ELSEVIER

Matrix Biology 18 (1999) 261–274

MATRIX BIOLOGY

Complete exon–intron organization and chromosomal location of the gene for mouse type XIII collagen (*col13a1*) and comparison with its human homologue

Ari-Pekka Kvist[a], Anne Latvanlehto[a], Malin Sund[a], Nina Horelli-Kuitunen[b], Marko Rehn[a], Aarno Palotie[b], David Beier[c], Taina Pihlajaniemi[a,*]

[a] *Collagen Research Unit, Biocenter and Department of Medical Biochemistry, University of Oulu, Kajaanintie 52 A, 90220 Oulu, Finland*

[b] *HUCS, Laboratory of Molecular Genetics, Hartmaninkatu 4, 00290 Helsinki, Finland*

[c] *Division of Genetics, Brigham and Women's Hospital, Harvard Medical School Boston, MA 02115, USA*

Received 27 October 1998; accepted 28 February 1999

Abstract

Recent findings indicate that type XIII collagen is a transmembrane protein with a short N-terminal sytocsolic domain, a single transmembrane domain and a large, mainly collagenous ectodomain. The complete exon–intron structure of the gene coding for the mouse α1(XIII) collagen chain, *col13a1*, has now been characterized from genomic clones spanning over 180 kilobases (kb) and shown to be approximately 135 kb in size and to contain 42 exons varying between 8 base pairs (bp), the shortest exon in the genes encoding the various collagens, and 836 bp. Nuclease S1 mapping and 5'RACE resulted in identification of multiple transcription initiation points in the mouse gene, ranging between 470 and 548 bp upstream from the initiation methionine. This is in good agreement with a recently identified human EST clone extending 537 bp upstream from the initiation methionine. The 836-bp first exon of the mouse gene covers both the long 5′ untranslated region and also a 36-residue cytosolic portion, a 23-residue transmembrane domain, and 37 residues of the 60-residue non-collagenous ectodomain immediately adjacent to the plasma membrane. One striking feature of the exons encoding solely collagenous sequences is the abundance of 27-bp exons, half the ancestral 54-bp size characteristic of fibrillar collagen genes, while the others vary between 8 and 144 bp, including instances of 36-, 45- and 54-bp exons. Determination of approximately 2.6 kb of sequences upstream of the initiation methionine of both the mouse and human genes and the identification of a clone containing four exons and spanning a gap in the previously characterized human clones allowed detailed comparison of the two genes. The exon–intron structures were found to be completely conserved between the species, and both genes have their 5′ untranslated region preceded by a highly homologous apparent promoter region of approximately 350 bp containing a modified TATAA motif and several GC boxes. The chromosomal location of the mouse gene was determined by SSCP and fluorescence in situ hybridization and found to be at chromosome 10, band 4, between markers *D10Mit5* -2.3 ± 1.6 cM — *col13a1* -3.4 ± 1.9 cM — *D10Mit15*. This result indicates that the mouse type XIII collagen gene and its human counterpart are located in chromosomal segments with conserved syntenies (The GenBank accession numbers for the mouse gene are AF063666–AF063693. The new GenBank accession number for the 5′ end of the human type XIII collagen gene is AF071009).

Keywords: Chromosomal localization; *Col13a1*; Gene structure

* Corresponding author. Department of Medical Biochemistry, University of Oulu, Kajaanintie 52 A, FIN-90220 Oulu, Finland. Tel.: +358-8-5375-801; fax: +358-8-5375-811.

E-mail address: taina.pihlajaniemi@oulu.fi (T. Pihlajaniemi)

0945-053X/99/$ - see front matter
PII: S0945-053X(99)00018-9

1. Introduction

The collagen family of proteins includes 19 collagen types and more than 30 genes that encode their constitutive α-chains, while several additional proteins have collagen-like domains (Vuorio and de Crombrugghe, 1990; Pihlajaniemi and Rehn, 1995; Prockop and Kivirikko, 1995). The collagens can be divided into two subgroups in terms of their structural and functional characteristics, fibril-forming and non-fibril-forming collagens. Members of the former group, i.e. types I–III, V and XI, aggregate into prominent fibrillar structures in many collagenous tissues. These molecules are structurally homologous and characterized by a long, uninterrupted collagen triple helix, and their genes are also highly homologous. The other collagens are unable to form fibrils and are very variable in structure, macromolecular organization, tissue distribution, function and genomic structure (Sandell and Boyd, 1990; Vuorio and de Crombrugghe, 1990; Li et al., 1991; Bateman et al., 1996).

Type XIII collagen belongs to the non-fibrillar group and has so far been characterized via human and mouse cDNA clones and partially via human genomic clones, but its function is still not known (Pihlajaniemi et al., 1987; Tikka et al., 1988, 1991; Pihlajaniemi and Tamminen, 1990; Pihlajaniemi and Rehn, 1995; Hägg et al., 1998). mRNAs encoding type XIII collagen have been found to occur widely in the human body, albeit in small amounts (Sandberg et al., 1989; Juvonen et al., 1993). The predicted α1(XIII) collagen polypeptide consists of N- and C-terminal non-collagenous domains, termed NC1 and NC4, respectively, and three collagenous domains COL1–COL3, separated by the non-collagenous domains NC2 and NC3. The precursor mRNAs encoding type XIII collagen undergo complex alternative splicing, affecting both the collagenous and non-collagenous sequences (Juvonen and Pihlajaniemi, 1992; Juvonen et al., 1992, 1993; Peltonen et al., 1997). Type XIII collagen is an unusual member of the collagen family in that it has a hydrophobic membrane-spanning segment in its NC1 domain which anchors the molecule to the plasma membrane of the cells that express this protein (Hägg et al., 1998).

Genomic clones encoding most of the human type XIII collagen have been characterized (Tikka et al., 1991). Recent characterization of mouse and human cDNA clones suggests a longer 5′ untranslated region and first exon than had previously been thought to exist (Hägg et al., 1998). In order to study further the promoter and 5′ flanking area and to characterize the complete exon–intron organization, we cloned the entire mouse type XIII collagen gene and studied the 5′ end of both the mouse and human gene, determining the transcription starting site of the mouse gene by S1 nuclease mapping and 5′RACE. To calculate the proper length of this gene, we produced a complete restriction map and sequenced approximately 30 kb around its exon–intron junctions. We also determined the location of this gene in the mouse genome by SSCP analysis and FISH. Four previously uncharacterized exons in the human gene are also described here.

2. Materials and methods

2.1. Isolation and characterization of genomic clones

A mouse genomic library (951303; Stratagene, La Jolla, CA) in the cosmid vector pWE15 was screened with the [α-^{32}P]dCTP-labeled mouse type XIII collagen cDNA clone G2 coding for the central portion of the polypeptide, i.e. exons 23–29 (Hägg et al., 1998). The library was plated as described in the manufacturer's manual and screened under stringent conditions (Sambrook et al., 1989). Hybridizations were carried out at +42°C in 50% (v/v) formamide, 5 × SSC (1 × SSC = 0.15 M NaCl, 0.015 M sodium citrate, pH 6.8), 1% (w/v) bovine serum albumin, 1% (w/v) Ficoll, 1% (w/v) polyvinylpyrrolidone, 0.25 mg of denatured salmon sperm DNA/ml and 0.1% (w/v) sodium dodecyl sulfate (SDS). The final washes for the filters were carried out in 0.5 × SSC and 0.1% SDS at +65°C. A 129 SVJ mouse genomic library (961301; Stratagene) was screened for the 5′ end using as a probe a 5′ PCR fragment generated with oligonucleotides 5′CCGGCTGTGTCAAGTTACAG and 5′CCTCTCAGTTTGGATCG to cover nucleotides 1–407 of the mouse cDNA sequence (GenBank accession no. U30292). The same library was also screened for the 3′ end using an oligonucleotide from exon 41 (5′GGAGAAGATGGCTTACCAGTCC) as the probe. Hybridizations were carried out at +37°C in 50% (v/v) formamide, 6 × SSC, 1% (w/v) bovine serum albumin, 1% (w/v) Ficoll, 1% (w/v) polyvinylpyrrolidone, 0.25 mg of denatured salmon sperm DNA/ml and 0.1% (w/v) SDS. Finally the filters were washed in 2 × SSC and 0.1% SDS at +42°C.

P1 clones were obtained from Genome Systems Inc. (St Louis, MO) by PCR screening of P1 libraries with oligonucleotides corresponding to intron sequences adjacent to exons 1 (5′GGCCCTAGGAAAGAGGAGTGCTGTGGTG and 5′GGCCGATTCTGGCGCCTCTGTAACT), 7 (5′GAGGTACCTGCTCACATCCCCTCTGTC and 5′TGAGTCTCACCCTCATTCCTCCAAAGC) and

15 (5′CTATCCTACTGCCATGTTTACACTGACG, 5′GGACCTTTCAGCTGAGGAAAGATAGAA). The 5′ end genomic clone of human type XIII collagen, CL412, was a generous gift from Dr Karl Tryggvason (University of Oulu, Oulu, Finland) (Tikka et al., 1991).

For restriction mapping, inserts released from the cosmid clones by *Not*I digestion as previously described (Evans et al., 1989) and from the P1 clones by *Not*I and *Sfi*I digestions were subjected to both partial and complete restriction enzyme digestion with the *Eco*RI, *Bam*HI, *Hind*III and *Xba*I enzymes and the fragments were fractionated by CHEF-Mapper electrophoresis (Bio-Rad, Hercules, CA) using in-built algorithms for optimal DNA fragment resolution and transferred to nitrocellulose filters (Millipore, Bedford, MA) for Southern hybridization. The partial digestions were analyzed using either T7 and T3 oligonucleotides (cosmid clones) or T7 and SP6 oligonucleotides (P1 clones), and the complete digestions were analyzed using oligonucleotides specific to each exon as probes.

2.2. Nucleotide sequencing and sequence analysis

The intron–exon boundaries were sequenced directly from the cosmid and P1 clones using the ^{T7}Sequencing™ Kit (Amersham Pharmacia Biotech, Uppsala, Sweden) and the Cycle Sequencing Kit (Amersham Pharmacia Biotech) or using an ABI automatic sequencer (Perkin-Elmer, Norwalk, CT). The gene areas that could not be sequenced directly from the P1 clones were examined by subcloning *Hind*III and *Eco*RI fragments to the Bluescript™ (Stratagene) vector and sequencing them with the ^{T7}Sequencing™ Kit (Amersham Pharmacia Biotech) (Table 2). Intron sizes were determined either by sequencing, by PCR using exon-specific oligonucleotides or from the restriction map (Fig. 1, Table 1).

DNASIS (Amersham Pharmacia Biotech) was used to analyze the nucleotide sequence data. Consensus sites for the binding of transcription factors were searched for in the Transcription Factor Database using the GCG Sequence Analysis Software Package, version 8.1 (Genetics Computer Group Inc., Madison, WI) (Devereux et al., 1984). Alignments and homology searches were performed with GCG's GAP, BESTFIT, FASTA and PLOTSIMILARITY programs. The human and mouse promoters were predicted with the PROSCAN (Prestridge, 1995, see http://bimas.dcrt.nih.gov/molbio/proscan/) and TSSG programs (Solovyev V.V., Salamov A.A., Lawrence C.B unpublished data, see http://dot.imgen.bcm.tmc.edu:9331/gene-finder/gf.html). A primer program was used to select the primer pairs (Rozen and Skaletsky, 1997).

2.3. 5′RACE and nuclease S1 protection

To prepare 5′RACE cDNA clones corresponding to the initiation of transcription, blunt-end cDNA was generated using a 5′ untranslated region primer (see below), poly A^+ RNA from an 18.5-day mouse embryo and the Time-Saver cDNA synthesis kit (Amersham Pharmacia Biotech) as described in Rehn and Pihlajaniemi (1995). Next, 0.5 μl of 5 μM linker solution was ligated to the beginning of the blunt-end cDNA and a 10-μl PCR reaction was performed in the presence of *Taq* polymerase buffer (Promega; 50 mM KCl, 1.7 mM $MgCl_2$, 0.1% Triton X-100 and 10 mM Tris–HCl, pH 9.0), 0.2 mM of each deoxynucleotide, 10 pmol of the gene-specific 20-mer primer 5′CCCCACTCTTTCCCTTCTCT, 1 pmol of a 25-mer linker primer and 1 unit of *Taq* polymerase (Promega, Madison, WI). The amplification conditions were 1 min at +94°C, 1 min at +65°C and 1 min at +72°C for 35 cycles. A second round of PCR was carried out using a 1:25 dilution of the first PCR reaction, the linker solution and a nested primer (5′GAATTCAGAGTTCTCGGGCTGGTGCG). The PCR product was then *Eco*RI digested and ligated to the pSP72 vector (Promega), and the mixture was used to transform *E. coli* DH5α cells, which were then plated and screened using a genomic sequence corresponding to the 5′ untranslated region of the mouse type XIII collagen gene as the probe. Positive colonies were picked out and the isolated DNA was sequenced using the ^{T7}Sequencing™ Kit (Amersham Pharmacia Biotech).

S1 nuclease protection was performed as described in Sambrook et al. (1989) and Pihlajaniemi and Myers (1987). A 859-bp PCR fragment generated using the oligonucleotides 5′TAAATAACCCCAAGTGTCACTTCAAAGG and 5′ACACAGCGCTAGAGCCAGCAGTC was used as a probe (open box in Fig. 2), having been 5′ end-labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP (3000 Ci/mmol, Amersham Pharmacia Biotech). The double-stranded probe (7×10^3 cpm) was hybridized to 20 μg of total RNA from mouse lung in the presence of 80% formamide, 40 mM Pipes, pH 6.4, 400 mM Nalco and 1 mM EDTA at +65°C for 18 h. After hybridization, 300 μl of buffer (280 mM Nalco, 50 mM sodium acetate, pH 4.5, and 4.5 mM $ZnSO_4$) was added, and the mixture was digested with 800 units of S1 nuclease (Boehringer Mannheim GmbH, Mannheim, Germany) at room temperature for 20 min. The protected fragments were analyzed on a 6% polyacrylamide sequencing gel. The total RNA from adult mouse lung was a generous gift from Dr Sirkku Peltonen (University of Turku, Turku, Finland) and was isolated by guanidium isothiocyanate–chloroform–phenol extrac-

Table 1
Exon–intron organization of the gene coding for mouse type XIII collagen and its comparison with the human gene[1]

Mouse exon/intron no.	Human exon/intron no.	Domains encoded by the exon	Mouse exon size	Human exon size	Mouse intron size	Human intron size	Features
1	1	NC1	836	1088	18000[a]	ND	Includes transmembrane domain
2	2	NC1/COL1	70	70	40000[a]	ND	Junction exon
3	3a	COL1	8	8	8300[a]	4372[c]	
4	3b	COL1	27	27	3000	2917[c]	Alternatively spliced (hu)
5	4a	COL1	36	36	3500	2865[c]	Alternatively spliced (hu/mo)
6	4b	COL1	27	27	2100	2436[c]	Alternatively spliced (hu/mo)
7	5	COL1	51	51	5500[a]	5500	Alternatively spliced (hu/mo), interruption
8	6	COL1	36	36	850	2100	Alternatively spliced (mo)
9	7	COL1	27	27	1000	2600	
10	8	COL1	27	27	1243[b]	5300	Alternatively spliced (mo)
11	9	COL1	27	27	90[b]	60	
12	10	COL1	27	27	937[b]	1500	
13	11	COL1/NC2	27	27	1693[b]	4100	Junction exon
14	12	NC2	57	57	1061[b]	800	Alternatively spliced (hu/mo)
15	13	NC2	66	66	2713[b]	2300	Alternatively spliced (hu/mo)
16	14	NC2/COL2	108	108	1800	2800	Junction exon
17	15	COL2	27	27	950	800	
18	16	COL2	36	36	8300[a]	7800	
19	17	COL2	45	45	1500	800	
20	18	COL2	33	33	700	800	Interruption
21	19	COL2	27	27	2800	3300	
22	20	COL2	45	54	1039[b]	1000	
23	21	COL2	87	87	583[b]	1200	Interruption
24	22	COL2	54	54	1643[b]	1200	
25	23	COL2	45	45	311[b]	1000	
26	24	COL2/NC3	69	69	1041[b]	1300	Junction exon
27	25	NC3	24	24	2300	2300	
28	26	COL3	63	63	855[b]	1000	
29	27	COL3	45	45	273[b]	400	
30	28a	COL3	144	153	515[b]	2100	
31		COL3	99	ND	1600	ND	Alternatively spliced (mo)
32	29	COL3	45	45	2700	2900	Alternatively spliced (hu)
33	30	COL3	42	42	1000	1600	Alternatively spliced (hu), interruption
34	31	COL3	27	27	400	1000	
35	32	COL3	81	81	2700	2300	Alternatively spliced (mo)
36	33	COL3	36	36	1600	1600	Alternatively spliced (mo/hu)
37	34	COL3	54	54	1100	4200	
38	35	COL3	54	54	1100	3900	
39	36	COL3	36	36	6200[a]	7900	
40	37	COL3/NC4	87	87	1100	5100	Alternatively spliced (hu), junction exon
41	38	NC4	39	39	1400	1700	
42	39	TER/3'UT	211	> 150			Termination signal

[1] *Note:* Based on the further characterization of the human gene performed here, it is proposed that exons 3a, 3b, 4a and 4b should be named 3–6, respectively. Furthermore, it is likely that mouse exon 31 also exists in the human gene. We thus propose that the numbering presented for the mouse gene should be adopted for the human gene as well. The size of the first exon of the mouse gene is calculated from the 5'RACE clones and includes a 288-bp coding region, while the size of this exon in the human gene is calculated based on the human EST clone (see text) and includes a 294-bp coding region. The last exon of the mouse gene contains a 3-bp termination codon and a 208-bp 3' untranslated region (3'UTR) while the human 3'UTR is over 150 bp in length.

[a] Intron lengths were assigned by PCR of genomic clones, except in some cases where the lengths were determined by restriction mapping.

[b] Intron lengths were assigned by PCR of genomic clones, except in some cases where the lengths were determined by nucleotide sequencing.

[c] Human intron sizes were determined from Tikka et al. (1991) and from GenBank accession no. U82211. All sizes are marked as base pairs. The sequences with alternative splicing in the human and mouse genes are indicated by hu and mo, respectively.

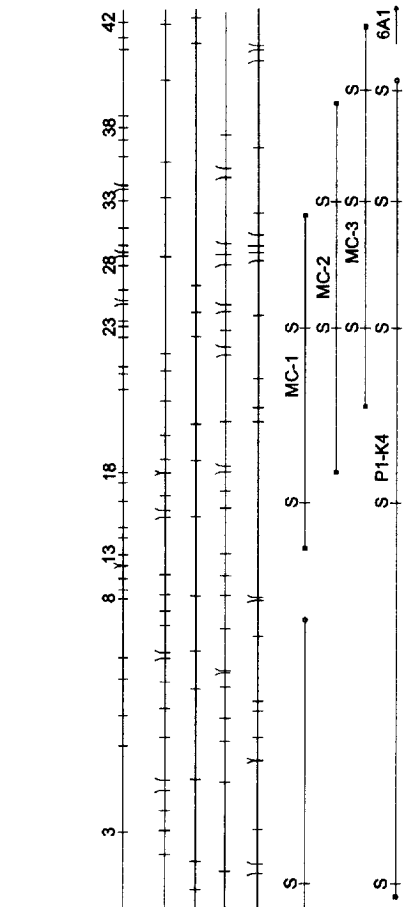

Fig. 1. Organization of the gene and isolated genomic clones encoding the mouse type XIII collagen α1 chain. The locations of the 42 exons are indicated by vertical bars and the intervening sequences by horizontal lines. The restriction sites for *Bam*HI, *Eco*RI, *Hind*III and *Xba*I are indicated with vertical bars. The genomic area is covered by cosmid and P1 clones, which include internal *Sfi*I (S) and *Not*I (N) restriction sites. The flanking T7, T3 and Sp6 oligonucleotide sequences are indicated by black squares, open squares and open circles, respectively. The orientation of clone 6A1 is not determined and the arrow indicates that the clone extends beyond the figure. The scale is marked with a bar indicating the length of 10 kb (kilobases).

tion (Sambrook et al., 1989). Yeast tRNA (20 μg) was used as a negative control.

2.4. Northern blotting

A mouse Multi-Tissue Northern blot (Clontech Laboratories, Inc. Palo Alto, CA) prepared by gel electrophoresis of 2 μg/lane of poly A^+ RNA isolated from various adult mouse tissues was hybridized with a α-^{32}P-labeled 756-bp *Pvu*II fragment of mouse type XIII cDNA with ExpressHyb™ (Clontech) hybridization solution according to protocols from the manufacturer's manual (Yang and Kain, 1995).

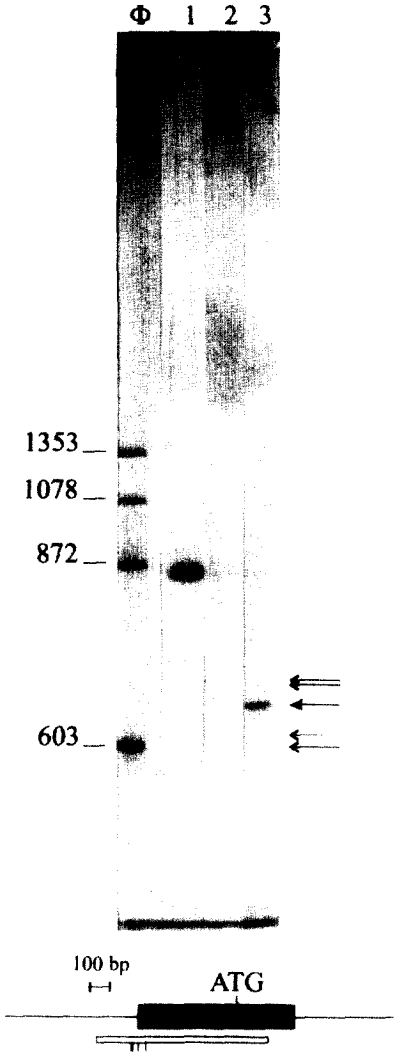

Fig. 2. Nuclease S1 mapping analysis to locate the transcription initiation site of the mouse type XIII collagen gene. A 859-bp PCR fragment of the gene was 5′-end-labeled and used for nuclease S1 digestion, and an autography of the digestion products fractionated by gel electrophoresis is shown. The probe is shown by an open bar below the corresponding genomic region, with exon 1 shown as a black bar. The lengths of the protected fragments ranged between 470 and 525 bp. The main protected fragment is marked with a filled arrow and several faint ones with open arrows. Lane 1, undigested probe; lane 2, probe with tRNA and S1-digestion; lane 3, S1-protected fragments obtained using the probe and mouse lung total RNA.

2.5. Determination of chromosomal location by single strand conformational polymorphism

Primers were designed to amplify a region corresponding to the intron 25 sequences of *col13a1* in order to test for single strand conformation polymorphisms (SSCPs) between the mouse strains. These were analyzed as described previously (Beier, 1993). Briefly, oligonucleotides were radiolabeled with [γ-^{32}P]ATP using polynucleotide kinase, and genomic DNAs from a series of mouse strains were amplified following standard protocols (denature at +94°C for 1 min, anneal at +55°C for 1 min and extend at +72°C for 2 min for 40 cycles, with a final extension at +72°C, 10 min). Of the amplified reaction material 2 μl was added to 8.5 μl USB (Amersham Pharmacia Biotech) stop solution, denatured at +94°C for 5 min and immediately placed on ice. Of each reaction 2 μl was loaded onto a 6% non-denaturing acrylamide sequencing gel and electrophoresed in a 0.05-M Tris–Borate and 1 mM EDTA, pH 8.3, buffer for 2–3 h at 40 W in a +4°C environment. A pair of primers with the sequences 5′TACGCGCTCCGTCTGAGCTCAAGGT and 5′AGGGGTGGCTGTCACCAAAAGGGTT were used to identify a series of polymorphisms between the C57BL/6J and *M. spretus*, which were then used to analyze DNA prepared from the BSS backcross (Rowe et al., 1994). The strain distribution pattern was analyzed using the Map Manager Program (Manly, 1993).

2.6. Fluorescence in situ hybridization

The mouse connective tissue fibroblast cell culture line L-929 (DSM, Germany) was used as a source of metaphase chromosomes. Cells were cultured according to standard protocols and treated with 5-bromodeoxyuridine (BrdU) to induce a banding pattern (Lemieux et al., 1992; Matsuda et al., 1992). Metaphase slides were pre-treated with RNAse (100 μg/ml) and pepsin (20 μg/ml). To produce a sharp banding pattern and to avoid unspecific hybridization, the slides were post-fixed with 1% formamide for 5 min, stained with Hoechst 33258 (1 μg/ml) and exposed to UV light (302 nm) for 30 min (Tenhunen et al., 1995).

Two genomic clones specific to the mouse *col13a1* gene (19A, 6A1, see Fig. 1) were labeled by nick translation with biotin 11-dUTP (Sigma Chemicals, St Louis, MO) according to standard protocols (Nick translation Reagent Kit, BRL, Gaithesburg, MD, USA) and the FISH procedure was carried out in 50% formamide and 10% dextran sulfate in 2 × SSC,

as described earlier (Lichter et al., 1988; Pinkel et al., 1988; Tenhunen et al., 1995). Repetitive sequences were suppressed with a 10-fold excess of Cot-1-DNA (BRL). After overnight incubation, unspecific hybridization signals were eliminated by washing the slides with 50% formamide/2 × SSC, 1 × 2 × SSC, 1 × 1 × SSC and 1 × 0.5 × SSC at +45°C. Specific hybridization signals were visualized using FITC-conjugated avidin (Vector Laboratories, Burlingame, CA) and the slides were counterstained with DAPI (4′-6′-diamino-2-phenylindole) (0.025 μg/ml). Multicolor digital image analysis was used for acquisition, display and quantification of the hybridization signals from the metaphase chromosomes. The system comprised a Photometrics PXL camera (Photometrics Inc., Tucson, AZ) attached to a PowerMac7100/Av workstation, IPLab software controls for camera operation, image acquisition and a Ludl Filter wheel (Heiskanen et al., 1996).

3. Results and discussion

3.1. Isolation of genomic clones encoding the mouse type XIII collagen gene

A total of seven overlapping genomic clones were obtained, covering the entire mouse type XIII collagen gene (Fig. 1). The first three clones (MC1–MC3) were identified by screening a genomic cosmid library with a mouse cDNA covering exons 23–29. These clones covered exons 14–41 of the type XIII collagen gene and extended over 50 kb. A cDNA-derived PCR fragment was used as a probe to screen for the 5′ end of the gene and resulted in the isolation of clone 19A, which contains the first exon, part of the intron 1 and the 5′ flanking sequences. The 3′ end was obtained by screening the same library with an exon 41-specific oligonucleotide. This clone, 6A1, included exons 41–42 and the 3′ flanking sequences. The rest of the type XIII collagen gene was obtained using a commercial screening service. Clone P1–K1, obtained by means of oligonucleotides from the 1st exon, covered exons 1–7, and clone P1–K4 was obtained by screening with oligonucleotides detecting exons 7 and 15. All in all, these clones cover over 180 kb of the mouse genomic DNA, including 140 kb corresponding to the type XIII collagen gene.

3.2. Restriction mapping by pulse field electrophoresis

Inserts were released from the cosmid clones by *Not*I digestion and from the P1 clones by *Not*I and *Sfi*I digestions, after which the DNAs were partly digested with the restriction enzymes *Bam*HI, *Eco*RI, *Hind*III or *Xba*I. The latter four enzymes were chosen because of the slightly different expected restriction fragment sizes. The partially digested fragments were separated by pulse field electrophoresis to achieve good resolution of the large fragments. The size and order of the fragments were ascertained by Southern blotting and hybridization with the insert-end specific oligonucleotides T7, T3 or SP6. The restriction fragment sizes were confirmed by complete digestion of the released inserts, and the final restriction map was achieved by subcloning and sequencing the difficult areas of DNA. The approximate exon locations were determined by Southern blotting and hybridization of complete digestions with exon-specific oligonucleotides. More accurate intron sizes were determined by amplifying the intron areas by PCR using exon-specific primers and separating the products by electrophoresis. Introns 1, 2, 3, 7, 18 and 39 were too large to be amplified by PCR, so that their sizes were determined from the restriction map alone.

The restriction map of the mouse type XIII collagen genomic area, with the four commonly cutting enzymes and the two rarely cutting ones, is shown in Fig. 1. *Not*I and *Sfi*I restriction sites of these rare cutting enzymes were identified unexpectedly close to the non-collagenous sequences encoding exon 1 (Fig. 1). The mouse type XIII collagen gene includes six *Sfi*I restriction sites, three of which were in exon areas (Fig. 1), which is not surprising in the case of a collagenous sequence, since every third amino acid is glycine, which provides a basis for the *Sfi*I recognition sequence GGCCNNNNNGGCC. The restriction map obtained here could be used as a tool in the future to produce various classic, knock-out and deletion-type transgenic mouse constructs.

3.3. Exon–intron structure of the mouse and human type XIII collagen genes

Nucleotide sequences were determined for all the exons, their intron junctions and intronic sequences of reasonable size, showing that the coding information in the mouse type XIII collagen gene is distributed among 42 exons that range in size between 8 and 836 bp, while the lengths of the introns vary between 90 bp and 40 kb (Table 1). The exon–intron boundaries conform well to the consensus sequence AG–exon–GT (Table 2). The human genomic clones described by Tikka et al. (1991) did not cover exons 3a–4b and were deduced by characterization of alternatively spliced cDNA clones (Juvonen and Pihlajaniemi, 1992; Juvonen et al., 1992), whereas scanning of the GenBank data bank for sequence similarities revealed one sequence identified in the course of sequencing the human genome, GenBank accession number U82211, which contained the human type XIII collagen gene region covering exons 3a–5 (Be-

nson et al., 1998). By analogy with the mouse gene, it is suggested that the human exons 3a–5 should be numbered 3–7, respectively (Table 1). The 99-bp mouse exon 31 has not been reported in the human species but is likely to exist there, too. Thus this new numbering of the mouse gene exons may be proposed as a replacement for the nomenclature used previously for the human gene (Tikka et al., 1991; Juvonen et al., 1992). The 8-bp exon 3 found both in the mouse clones and in the human high throughput genomic sequencing clone is the shortest exon known in collagen genes to date, superseding the 11 bp exon found in both the human and chicken gene for the α2 chain of type I collagen (Sandell and Boyd, 1990).

The coding information in the mouse type XIII collagen gene is unevenly distributed, since the 5′ half of the genomic sequences contains only the first two exons (Fig. 1), one of which, exon 1, is the longest of all, and the two longest introns, whereas the other half contains exons 3–42, separated by introns that

Table 2
Exon–intron junctions in the type XIII collagen gene. The 5′ end of exon 1 is not shown here since it represents the 5′ end of the corresponding mRNA, which was detected by S1 nuclease assay and by 5′RACE (Figs. 2 and 3)

Exon no.	Exon–intron boundaries			Exon no.	Exon–intron boundaries		
1	...	ATG GTG..GAC GAG M V D E	gtttgtgcc	22	ctcctgcag	GGT GAG..CAG AAG G E Q K	gtaggtgga
2	tctgtgcag	AAA TGG.CCA CCA G K W P P	gtgagttgg	23	ccccacgag	GGA GAA..CCA AAG G E P K	gtgagtttt
3	ctctttcag	GA CCT CCT (G) W P	gtaagtaca	24	tcctcccag	GA GAA..GAC AAG G E P K	gtagagtcc
4	ttctttcag	GGT CCC..CTC CCA G P L P	gtaagtaac	25	atgcaacag	GGG CAG..CCC AAG G Q P K	gtacgttca
5	cttcatcag	GGG GAC..CGT GTG G A R V	gtgagttgg	26	gctttctag	GGT GCC..CTT CAG G A L Q	gtgagtggg
6	cctgtccag	GGT ATC..GAG AAG G I E K	gtgagtctt	27	tgttttcag	GAG ATC..TTG ATG E I L M	gtgagtcac
7	tccccacag	GGT GCC..CCC CCT G A P P	gtaagttgc	28	ccttttag	GGG CCT..CAA AGG G P Q R	gtaaagctt
8	tcttttcag	GGC CAG..TTC CCG G Q F P	gtaagtgca	29	ctgctgcag	GGG GAG..GAC AAG G E D K	gtgagcaca
9	tatttgcag	GGT CCC..AGG CCG G P R P	gtaagtgga	30	tttctatag	GGG CCT..CCC CAG G P C Q	gtaagtgtc
10	tttcttcag	GGC CAC..GAG ATG G H E M	gtaagcccc	31	aatctatag	GGT CTA..TTC ACA G L F T	gtaaggcga
11	tctcccaca	GGC CTG..CAA CCG G L Q P	gtttgtaac	32	attccacag	GGA GAG..GCA GAG G E A E	gtaagtcaa
12	tctttgtag	GGA CCT..GAA AAG G P Q K	gtaagtgcc	33	ggcttctag	GTT CCT..CCT CCG V P P P	gtaagtatc
13	ccccctag	GGT CAG..CAC AGG G E H R	gtaagtgat	34	ccaacacag	GGC CTC..CCA AAG G L P K	gtaaggaaa
14	ccaaaccag	GAG TAC..CTA AAG E Y L K	gtttgtcgg	35	tccatgcag	GGA GAA..TTG CCT G E L P	gtaagtacc
15	tttcctcag	CTG TTG..TTT CAG L L F Q	gtagatacc	36	ctgagcgag	GGA GCT..CCA CCG G A P P	gtgagtgtc
16	caacttcag	GGG GAA..CCA CCT G E P P	gtaagtatc	37	cctccccag	GGT ACT..GAA AGG G T E R	gtgagtgct
17	cttccacag	GGT TTA..CCA AAG G L P K	gtaagtgct	38	ctctttcag	GGC AGC..CTG CCT G S L P	gtgagtcta
18	tccctgcag	GGA GAC..CGC AAG G D R K	gtaagcagg	39	ttaccgcag	GGT TTA..AAC CGG G L N R	gtaagtctg
19	attttccag	GGA GAG..GCC AAG G E A K	gtacccatg	40	tgctcacag	GGG GAG..CCG CTG G E P L	gtatgtttc
20	ctctaccag	GGG GTT..ATG AAG G V M K	gtaggtaaa	41	tccatgcag	GGA GAA..AAC AAG G E N K	gtaagattc
21	cccacacag	GGT GAG..ACC AAG G E T K	gtactggtg	42	gttttgcag	TGA ter	

are in most cases approximately 1 kb in length (Table 1). Exons 1 and 2 cover the NC1 domain, exons 3–13 the COL1 domain, exons 14–16 the NC2 domain, exons 17–26 the COL2 domain, exon 27 the NC3 domain, exons 28–40 the COL3 domain and exon 41 the NC4 domain (Table 1). Domain transitions occur within the junction exons 2, 13, 16, 26 and 40, while exon 42 consists only of the termination codon TGA and 3′ untranslated sequences. Exons 3–12, 17–25 and 28–39 code for purely collagenous sequences, their sizes varying from 8 to 144 bp. Most of the exons encoding collagenous sequences are 27 bp in length, and only three of them are of the 54-bp in size characteristic of genes encoding collagenous sequences of fibrillar collagens, while the 36- and 63-bp exons are of a size found in some of the other genes encoding non-fibrillar collagens also (Lozano et al., 1985; Soininen et al., 1989; Hayman et al., 1991; Li et al., 1991; Saitta et al., 1991; Tikka et al., 1991; Wälchli et al., 1992; Christiano et al., 1994; Perälä et al., 1994; Zhou et al., 1994). With the exception of exons 3 and 33, the exons encoding only collagenous sequences begin with a complete codon for glycine, which is characteristic of collagen genes (Table 1). Exon 3 begins with a split codon which codes for glycine, while exon 33 begins with an interruption in the collagenous sequence where a codon for glycine is missing.

All in all, comparison of the mouse and human type XIII collagen genes in terms of their exon–intron organization indicates that the exon sizes are totally conserved with the exception of exons 22 and 30, where a 9-bp Gly–X–Y-coding motif is lacking in the mouse gene, and exon 1, where there is a 6-bp insertion in the human gene.

3.4. Transcription start site

Recent cloning of mouse type XIII collagen cDNAs and subsequent further characterization of human cDNAs (Hägg et al., 1998) has indicated that the initiation of translation begins at an ATG further upstream than was initially thought (Tikka et al., 1991). Our S1 mapping showed the main initiation of transcription to take place 500 nucleotides upstream from this ATG codon in the case of the mouse gene (Fig. 2), although there are several other minor initiation sites between 470 and 525 bp upstream of the initiation ATG codon. We have also isolated six 5′RACE cDNA clones corresponding to the 5′ end of the mouse mRNAs encoding the first exon of the polypeptide and have found that all of these end at the same position, namely 548 bp before the initiation ATG (Fig. 3). Subsequently, a human cDNA clone (GB accession no. R25685) extending 537 bp in the 5′ direction with respect to the upstream ATG was identified in the course of the EST project (Fig. 3). This justifies further characterization of the human 5′ flanking sequences as well, since the promoter in the human gene was reported to be located in sequences that we now think are part of the 5′ untranslated area (Tikka et al., 1988).

3.5. Comparison of the promoter and 5′ flanking regions of the mouse and human genes

To enable comparison between human and mouse promoter and 5′ flanking areas, 4932 bp of the 5′ flanking areas of the mouse type XIII collagen gene and 2642 bp of those in the human gene were sequenced and compared. Analysis of 2642 bp of both flanking regions shows an overall identity of 58.3%, but this is unevenly distributed, as 77.13% identity was found in the computationally proposed promoter areas (see below) and the 5′ untranslated areas but no significant homologies were detected upstream of the putative promoters. To see how the identity between the 5′ untranslated and 5′ flanking regions of the human and mouse genes is distributed, the optimally aligned sequences were plotted with PLOTSIMILARITY. This showed the identity between these two 5′ regions to be quite low but to increase just before the suggested promoter areas (Fig. 3).

A search for potential promoters with the PROSCAN v.1.6 and TSSG programs pointed at an identical position in both genes (Fig. 3), the predicted direction of the promoter being the same with both programs. More precisely, the PROSCAN program proposed that the mouse promoter must lie between −834 and −584 bp before the initiation ATG in the 5′ direction, whereas TSSG gave a position between −864 and −568. Collectively, these data suggest that the promoter is located between −864 and −568 bp upstream of the initiation ATG. The 5′ flanking region of the gene before the initiation of transcription contains a modified TATA-like box (TAAATAAACAAA) in the position −601 bp before the initiation ATG, which is in agreement with the presence of multiple transcription start sites observed downstream of this site. The extreme 5′ cDNA clone found here starts 53 bp downstream of the proposed TATA box. Furthermore, the 5′ flanking region contains numerous consecutive GC boxes (a total of 20 in the predicted promoter area). The absence of the traditional TATAA motif and the multiple presence of GC motifs render the type XIII collagen promoter similar to the promoters of genes transcribed widely but at low RNA levels in many tissues.

In the case of the human gene, PROSCAN proposed a promoter area from −855 to −605 bp in the 5′ direction with respect to the initiation ATG and TSSG an area from −905 to −607 bp. Taken together,

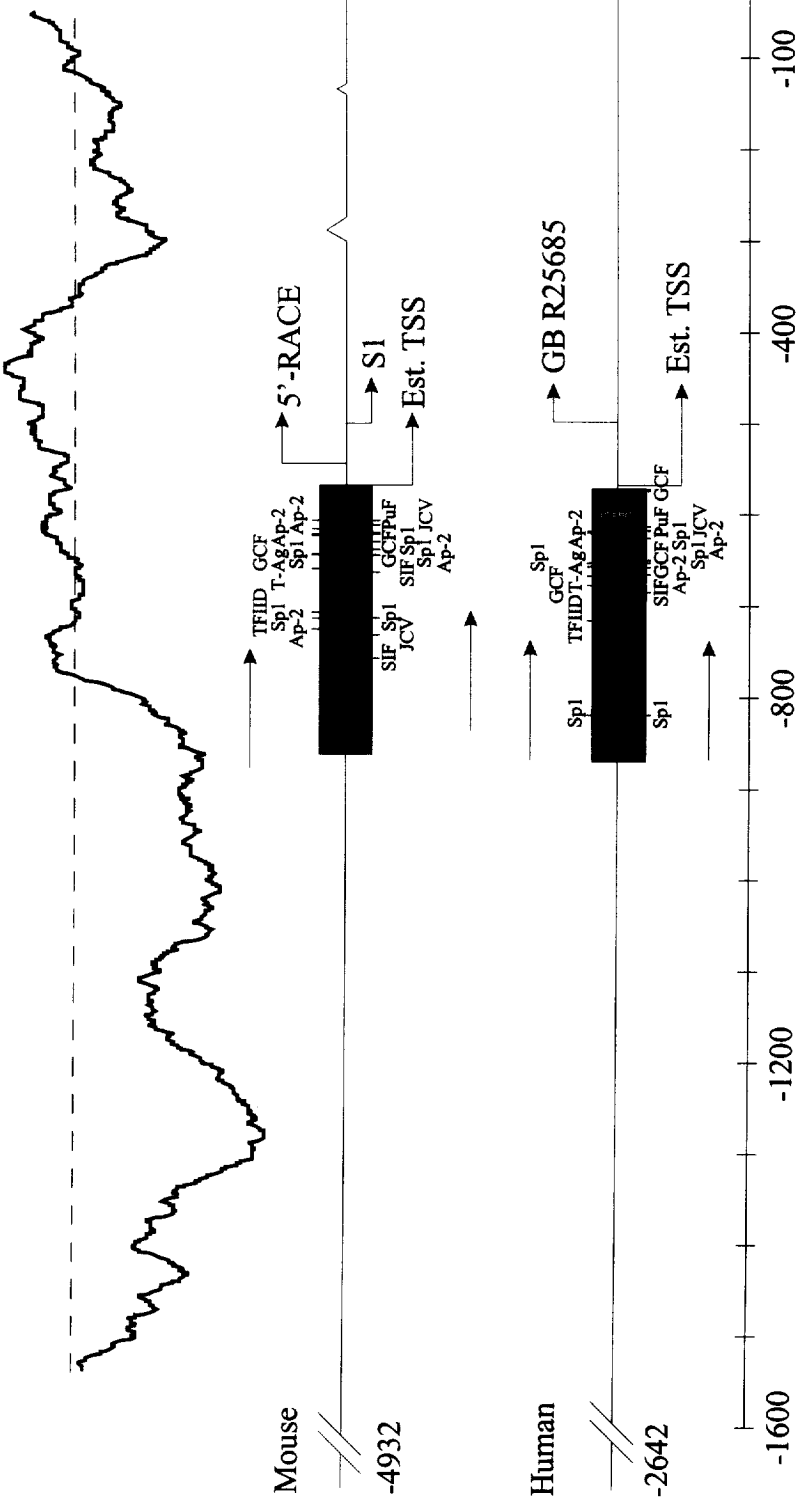

Fig. 3. Comparison of the mouse and human type XIII collagen gene sequences upstream of the initiation ATG codon. 4932 bp of the mouse gene and 2642 bp of the human gene preceding the initiation methionine were sequenced and compared. Promoter regions (boxed areas) and the directions of transcription were predicted by computer, with the PROSCAN program, shown above the boxes, and the TSSG program, shown below them. Promoters were predicted to occur at identical locations in the two genes. The theoretical transcription initiation sites (Est. TSS) for both genes, those observed for the mouse gene by 5′RACE and S1 mapping and the starting point for the human EST clone (GB accession no. R25685) are indicated with arrows. The transcription factor recognition sequences, to be found in both promoters, are marked. These sites are marked above the promoters when they occur in the plus strand and below the box when in the minus strand. TATA-like sequences are marked with light boxes. The curve at the top of the figure shows the identity between the mouse and human sequence. The dashed line indicates 70% identity when the comparison was performed with a window size of 100 bp. The scale is indicated with the first nucleotide of the codon for the initiation methionine as +1 and the upstream sequences as negative.

these analyses suggest that the promoter in the human type XIII collagen gene lies between −905 and −605 bp in the 5′ direction from the initiation ATG. Both programs proposed a TATA box at −636 bp. The difference in the positions of the TATA motifs between the mouse and human genes can be explained by two gaps present in the mouse 5′UTR sequence (Fig. 3).

Comparison of these regions with the transcription factor data base (TFDB) leads to the identification of consensus motifs for several transcription factors. Those that are identical between the two species are shown in Fig. 3, although their function remains to be determined.

3.6. Chromosomal localization

Single strand conformation polymorphism (SSCP) analysis was used to map *col13a1*, as described previously (Beier et al., 1992; Beier, 1993). Primers corresponding to intron 25 of *col13a1* were analyzed and found to identify a SSCP between inbred mouse strains (see Section 2). The BSS interspecific backcross was genotyped and the strain distribution pattern analyzed using the Map Manager program. *Col13a1* was found to map to chromosome 10 with a LOD likelihood score of 25. The position of *col13a1* with respect to the flanking microsatellite markers is: *D10Mit5* −2.3 ± 1.6 cM — *col13a1* − 3.4 ± 1.9 cM — *D10Mit15*. Thus SSCP analysis localized the *col13a1* to mouse chromosome 10, tightly linked to the microsatellite marker *D10Mit5*.

As an alternative approach, the location of the mouse *col13a1* gene was also determined by FISH. Two specific genomic probes, 19A and 6A1, were hybridized on metaphase chromosomes derived from the mouse connective tissue fibroblast cell culture. The identification of the mouse chromosomes was based on their G-banding pattern (Cowell, 1984). Forty-two out of 48 metaphase spreads from the 19A hybridizations and 27 out of 30 from the 6A1 hybridizations showed a specific location on chromosome 10, band B4 (Fig. 4).

4. Conclusions

The gene encoding the mouse type XIII collagen is approximately 135 kb in size and the ensuing mRNAs are approximately 3000 nucleotides in length. *col13a1* contains 42 exons, which vary in size between 8 bp, the shortest exon in the genes encoding the various collagens, and 836 bp. The coding sequences are unevenly distributed in the gene, as the first two exons and introns encompass the entire 5′ half. These two exons cover the 119-residue non-collagenous NC1 domain and the beginning of the COL1 domain. Particularly striking is the first exon, which encodes an extensive 5′ untranslated region of at least 548-bp and the 96 extreme N-terminal amino acid residues of the

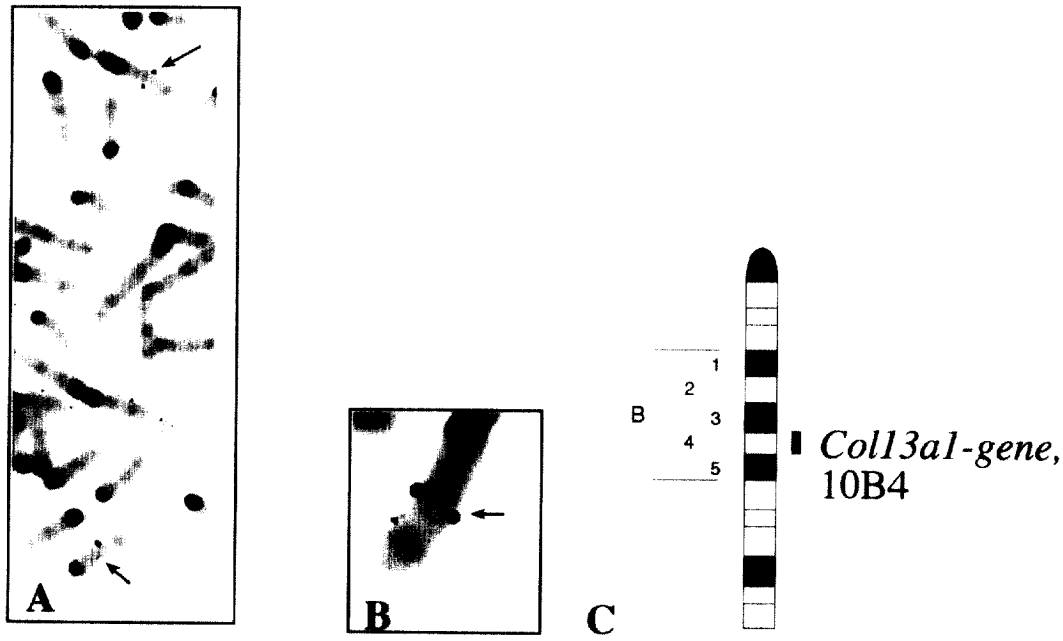

Fig. 4. (a) A partial mouse metaphase spread showing specific hybridization signals for the biotin-labelled cos 19B probe on two chromosomes. (b) A Grayscale image of the specific *col13α1* probe (cos19B) located on mouse chromosome 10, band B4. (c) An idiogram of the mouse chromosome 10 showing location of the type XIII collagen gene.

NC1 domain, including a 36-residue cytosolic portion, a 23-residue transmembrane domain and 37 residues of the non-collagenous ectodomain immediately adjacent to the plasma membrane. The exons of this gene solely encoding collagenous sequences show both features characteristic of fibrillar collagens, in which the triple-helix is encoded predominantly by exons of 54 bp or derivatives of this, and the presence of exons of more variable size, as is found in non-fibril-forming collagen genes (for reviews, see Sandell and Boyd, 1990; Vuorio and de Crombrugghe, 1990). Particularly notable in *col13a1* is the abundance of 27-bp exons, half of the proposed ancestral 54-bp size, as these comprise nearly one-third of the 31 exons solely encoding collagenous sequences. Five of the others are of 36 bp, another five of 45 bp and three of 54 bp, while the remainder vary between 8 and 144 bp. The last exon begins with the stop codon for translation.

A gap in the previously characterized clones encoding the human type XIII collagen (Tikka et al., 1991) was covered by a clone identified when sequencing the human genome and found to contain four exons. All in all, comparison of the mouse and human type XIII collagen genes indicates complete conservation of exon–intron structures between the two species, excluding the exons 1, 22 and 30. Recent characterization of mouse and human cDNA clones (Hägg et al., 1998) predicted a longer N-terminal non-collagenous NC1 domain than had previously been thought (Pihlajaniemi and Tamminen, 1990), which prompted us to study further the 5′ flanking sequences of the mouse and human type XIII collagen genes. 2642 bp of sequences preceding the initiation methionine for this extended NC1 domain were compared between the two genes and their homology was found to be low with the exception of the 5′ untranslated region of approximately 550 bp and an adjacent apparent promoter region of approximately 350 bp containing a modified TATAA motif, several GC boxes and other conserved putative *cis*-acting elements. Previous work on the human gene (Tikka et al., 1991) has suggested that the promoter lies within the sequences that we show here to represent 5′ untranslated sequences, and thus the data point to a revised understanding of the promoter and 5′ untranslated sequences of the human type XIII collagen gene.

The gene encoding mouse type XIII collagen was found to be located in band B4, at *D10Mit5* − 2.3 + 1.6 cM — *col13a1* — 3.4 + 1.9 cM − *D10Mit15* in chromosome 10. According to the consensus genetic map data maintained in the Mouse Genome Database at the Jackson Laboratory (http://www.informatics.jax.org), sub-chromosomal linkage relationships in this region have been conserved on the human chromosome 10q21–22. This result is consistent with the evidence that the human homologue of *col13a1* has been mapped to chromosome 10q22 (Shows et al., 1989). Interestingly, the gene encoding the catalytically active α subunit of prolyl 4-hydroxylase, the key enzyme of collagen biosynthesis, is located approximately 550 kb from the type XIII collagen gene in the human genome (Horelli-Kuitunen et al., 1997). A few cloned mouse genes have been mapped between the *D10Mit5* and *D10Mit15* markers, namely HK1, Prkacn2, Skn2, Gp49b and Nodal (Arora et al., 1990; Zhou et al., 1993; Castells et al., 1994; Rees et al., 1994; Scarpetta et al., 1996). Furthermore, genetic linkage mapping has resulted in the assignment of two additional phenotypes, namely a kidney disease (kd) and valzer (v), a mutation causing hearing problems and circling behavior (Lyon and Hulse, 1971; Bryda et al., 1997). Due to the lack of sequence data on these disorders, we do not know their possible relation to type XIII collagen. All in all, the genomic area where the type XIII collagen gene is located on chromosome 10 is only partly known and is likely to contain several other genes in addition to those indicated above.

Acknowledgements

We thank Mervi Eeva, Aila Jokinen, Maija Seppänen, Sirkka Vilmi and Jaana Väisänen for their technical assistance. This work was supported by grants from the Medical Research Council of the Academy of Finland, Fibrogen Inc., South San Francisco, CA, the Sigrid Juselius Foundation, the Emil Aaltonen Foundation, RO1 HD29028 from NICHD, and RO1 HG00951 from the NCHGR.

References

Arora, K.K., Fanciulli, M., Pedersen, P.L., 1990. Glucose phosphorylation in tumor cells. Cloning, sequencing, and overexpression in active form of a full-length cDNA encoding a mitochondrial bindable form of hexokinase. J. Biol. Chem. 265, 6481–6488.

Bateman, J.F., Lamandé, S.R., Ramshaw, J.A.M., 1996. Collagen superfamily. In: Comper, W.D. (Ed.), Extracellular Matrix. Harwood Academic Publishers, Amsterdam, The Netherlands, pp. 22–67.

Beier, D.R., Dushkin, H., Sussman, D.J., 1992. Mapping genes in the mouse using single-strand conformation polymorphism analysis of recombinant inbred strains and interspecific crosses. Proc. Natl. Acad. Sci. USA 89, 9102–9106.

Beier, D.R., 1993. Single-strand conformation polymorphism (SSCP) analysis as a tool for genetic mapping. Mamm. Genome. 4, 627–631.

Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F., 1998. GenBank. Nucleic. Acids Res. 26, 1–7.

Bryda, E.C., Ling, H., Flaherty, L., 1997. A high-resolution genetic map around waltzer on mouse chromosome 10 and identification of a new allele of waltzer. Mamm. Genome. 8, 1–4.

Castells, M.C., Wu, X., Arm, J.P., Austen, K.F., Katz, H.R., 1994. Cloning of the gp49B gene of the immunoglobulin superfamily and demonstration that one of its two products is an early-ex-

pressed mast cell surface protein originally described as gp49. J. Biol. Chem. 269, 8393–8401.

Christiano, A.M., Hoffman, G.G., Chung-Honet, L.C. et al., 1994. Structural organization of the human type VII collagen gene (COL7A1), composed of more exons than any previously characterized gene. Genomics 21, 169–179.

Cowell, J.K., 1984. A photographic representation of the variability in the G-banded structure of the chromosomes in the mouse karyotype. A guide to the identification of the individual chromosomes. Chromosoma (Berl.) 89, 294–320.

Devereux, J., Haeberli, P., Smithies, O., 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.

Evans, G.A., Lewis, K., Rothenberg, B.E., 1989. High efficiency vectors for cosmid microcloning and genomic analysis. Gene 79, 9–20.

Hägg, P., Rehn, M., Huhtala, P., Väisanen, T., Tamminen, M., Pihlajaniemi, T., 1998. Type XIII collagen is identified as a plasma membrane protein. J. Biol. Chem. 273, 15590–15597.

Hayman, A.R., Koppel, J., Trueb, B., 1991. Complete structure of the chicken α2(VI) collagen gene. Eur. J. Biochem. 197, 177–184.

Heiskanen, M., Kallioniemi, O., Palotie, A., 1996. Fiber-FISH: experiences and a refined protocol. Genet. Anal. 12, 179–184.

Horelli-Kuitunen, N., Kvist, A.P., Helaakoski, T., Kivirikko, K., Pihlajaniemi, T., Palotie, A., 1997. The order and transcriptional orientation of the human COL13A1 and P4HA genes on chromosome 10 long arm determined by high-resolution FISH. Genomics 46, 299–302.

Juvonen, M., Pihlajaniemi, T., 1992. Characterization of the spectrum of alternative splicing of α1(XIII) collagen transcripts in HT-1080 cells and calvarial tissue resulted in identification of two previously unidentified alternatively spliced sequences, one previously unidentified exon, and nine new mRNA variants. J. Biol. Chem. 267, 24693–24699.

Juvonen, M., Sandberg, M., Pihlajaniemi, T., 1992. Patterns of expression of the six alternatively spliced exons affecting the structures of the COL1 and NC2 domains of the α1(XIII) collagen chain in human tissues and cell lines. J. Biol. Chem. 267, 24700–24707.

Juvonen, M., Pihlajaniemi, T., Autio-Harmainen, H., 1993. Location and alternative splicing of type XIII collagen RNA in the early human placenta. Lab. Invest. 69, 541–551.

Lemieux, N., Dutrillaux, B., Viegas-Pequignot, E., 1992. A simple method for simultaneous R- or G-banding and fluorescence in situ hybridization of small single-copy genes. Cytogenet. Cell Genet. 59, 311–312.

Li, K., Sawamura, D., Giudice, G.J. et al., 1991. Genomic organization of collagenous domains and chromosomal assignment of human 180-kDa bullous pemphigoid antigen-2, a novel collagen of stratified squamous epithelium. J. Biol. Chem. 266, 24064–24069.

Lichter, P., Cremer, T., Tang, C.J., Watkins, P.C., Manuelidis, L., Ward, D.C., 1988. Rapid detection of human chromosome 21 aberrations by in situ hybridization. Proc. Natl. Acad. Sci. USA 85, 9664–9668.

Lozano, G., Ninomiya, Y., Thompson, H., Olsen, B.R., 1985. A distinct class of vertebrate collagen genes encodes chicken type IX collagen polypeptides. Proc. Natl. Acad. Sci. USA 82, 4050–4054.

Lyon, M.F., Hulse, E.V., 1971. An inherited kidney disease of mice resembling human nephronophthisis. J. Med. Genet. 8, 41–48.

Manly, K.F., 1993. A Macintosh program for storage and analysis of experimental genetic mapping data. Mamm. Genome. 4, 303–313.

Matsuda, Y., Harada, Y.N., Natsuume-Sakai, S., Lee, K., Shiomi, T., Chapman, V.M., 1992. Location of the mouse complement factor H gene (cfh) by FISH analysis and replication R-banding. Cytogenet. Cell Genet. 61, 282–285.

Peltonen, S., Rehn, M., Pihlajaniemi, T., 1997. Alternative splicing of mouse α1(XIII) collagen RNAs results in at least 17 different transcripts, predicting α1(XIII) collagen chains with length varying between 651 and 710 amino acid residues. DNA Cell Biol. 16, 227–234.

Perälä, M., Elima, K., Metsäranta, M., Rosati, R., de Crombrugghe, B., Vuorio, E., 1994. The exon structure of the mouse α2(IX) collagen gene shows unexpected divergence from the chick gene. J. Biol. Chem. 269, 5064–5071.

Pihlajaniemi, T., Myers, J.C., 1987. Characterization of a pro-α2(I) collagen gene mutation by nuclease S1 mapping. Meth. Enzymol. 145, 213–222.

Pihlajaniemi, T., Rehn, M., 1995. Two new collagen subgroups: Membrane associated collagens and types XV and XVIII. Prog. Nucleic Acid Res. Mol. Biol. 50, 225–262.

Pihlajaniemi, T., Tamminen, M., 1990. The α1 chain of type XIII collagen consists of three collagenous and four noncollagenous domains, and its primary transcript undergoes complex alternative splicing. J. Biol. Chem. 265, 16922–16928.

Pihlajaniemi, T., Myllylä, R., Seyer, J., Kurkinen, M., Prockop, D.J., 1987. Partial characterization of a low molecular weight human collagen that undergoes alternative splicing. Proc. Natl. Acad. Sci. USA 84, 940–944.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., Gray, J., 1988. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc. Natl. Acad. Sci. USA 85, 9138–9142.

Prestridge, D.S., 1995. Predicting Pol II promoter sequences using transcription factor binding sites. J. Mol. Biol. 249, 923–932.

Prockop, D.J., Kivirikko, K.I., 1995. Collagens: molecular biology, diseases, and potentials for therapy. Annu. Rev. Biochem. 64, 403–434.

Rees, D., Nesbitt, M.N., Goldberg, E.H., 1994. Skn 2 is linked to Myb on chromosome 10 of the mouse. Immunogenetics 39, 363–366.

Rehn, M., Pihlajaniemi, T., 1995. Identification of three N-terminal ends of type XVIII collagen chains and tissue-specific differences in the expression of the corresponding transcripts. The longest form contains a novel motif homologous to rat and Drosophila frizzled proteins. J. Biol. Chem. 270, 4705–4711.

Rowe, L.B., Nadeau, J.H., Turner, R. et al., 1994. Maps from two interspecific backcross DNA panels available as a community genetic mapping resource [published erratum appears in Mamm Genome 1994 July;5(7):463]. Mamm. Genome. 5, 253–274.

Rozen, S., Skaletsky, H.J., 1997. Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

Saitta, B., Wang, Y.-M., Renkart, L. et al., 1991. The exon organization of the triple-helical coding regions of the human α1(VI) and α2(VI) collagen genes is highly similar. Genomics 11, 145–153.

Sambrook, J., Fritch, E.F., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sandberg, M., Tamminen, M., Hirvonen, H., Vuorio, E., Pihlajaniemi, T., 1989. Expression of mRNAs coding for the α1 chain of type XIII collagen in human fetal tissues: comparison with expression of mRNAs for collagen types I, II, and III. J. Cell. Biol. 109, 1371–1379.

Sandell, L.J., Boyd, C.D., 1990. Conserved and divergent sequence and functional elements within collagen genes. In: Sandell, L.J., Boyd, C.D. (Eds.), Extracellular Matrix Genes. Academic Press, Inc, San Diego, pp. 1–56.

Scarpetta, M.A., Uhler, M.D., Meisler, M.H., 1996. The mouse Prkacn2 gene encoding protein kinase A inhibitor 2 is located on proximal chromosome 10. Mamm. Genome. 7, 921–922.

Shows, T.B., Tikka, L., Byers, M.G. et al., 1989. Assignment of the human collagen α1(XIII) chain gene (COL13A1) to the q22 region of chromosome 10. Genomics 5, 128–133.

Soininen, R., Huotari, M., Ganguly, A., Prockop, D.J., Tryggvason, K., 1989. Structural organization of the gene for the α1 chain of human type IV collagen. J. Biol. Chem. 264, 13565–13571.

Tenhunen, K., Laan, M., Manninen, T., Palotie, A., Peltonen, L., Jalanko, A., 1995. Molecular cloning, chromosomal assignment, and expression of the mouse aspartylglucosaminidase gene. Genomics 30, 244–250.

Tikka, L., Elomaa, O., Pihlajaniemi, T., Tryggvason, K., 1991. Human α1(XIII) collagen gene. Multiple forms of the gene transcripts are generated through complex alternative splicing of several short exons. J. Biol. Chem. 266, 17713–17719.

Tikka, L., Pihlajaniemi, T., Henttu, P., Prockop, D.J., Tryggvason, K., 1988. Gene structure for the α1 chain of a human short-chain collagen (type XIII) with alternatively spliced transcripts and translation termination codon at the 5′ end of the last exon. Proc. Natl. Acad. Sci. USA 85, 7491–7495.

Vuorio, E., de Crombrugghe, B., 1990. The family of collagen genes. Annu. Rev. Biochem. 59, 837–872.

Wälchli, C., Koller, E., Trueb, J., Trueb, B., 1992. Structural comparison of the chicken genes for α1(VI) and α2(VI) collagen. Eur. J. Biochem. 205, 583–589.

Yang, T.T., Kain, S.R., 1995. Fast hybridization solution for the detection of immobilized nucleic acids. Biotechniques 18, 498–503.

Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L., Kuehn, M.R., 1993. Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. Nature 361, 543–547.

Zhou, J., Leinonen, A., Tryggvason, K., 1994. Structure of the human type IV collagen COL4A5 gene. J. Biol. Chem. 269, 6608–6614.