACGGCGCA

COL4A4 exon 1
 TCCCGCCTCCAGCCCGCTCCAATTGCGGCCGCGGGCTTGGCTGCGAGAGCGCCGGAAGAGGCAGCGGGGCGGGTGGATTGGGG 1170
 AGGGGCGGAGGTTCGGGCGGAGGTGAACGCCGCGGCCCGAACCAGCGCTCTCGCGGGCTTCTCCGTCGCCCGCGCCACCTAACCC

CTGGAGGTGCGCGTCCCGTGGGGTGGCAAGCGCGCAGCTCTGGCGCTGCGGGCGTCCCCACAGTAAAGCGCCATAGGTGGGGCAACC 1260
 GACCTCCACGCGCAGGGCACCCACCGTCCGCGCTCGAGGACCGGACGCGCCGACGGGTGTCCATTCCGCGGTATCCACCCGTTGG

u2
 GCTGTCCGGGGCCCCGACCTGCCTGGGAGGGCAGTGGAGGAGCGTGGCATTCAAGGGGTGGTGGCGGGCCAGCTGGGGAGTTGT 1350
 CGACAGGGCCCCGGGGCGGTGGACGGACCCCTCCCGTCACTCCCTCGCACCGTAAGTCCCCACCACCGCCGGGTGACCCCTCAACA

GGCCCGGGGTCAACGAGGAGAGCGCGTCTGGCTCCTGGGCTTGTCTGCCTGGGCTTTAACTGGGTGCTGTGTGTTTTGATTAATTTGG 1440
 CCGGGCCCCAGTGGTCTCCTCTCGCGCAGACCGAGGACCCGAACAGACGGACCCGAAATTGACCCACGACACAAAACTAATTAACC

AGGCTCGAG 1449
 TCCGAGCTC

B

ATGAAAGTAATTTATTTTCAACATTATAATGCATGCTGTGTTTTAAGTGATATATTTATATCTAACAGCAACAGACTTTGACCCAGAACAC 90
 TACTTTTCAATAAAAGTTGTAATATTACGTACGACAAAAATCACTATATAAATATAGATTGCTTGTCTGAAACTGGGTCTTGTG

COL4A4 exon 2 ←
 AGAACCTCACTTGTCAACAAGAACCTTCTGGAAGAGAAGACTGGCAGAATTTTTTAAGTACTAAGACTTGCCTGCGATGTGGTCTCTG 180
 TCTTGGAGTGAACAGTTGTTCTTGGAAAGACCTTCTCTTCTGACCGTCTTATAAAAAATTCATGATTCTGAACGGACGCTACACCAGAGAC

R4 ← R2
 CACATAGTACTAATGAGGTGCTCCTCAGATTGACCAAGTCCCTGGCCACAGGTCCCTGGTGGAGTATCCTCAGTGTGATTGCCTGCTTTT 270
 GTGTATCATGATTACTCCAGGGAAGTCTAACTGGTTCAGGAACCGGTGTCCAGGGACCACTCATAGGAGTCACTAACGGACGAAAA

M W S L H I V L M R C S F R L T
 ATTTAGTAATTTCTACTCAATGTTTGTCTATGGATGTAA 310
 TAAATCATTAAGATGAGTTACAACAAGATACCTACATT

3.2. 5' RACE for $\alpha 4(IV)$ cDNA

To identify the 5' termini of the $\alpha 4(IV)$ transcripts, we performed 5' RACE using RNAs from cultured human RPE cells. The 5' RACE products hybridized with the

cDNA probe MT107Sau96I were sequenced. Three strongly hybridized clones shared the nucleotide sequence with the same 5' sequence in the non-translated region of MT107 cDNA, but their sequences were extended beyond the pre-

Fig. 3. Nucleotide sequence of the 1.4 kb *PstI-XhoI* fragment of RMg3 and exon 2 sequence of RMg2. A: A *PstI-XhoI* fragment of RMg3 (1.4 kb) that contains exon 1 for *COL4A3* and exons 1 and 1' for *COL4A4* was sequenced. Potential CCAAT box, GC boxes, CTC boxes and 19 nt tandem repeat are indicated in small shaded areas. Exons are indicated by boxed-in areas. The 5' ends of exon 1' and exon 1 of *COL4A4* are those of most extended clones. Transcription start sites estimated by RNase protection assay are indicated by vertical small arrows. Transcriptional directions are indicated by arrows. Deduced amino acid residues for the *COL4A4* translation product are shown below the boxes in the one-letter code, whereas those for the *COL4A3* product are shown above the boxes in inverted letters. Note that the translation starts within exon 1 for *COL4A3*, whereas both exons 1' and 1 for *COL4A4* contain a 5' UTR sequence. Three bases at the nucleotides 531, 533, and 535 were reported as A but we confirmed as C from both strands. B: Nucleotide sequence containing exon 2 for *COL4A4* was determined from a RMg2 fragment. Exon 2 is indicated in a boxed-in area. Exon 2 contains the codon for the first methionine of the $\alpha 4(IV)$ chain.

vious report [18]. The longest was more extended with additional 74 nt, which was found in the genomic DNA sequence of RMg3 that was described above. We designated this 5' UTR with additional 74 nt as exon 1 (178 bp) for *COL4A4*.

There was a group of clones that hybridized weakly to the cDNA MT107Sau961. Nucleotide sequence analysis demonstrated that the 3' 35 bp of the clones was the same as that of an other group of clones found in exon 2 in the genomic DNA sequence; however, the 5' side sequence of the clones was different. The longest three clones contained 164 bp at the 5' side, which was located between the exon 1 for *COL4A4* and the exon 1 for *COL4A3* in the genomic DNA (Fig. 2A). This indicates the presence of the alternative exon 1 for *COL4A4* and we designated this alternative 164 bp exon as exon 1'. No clone had both exon 1' and exon 1 at the same time, indicating that they were exclusively selected.

3.3. RNase protection assay

To confirm where the two transcripts start, we performed RNase protection assay. As shown in Fig. 4, both probes were protected at the predicted places: 720–730 for exon 1', 1050–1080 for exon 1. The assay also revealed the multiple protected bands for both transcripts. Multiple starting is a characteristic feature of housekeeping genes that contain GC-rich promoters with no TATA box. The transcription initiation sites were thus determined for the two first exons in the 5' RACE experiments. From comparison of the sequence between the genomic clone and the cDNA clones, we conclude that there are two kinds of transcripts that differ in their 5' end and are presumably driven from the two alternative promoters.

3.4. Expression of the two transcripts of $\alpha 4(IV)$ gene in various human tissues and cultured cells

To examine the expression of the two transcripts of *COL4A4*, we used Southern-blot analysis for 5' RACE products from RNAs of various human tissues and various kinds of cultured cells. Genomic fragments of RMg3 were used as probes for specific detection of the two transcripts: probe 1 for transcripts starting from exon 1, and probe 1' for newly found transcript starting from exon 1' (Fig. 2A). These two probes determined the level of expression of the two alternative transcripts since they hybridized only with exon 1 or exon 1'. Interestingly, the newly found transcript was widely detected by this method, whereas the level of transcript that contained exon 1 was quite low in cultured fibroblasts (Fig. 2B). When the sensitivity of this method is considered, cultured fibroblasts do not appear to represent major producing cells of the transcript containing exon 1. Quantitative PCR using the primer in exon 1 or exon 1' with reverse primer in exon 2 (see Section 2 and Fig. 3) revealed that TCO-1 cells

preferentially utilized exon 1 promoter by 100-fold of exon 1' promoter (data not shown).

3.5. *COL4A3* and *COL4A4* genes are transcribed to opposite direction

As shown in Fig. 1, RMg3 contained a part of intron 1 and exon 1 for *COL4A3*, the 5' termini of intergenic region for both *COL4A3* and *COL4A4*, exon 1', exon 1 and a part of intron 1 for *COL4A4*, and RMg2 contained exon 2 for *COL4A4* and introns. The nucleotide sequences of the promoter region of the two genes and exon 2 for *COL4A4* and

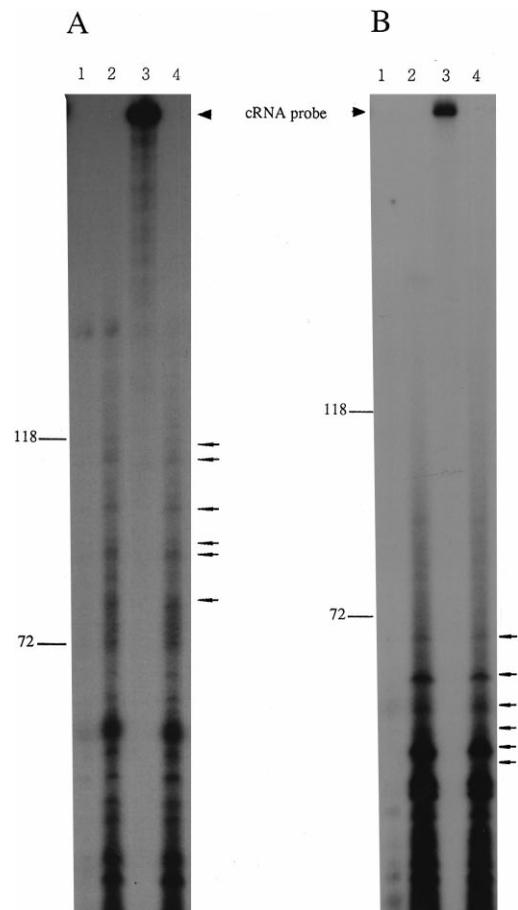


Fig. 4. RNase protection assay was performed to determine the 5' end structure of the two alternative transcripts, started from exon 1' (panel A) and exon 1 (panel B). [α - 32 P]UTP labeled cRNA was synthesized and hybridized to 10 μ g of total RNA derived from thyroid carcinoma cell, TCO-1 (lane 2) and retinal pigment epithelial cell, K1034 (lane 4). Yeast tRNA (10 μ g) was hybridized (lane 1) and non-treated cRNA probe (lane 3) was included in each run to refer the non-specific signals. Protected bands, indicated by arrows, were identified by comparison with exposed autoradiographs.

neighboring introns are shown in Fig. 3. Nucleotide sequence comparison between the genomic DNA and cDNA for $\alpha 4(\text{IV})$ and $\alpha 3(\text{IV})$ chains thus demonstrated that the two genes were arranged head-to-head and transcribed in opposite directions. A translation start codon was found in exon 2 of the *COL4A4* gene, whereas it occurred in exon 1 for the *COL4A3* gene. Intriguingly, the *COL4A4* gene generated two kinds of transcripts that differed in the 5' ends driven from the two presumptive alternative promoters, as mentioned above. An alternative promoter is also found in the *COL4A6* gene and initiates transcription from different first exons, which results in different signal peptides at the 5' termini [7]. As the two transcripts of *COL4A4* differed only in the 5' untranslated region, the deduced amino acid sequences in $\alpha 4(\text{IV})$ polypeptides would be the same.

Quite a few vertebrate genes have alternative promoters [19]. The structures are different in each case and alternative usage can influence gene expression in diverse ways. The level of transcription initiation can vary between alternative promoters which can have different tissue-specificity and respond to extracellular signals. The shift in promoter usage sometimes causes diseases [19]. Our data suggest that the transcripts containing exon 1 show specific expressions in some epithelial cells, whereas those having exon 1' are ubiquitously expressed. So similar utilization of $\alpha 4(\text{IV})$ promoters may be expected. It is interesting to know whether the change in expression pattern relates to abnormal conditions of basement membrane in some diseases.

3.6. The 5' flanking region of *COL4A3* and *COL4A4*

The 5' flanking region of both genes was composed of GC-rich sequences with no canonical TATA box. The GC-rich region spanned more than 1000 nucleotides with a GC content of 60–80% and contained many CpG dinucleotides, forming a CpG island, in which the first exons of both genes were embedded. The density of CpG dinucleotides was conspicuously higher in exon 1 of *COL4A3* and *COL4A4* compared with exon 1' of *COL4A4*. When compared with promoters of other type IV collagen genes, high content of CpG dinucleotides had been found in the *COL4A1/A2* promoter as the *COL4A3/A4* promoter, while the *COL4A5/A6* promoter contained much less CpG dinucleotides (data not shown). Some other consensus sequence motifs were found such as GC boxes, consensus transcription factor Sp1 binding site, a CCAAT box for *COL4A3* and CTC boxes. Especially, the most conspicuous one was a tandem repeat of 19 nt at –120 to –81 of exon 1. Although how these elements regulate the transcription of both *COL4A3* and *COL4A4* genes is yet to be determined, they seem essential for appropriate transcription, considering their roles in other collagen genes [5,6,20–22]. It is true that the common promoter region of the two genes is not sufficient for the proper expression of both genes as *COL4A1* and *COL4A2* genes, since this region alone is inactive [6,23]. Each gene contains an activator region which activates the promoter either in the *COL4A1* or the *COL4A2* direction [24,26]. There is a silencer element within the third intron of *COL4A2* which influences transcription in both directions [25,26]. Although a head-to-head arrangement

may not be prerequisite for tight and synchronous regulation of two genes, such an arrangement may be beneficial for many functionally related genes.

Acknowledgements: This work was supported in part by Scientific Research Grant, Grant-In-Aid for Exploratory Research 08877098 (to R.M.) and Grant-In-Aid for Scientific Research (B) 08457154 (to Y.N.) from the Ministry of Education, Science, and Culture of Japan, and The Nakatomi Foundation.

References

- [1] Prockop, D.J. and Kivirikko, K.I. (1995) *Annu. Rev. Biochem.* 64, 403–434.
- [2] Fukai, N., Apte, S.S. and Olsen, B.R. (1994) *Methods Enzymol.* 245, 3–28.
- [3] Hudson, B.G., Reeders, S.T. and Tryggvason, K. (1993) *J. Biol. Chem.* 268, 26033–26036.
- [4] Ninomiya, Y., Kagawa, M., Iyama, K., Naito, I., Kishiro, Y., Seyer, J.M., Sugimoto, M., Oohashi, T. and Sado, Y. (1995) *J. Cell Biol.* 130, 1219–1229.
- [5] Burbelo, P.D., Martin, G.R. and Yamada, Y. (1988) *Proc. Natl. Acad. Sci. USA* 85, 9679–9682.
- [6] Pöschl, E., Pollner, R. and Kühn, K. (1988) *EMBO J.* 7, 2687–2695.
- [7] Sugimoto, M., Oohashi, T. and Ninomiya, Y. (1994) *Proc. Natl. Acad. Sci. USA* 91, 11679–11683.
- [8] Kamagata, Y., Mattei, M.-G. and Ninomiya, Y. (1992) *J. Biol. Chem.* 267, 23753–23758.
- [9] Mariyama, M., Zheug, K., Yang-Feng, T.L. and Reeders, S.T. (1992) *Genomics* 13, 809–813.
- [10] Higashi, T., Sasai, H., Suzuki, F., Miyoshi, J., Ohuchi, T., Takai, S. and Kakunaga, T. (1990) *Proc. Natl. Acad. Sci. USA* 87, 2409–2413.
- [11] Ohashi, T., Aizawa, S., Ooka, H., Ohsawa, T., Kaji, K., Kondo, H., Kobayashi, T., Noumura, T., Matsuo, M., Mitsui, Y., Yamamoto, K., Murota, S., Ito, H. and Utakoji, T. (1980) *Exp. Gerontol.* 15, 121–133.
- [12] Kigasawa, K., Souchi, S., Tanaka, Y. and Obazawa, H. (1994) *Jpn. J. Ophthalmol.* 38, 10–15.
- [13] Sugimoto, M., Oohashi, T., Yoshioka, H., Matsuo, N. and Ninomiya, Y. (1993) *FEBS Lett.* 330, 122–128.
- [14] Mariyama, M., Leinonen, A., Mochizuki, T., Tryggvason, K. and Reeders, S.T. (1994) *J. Biol. Chem.* 269, 23013–23017.
- [15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, 2nd Edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [16] Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA* 85, 8998–9002.
- [17] Fend, J.Q., Chen, D., Cooney, A.J., Tsai, M.-J., Harris, M.A., Tsai, S.Y., Feng, M., Mundy, G.R. and Harris, S.E. (1995) *J. Biol. Chem.* 270, 28364–28373.
- [18] Leinonen, A., Mariyama, M., Mochizuki, T., Tryggvason, K. and Reeders, S.T. (1994) *J. Biol. Chem.* 269, 26172–26177.
- [19] Ayoubi, T.A.Y. and van de Van, W.J.M. (1996) *FASEB J.* 10, 453–460.
- [20] Fischer, G., Schmidt, C., Opitz, J., Cully, Z., Kühn, K. and Pöschl, E. (1993) *Biochem. J.* 292, 687–695.
- [21] Heikkilä, P., Soininen, R. and Tryggvason, K. (1993) *J. Biol. Chem.* 268, 24677–24682.
- [22] Schmidt, C., Fischer, G., Kadner, H., Genersch, E., Kühn, K. and Pöschl, E. (1993) *Biochim. Biophys. Acta* 1174, 1–10.
- [23] Kühn, K. (1994) *Matrix Biol.* 14, 439–445.
- [24] Tanaka, S., Kaytes, P. and Kurkinen, M. (1993) *J. Biol. Chem.* 268, 8862–8870.
- [25] Haniel, A., Lüben, U.W., Kühn, K. and Pöschl, E. (1995) *J. Biol. Chem.* 270, 11209–11215.
- [26] Pollner, R., Schmidt, C., Fischer, G., Kühn, K. and Pöschl, E. (1997) *FEBS Lett.* 405, 31–36.