The gene structure and organization of mouse PG-Lb, a small chondroitin/dermatan sulphate proteoglycan

Yukiko IWATA*, Tamayuki SHINOMURA*, Kazuhiro KURITA*†, Masahiro ZAKO* and Koji KIMATA*[1]

*Institute for Molecular Science of Medicine, Aichi Medical University, Nagakute, Aichi 480-11, Japan, and †Department of Orthopaedic Surgery, Nagoya University School of Medicine, Showa, Nagoya 466, Japan

PG-Lb was originally characterized as a small chondroitin/dermatan sulphate proteoglycan expressed preferentially in the zones of flattened chondrocytes in developing chick limb cartilage. The occurrence of this proteoglycan in mammalian cartilage has been shown by the isolation of a cDNA clone from mouse cartilage cDNA library [Kurita, Shinomura,Ujita, Zako, Kida, Iwata and Kimata (1996) Biochem. J. **318**, 909–914]. To understand the regulation mechanisms for such a unique expression, we have investigated a genomic DNA structure of the PG-Lb gene. The gene is composed of seven exons and six introns spanning more than 50 kb. The leucine-rich repeats are encoded from exon V to exon VII. The transcription initiation site has been determined by rapid amplification of the cDNA ends ('5′-RACE'). The possible TATA box was detected about 90 bp upstream of the adenosine residue that was numbered as position +1. Further analyses of 1.5 kb of the 5′ flanking region and 2.2 kb of the first intron have revealed several potential binding motifs for transcription factors such as Sox 5 and 9. The presence of those sequences in the PG-Lb gene was discussed in relation to the unique expression of this proteoglycan. The chromosomal localization of the murine PG-Lb gene was determined to be on the mouse chromosome 10 by the fluorescence-*in-situ*-hybridization ('FISH') method.

INTRODUCTION

PG-Lb was for the first time isolated from chick embryonic epiphyseal cartilage [1]. Biochemical analysis showed that this proteoglycan is composed of a core molecule (the core protein plus oligosaccharides) of M_r 43000 and chondroitin/dermatan sulphate chains of M_r 52000 [1]. Subsequent cDNA analyses of chicken PG-Lb [2] revealed the structure of the core protein which was unique but related to other small chondroitin/dermatan sulphate proteoglycans in cartilage such as decorin and biglycan [3,4]. Recently, we have, for the first time, isolated a cDNA clone for mammalian PG-Lb from a cDNA library of newborn-mouse epiphyseal cartilage [5]. The cloned cDNA was 1430 bp long and contained a 966 bp open reading frame encoding the core protein of 322 amino acid residues. The deduced amino acid sequence showed a high identity with chicken PG-Lb (about 62% on the whole, and about 80% on the C-terminal two-thirds, of the molecule). The sequence further revealed the occurrence of six cysteine residues and leucine-rich repeats at the invariant relative positions to chicken PG-Lb. In addition, the occurrence of potential glycosaminoglycan-attachment and Asn-glycosylation sites suggested that the cloned molecule is a proteoglycan and corresponds to mouse PG-Lb.

Recent cDNA analyses for small proteoglycans have shown that many of them have leucine-rich repeats and belong to a family of small leucine-rich proteoglycans [6–8]. Three of them have keratan sulphate chains and are named fibromodulin, lumican and keratocan. Fibromodulin was found in cartilage [9], and has been shown to bind both type I and II collagen *in vitro* [10]. Lumican is present in several tissues, including cartilage, in addition to cornea in a non-sulphated form [11]. Keratocan is the 'newest' member of the family [12]. The distribution is more limited than that of lumican or fibromodulin, and is abundant in cornea and sclera, but much less in skin, ligament and cartilage [12]. All of the six different small proteoglycans so far reported in the literature (decorin, biglycan, fibromodulin, lumican, keratocan and PG-Lb) were detected in cartilage, but their localization and expression are quite different from each other. We previously showed by immunohistochemical analysis that PG-Lb was expressed exclusively in the middle zone of embryonic epiphyseal cartilage consisting of flattened chondrocytes [13]. In contrast, other small chondroitin/dermatan sulphate proteoglycans were shown to be expressed rather widely. Decorin is in the whole zones of developing limb cartilage, biglycan is localized in the pericellular space in a variety of cell types, including chondrocytes [14]. Fibromodulin is rather localized in the interterritorial matrix [15]. In these regards, PG-Lb appears to be different in function from the other small proteoglycans. The physiological function of PG-Lb is unknown, but its unique expression suggests that this molecule may be involved in cytodifferentiation in skeletal cartilage tissues. We are, therefore, interested in the molecular mechanism of PG-Lb gene expression, which might help us to understand the function of PG-Lb.

In the present study we isolated and characterized several genomic clones encompassing the entire mouse PG-Lb gene. Comparison of the genomic structure with those of the other murine small proteoglycan genes indicates that the intron/exon organization and boundaries are highly conserved, although their length and sequences significantly vary from gene to gene [16–19], suggesting their evolutionarily close relationship. How-

Abbreviations used: PG-Lb, proteoglycan-Lb; 5′-RACE; **r**apid **a**mplification of **c**DNA 5′-**e**nds; FISH, fluorescence *in situ* hybridization; LA-PCR, long and accurate PCR; poly(A)$^+$, polyadenylated; DAPI, 4,6-diamidino-2-phenylindole.

[1] To whom correspondence should be addressed (e-mail kimata@amugw.aichi-med-u.ac.jp).

The nucleotide sequence data reported here will appear in the DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number D87187 [5′-flanking region, first exon, first intron, second exon and partial second intron of the mouse gene for the PG-Lb core protein (4002 bp)].

ever, comparison of the promoter structure with those of the murine decorin and biglycan genes revealed few potential binding sites for *trans*-acting factors commonly observed among those proteoglycans. Taken together, it is likely that PG-Lb is closely related to biglycan and decorin, but its function is different from theirs. We also show by the fluorescence-*in-situ*-hybridization (FISH) method that the PG-Lb gene is located on the mouse chromosome 10.

EXPERIMENTAL

Isolation and characterization of the mouse PG-Lb gene

A 129/sv mouse genomic DNA library in the λFix II Vector (Stratagene, La Jolla, CA, U.S.A) was used for screening. About 5×10^5 plaques were screened by standard plaque hybridization with various ^{32}P-labelled cDNA probes encoding mouse PG-Lb core protein as follows: The mouse cDNA clone of 1.4 kb containing the entire open reading frame [5] was used for the first cloning. The PCR product corresponding to the 5′- and 3′-terminal regions (nucleotide positions 1–204 and 770–1430 of the mouse PG-Lb cDNA respectively) were used for the second cloning. Restriction-enzyme mapping and LA-PCR (long and accurate PCR) (Takara Shuzo Co., Ltd., Otsu, Japan) analyses were performed to identify three positive clones. The genomic DNA inserts were obtained from the large-scale preparation of the phage DNA by digestion with *Sal*I and then subcloned into plasmid vector pGEM3Zf(−) (Promega Corp., Madison, WI, U.S.A.) for further analyses. The exons were mapped by the combination of hybridization with the cDNA clones and restriction-enzyme mapping. The size of each intron was determined by restriction-enzyme mapping and, in some cases, by PCR using primers located in the exons.

DNA sequencing and analysis

Nucleotide sequence was determined by the dideoxy-chain-termination method using 7-DEAZA Sequencing Kit Ver. 2.0 (Takara) [20] and [α-^{32}P]dCTP (Amersham International, Amersham, Bucks., U.K.) as a tracer. Primers were prepared on the basis of both the cDNA sequence and the proceeding sequence. The DNA sequences were compiled and analysed using GENETYX-MAC computer software program (Software Development Co., Ltd., Tokyo, Japan).

RACE PCR for the determination of 5′-RNA terminals

Rapid **a**mplification of **c**DNA **e**nds (RACE) using a 5′-Ampli FINDER RACE kit (Clontech, Palo Alto, CA, U.S.A) was applied to identify transcription initiation sites [21,22]. The outline is shown in Figure 3(a) (below). The sequence of an antisense gene-specific primer (primer a, 5′-dGCCTCTCAGACACTTTCTCGTTCTCTAC-3′) is derived from the cDNA sequence corresponding to the junction of the second and third exons. To prevent an artificial termination of the reaction due to secondary structures of mRNA, reverse transcription was carried out with avian myeloblastosis-virus reverse transcriptase at 52 °C for 30 min. A specially designed single-stranded anchor oligonucleotide provided in the kit was ligated to the 3′-end of the cDNA using T4 RNA ligase. In subsequent anchor ligation, a portion of the reaction product was used as a template for PCR amplification using a primer complementary to the anchor sequence (primer d) and a nested gene-specific primer corresponding to the cDNA sequence derived from the second exon (primer b, 5′-dTCCCTGTGTCCTCCAAGATGGCATCATACA-3′). The reaction was performed in a Model PJ 9600 DNA thermal cycler (Perkin–Elmer, Norwalk, CT, U.S.A) using a GeneAmp DNA amplification reagent kit (Takara) under the conditions of 40 cycles at 95 °C for 1 min, 65 °C for 2 min, 72 °C for 3 min and an additional 15 min at 72 °C for extension. The product was then used as a template for the second PCR amplification. The reaction was performed using the combination of the anchor primer (primer d) and another antisense gene-specific primer corresponding to the second exons (primer c, 5′-dGAGCTCTGTAGTTGGTGCAGCCA-3′). PCR cycling parameters were 30 cycles at 95 °C for 1 min, 54 °C for 2 min, 72 °C for 3 min and another 15 min at 72 °C for a final extension. The amplified products were analysed on 3% NuSieve 3:1 agarose gel (FMC Bio Products, Rockland, ME, U.S.A) [Figure 3(b) below]. Then specific products were isolated, and the DNA fragments were purified and then cloned into a pGEM3Zf(−) vector after blunt-end formation and phosphorylation. The 5′-ends of individual clones were sequenced by the dideoxy-chain-termination method [20].

Primer extension

The transcription initiation site of the mouse PG-Lb gene was also analysed by the primer extension assay. Polyadenylated RNA [poly(A)$^+$] prepared from cartilages of mice [5] was served as a template for reverse transcriptase and a specific oligonucleotide (primer e, 5′-GACTGAGCATACAGAGTCCTTGGTATTTGCC-3′). The 5′-end of this primer was end-labelled using [γ-^{32}P]ATP (5000 Ci/mmol; Amersham) and T4 Polynucleotide Kinase. About 3 μg of poly(A)$^+$ RNA from mouse cartilage was hybridized with the ^{32}P-labelled primer in a volume of 15 μl containing 150 mM KCl, 10 mM Tris/HCl, pH 8.3, and 1 mM EDTA. The sample was heated at 65 °C for 60 min, and then cooled slowly to room temperature. This reaction mixture was then incubated at 42 °C for 60 min, with 30 μl of extension mixture [30 mM Tris/HCl (pH 8.3)/15 mM $MgCl_2$/8 mM dithiothreitol/1 mM dNTPs/80 μg/ml actinomycin D] containing 20 units of avian-myeloblastosis-virus reverse transcriptase. After the reaction, RNase reaction mixture [25 μg/ml RNase A/5 μg/ml salmon sperm DNA/0.1 M NaCl/10 mM Tris/HCl (pH 7.5)/1 mM EDTA] were added and the mixture was further incubated at 37 °C for 30 min. The reaction mixture was then subjected to phenol/chloroform/3-methylbutan-1-ol extraction and the cDNAs were precipitated with ethanol. The products were analysed on a 6%-polyacrylamide gel in 7 M urea.

FISH

The genomic clones in the PG-Lb gene were used as a probe to map the gene to mouse chromosomes by using FISH [23]. The clones were labelled with digoxigenin-dUTP using nick translation, and were then combined with sheared mouse DNA and hybridized to normal metaphase chromosomes derived from mouse-embryo fibroblast cells in a solution containing 50% formamide, 10% dextran sulphate and 2 × SSC (1 × SSC is 0.15 M NaCl/0.015 M sodium citrate). Specific hybridization signals were detected by incubating the hybridized slides in fluorescence-conjugated anti-digoxigenin antibodies followed by counterstaining with DAPI (4,6-diamidino-2-phenylindole), according to the manufacturer's (Genome Systems Inc., St. Louis, MO, U.S.A) protocol.

RESULTS

Isolation of the mouse PG-Lb gene

A 129/sv mouse genomic DNA library was screened with the mouse PG-Lb cDNA clone that we had obtained previously [5].

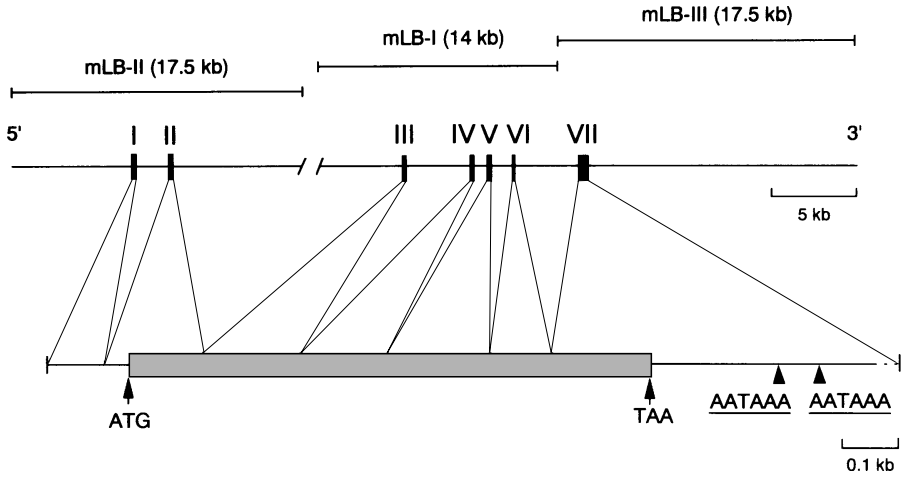

Figure 1 Organization of the mouse PG-Lb gene, alignment of isolated genomic DNA clones and its relationship to structural domains

The position and size (in kb) of the three phage clones representing ≈ 50 kb genomic section are shown. Exons are indicated by closed boxes and are numbered from I to VII. The scale for 5 kb is indicated. Intron as well as 5′- and 3′-flanking region are indicated by lines. In the transcript the initiation (ATG) and stop (TAA) codons are indicated by arrows. Two polyadenylation sites are shown by arrowheads. The scale for 0.1 kb is indicated.

```
5'.. AAG gtaagt ............ 10(py)ttncag G .... 3'  consensus
   (donor site)              (acceptor site)

Exon                                                          Exon
 I ..   GCCAAAG gtaagt ............ acattttctgtttaag AAATCTTG    .... II

 II .. T GTA GAG gtaatg ............ ttctgtttgtttccag AAC GAG AAA .... III
       Val Glu                                      Asn Glu Lys
III .. AAC GAA G gtcaaa ............ cattttctgtttcag AC TTT CCA  .... IV
       Asn Glu A                                    sp Phe Pro
 IV .. AGC TTA A gtatgg ............ atccctttcttctag AT GAT TTA  .... V
       Ser Leu A                                    sn Asp Leu
  V .. T TTT AAA gtaagt ............ atatggttgcttacag GAT ATG TAT .... VI
       Phe Lys                                      Asp Met Tyr
 VI .. C CTC CAG gtaggc ............ ttttttttttttag AAT AAT GAC .... VII
       Leu Gln                                      Asn Asn Asp
```

Figure 2 Alignment of exon/intron boundaries and comparison with consensus sequences for splicing

Exon sequences are indicated by UPPER-CASE letters and intron sequences by lower-case letters. Exon sequences for the 5′-untranslated region are shown in *italics*. Amino acids encoded near and at splice junctions are indicated in three-letter code below their codons. Splice-site consensus sequences are shown at the top and were adapted from the work of Shapiro and Senapathy 22 [16]. Py, pyrimidine.

Only one positive clone, which we designated mLB-I, was obtained. The same library was again screened with the 5′- or 3′-terminal regions of the mouse PG-Lb cDNA as probes. The second screening gave two additional positive clones, mLB-II and mLB-III, of the same (17.5 kb) size. The three clones thus obtained represented a more-than-50-kb genomic gene as a total, although they still did not cover the whole expanse (Figure 1).

Structural organization of the mouse PG-Lb gene

Every exon, including part of the exon-flanking regions, was sequenced bidirectionally. The exon–intron organization of the gene is depicted in Figure 2, and exon sizes are listed in Table 1.

Table 1 Exon sizes of mouse PG-Lb gene

The size of exon I was based on the major transcription start site shown in Figure 4. The size of exon VII was evaluated from the cDNA sequence.

Exon number	Size (bp)
I	93
II	177
II	175
IV	159
V	203
VI	96
VII	> 593

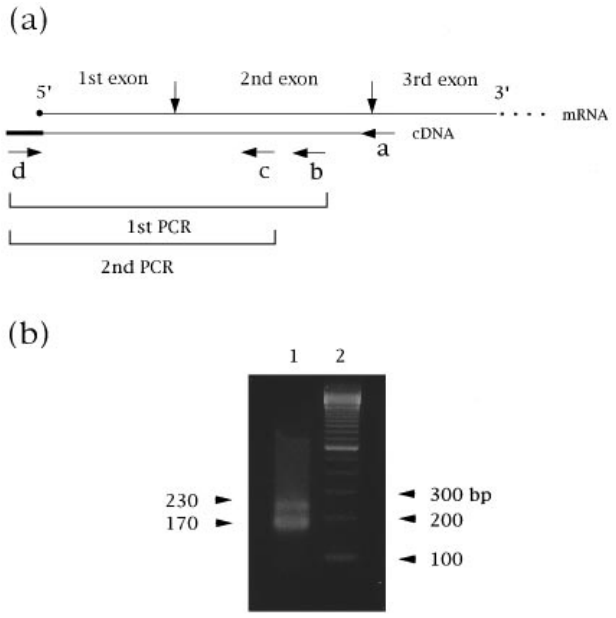

Figure 3 Determination of mouse PG-Lb gene transcription initiation site by a modified 5′-RACE method

(**a**) The location of four different primers was shown by arrows a–d. Primer a was used for the reverse transcription reaction and the others were used for PCR amplification. Different exon boundaries are indicated by vertical arrows. The mRNA and first-strand cDNA are indicated by thin lines. Thick line indicates the anchor DNA. (**b**) A portion of the final PCR product was analysed by agarose-gel electrophoresis. Lane 1, the final PCR product; lane 2, 100 bp DNA ladder size markers. The amplified DNA fragments are indicated by arrowheads.

Sequences for all exon–intron splice junctions matched with consensus sequences for intron splice donor and acceptor sites [24]. Exon I (93 bp) codes for only part of the 5′-untranslated sequences. Exon II (177 bp) is composed of the rest of 5′-untranslated sequences, the starting codon, the putative signal peptide and translated sequences. Exon III (175 bp) has two potential chondroitin sulphate attachment sites. Exon IV (159 bp) contains the consensus amino-flanking region which contains cysteine-rich clusters and is adjacent to the N-terminal side of the leucine-rich repeats [25]. Exons V–VII code for six leucine-rich repeats. There are three leucine-rich repeats in exon V, one in exon VI, one in exon VI–VII, where an intron splits the middle of the leucine-rich repeat 5, and one in exon VII. Exon VII (> 593 bp) also codes the termination codon and the 3′-untranslated region, including two potential polyadenylation signals. Two potential Asn-glycosylation sites are located in exon VII.

Transcription initiation sites

To identify the transcription initiation site of the mouse PG-Lb gene, two different analyses by primer extension and modified 5′-RACE methods were performed. The former method gave no conclusive results inasmuch as it gave multiple signals (results not shown) that may have been due to some steric hindrance *in vitro*. However, the modified 5′-RACE method amplified two DNA fragments (Figure 3). Sequence analysis of the fragments suggested the two candidate sites which were located 94 and 39 bases upstream from the 3′-end of the first exon. However, the latter site does not seem to fit a canonical rule for the transcription initiation site [26]. We therefore predicted that the transcription starts at an adenosine residue of the former site (Figure 4). A possible TATA box was observed about 90 bp upstream of the adenosine that was numbered as position +1 (Figure 4).

```
-280  cacctactaa cctcgtgtta attcccttcc tttcttaact
-240  ggtgaatttg ttagagaaga gacctacgaa aactagcctt
-200  tttctactta aatattttaa ggtgacattt tggcaattac
-160  agctaattag catcgctgca agctttgtgt actgtgaaat
-120  tcggattatt tagcttagct tgtgtgacta taatataatg
 -80  aacggttttt aaagacagag aagcattctc caggttaggt
 -40  ttttcaaaca gaaagttgag tgttacaaat tacacctccc
  +1  ACAGACCAGA ATTCCTATCC ATTGGTCAGG GGCAAATACC
 +41  AAGGACTCTG TATGCTCAGT CACTCTGTGC GACTTCATCA
 +81  AGTCGGAGCC AAAGgtaagt ttttacacat gctttcttct
```

Figure 4 Nucleotide sequence of exon 1 and the 5′-flanking region of mouse PG-Lb gene

Numbering is relative to the start site of the transcription. The complete exon 1 sequence is shown by UPPER-CASE letters. Part (280 bp) of the 5′-flanking region sequence and the first 26 bp of intron are indicated by lower-case letters. The possible TATA sequence is boxed.

Table 2 Location of recognition sequences for transcription factors in the promoters of the mouse PG-Lb gene

Numbers in the location of 5′-nucleotide refer to nucleotide position relative to the major transcription site. The nucleotide sequence data are available from the DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession number D87187. The following putative *cis*-acting elements were identified by using GENETYX-MAC program and computer-assisted analysis: primary target of signal transduction responding to PMA, epidermal growth factor and serum (PEA3), glucocorticoid-receptor-binding element (GR), members of a family of proteins which harbour a DNA-binding domain similar to that of sex-determining region Y (SRY) (Sox 5, Sox 9), two ets (**E**26 **t**ransformation-**s**pecific gene product)-like factors (PU), a binding site for a H4-histone-gene-inducing element (H4TF-1), a homoeodomain protein-binding element (Hox-1) and murine paired-related homoeobox gene (S8). N, any in sequence.

Sequence	Description of factor	Location of 5′-nucleotide
AGGAAA/G	PEA3	−1415
		−1164
		−835
		−563
TGTTCT	GR	−1378
		−606
TGTCCT	GR	−1142
		−925
AGAA/TCAGA/T	GR	−1000
		−918
AACAAT	Sox 5	−1039
		+1323
		+1461
		+1532
GAGGAA	PU	−564
CATTTC	H4TF-1	−377
CATTAGTG	Hox-1	−365
ANC/TC/TAATTAA/GC	S8 homoeoprotein	−160
AACAAAG	Sox 9	+1779

Promoter region

The 5′ flanking region of mouse PG-Lb (1461 bp upstream of the adenosine), the first intron, the second exon, and the partial second intron, which consist of 4002 bp were sequenced (refer to the Nucleotide Sequence Database under the accession number D87187) and examined for the possibility that they contain sites with any structual or functional features of mammalian pro-

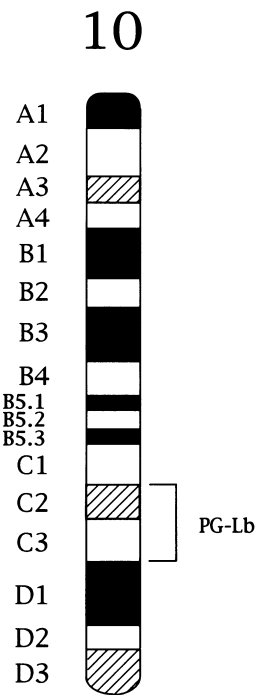

Figure 5 FISH mapping of the mouse PG-Lb gene

A mouse chromosome 10 idiogram shows the location of the mouse PG-Lb gene at the region of 10 C2–C3.

moters. The following putative *cis*-acting elements were detected by using GENETYX-MAC programs (Table 2): H4-histone-gene-inducing element (H4TF-1), glucocorticoid-receptor binding element (GR) and a primary target of signal transduction responding to PMA, epidermal growth factor (EGF) and serum (PEA-3) were found. In addition, there was a homoeodomain protein-binding element (Hox-1). Comparison of putative *cis*-acting elements in the 5′ flanking region of mouse PG-Lb with those of the murine decorin and biglycan genes [16,18,19] revealed a lack of AP-1, AP-2 and SP-1, suggesting some difference from biglycan and decorin in respect of the promoter activity.

In relation to the unique expression of PG-Lb in testis [5], it was considered worthwhile to examine whether there were some putative binding sites for Sox proteins. Computer-assisted analysis of the 5′ flanking region including the first intron of mouse PG-Lb gene revealed the presence of Sox 5 and Sox 9 binding sequences [27,28], which appeared four times in the 5′-flanking region and the first intron and once in the first intron, respectively (Table 2).

Chromosomal location

In the first experiment, fluorescence labelling was observed in the distal portion of a medium-sized chromosome which was believed to be choromosome 10 on the basis of DAPI staining. A second experiment was conducted in which a probe that is specific for the centromeric region of chromosome 10 was co-hybridized with the clones. This experiment resulted in the specfic labelling of the centromere and the distal portion of chromosome 10. Precise measurements of the labelled chromosomes demonstrated that this clone is located at a position which is 73% of the distance from the heterochromatic/euchromatic boundary to the telomere of chromosome 10, an area that corresponds to band 10 C2–C3 (Figure 5). A total of 80 metaphase cells were analysed with 71 exhibiting specific labelling.

DISCUSSION

We report here the genomic DNA structure encoding mouse PG-Lb. The mouse PG-Lb gene cloned from a 129/sv mouse genomic library spans at least 50 kb and consists of seven exons. Exon III encodes the chondroitin sulphate attachment region. Three exons, V, VI and VII encode six leucine-rich repeats. The modified 5′-RACE method gave two candidate sites for the transcription initiation (94 and 39 bases upstream from the 3′-end of the first exon). The site at base 39, however, does not conform to the canonical rule for a transcription initiation site and might be due to an artificial termination of the reverse-transcription reaction by the secondary structure rather than to the real 5′-end of the mRNA. Although further study will be needed for this alternative, a total size of the putative exons corresponds to the observed size of PG-Lb mRNA [5].

Comparison of the exon–intron organization and junctions among small leucine-rich proteoglycan has revealed that they are classified into two groups. Human decorin [17] and biglycan [16] are encoded by eight distinct exons, but human fibromodulin [29] and lumican [11] are encoded by three exons. In the first group the leucine-rich repeats consist of multiple exons, but those in the second group are contained within a single exon. The exon–intron organization and junctions of the PG-Lb gene showed some similarity to those of the first group of biglycan and decorin. However, the expression pattern of PG-Lb [13] completely differs from that of biglycan and decorin [14,18]. Both Northern-blot hybridization and reverse transcription-PCR analyses showed a specific expression of mouse PG-Lb mRNA in cartilage and testis [5] and in the primary culture chondrocytes obtained from mouse embryonic epiphyseal cartilage (results not shown). It is noteworthy that the expression pattern of PG-Lb is very similar to that of a cartilage homoeoprotein 1, Cart-1 [30], in that both molecules are exclusively expressed in cartilage and testis. However, it remains uncertain whether PG-Lb and Cart-1 are expressed in the same cell in these tissues.

Comparison of the sequences of the 5′-flanking promoter region between chicken and mouse showed a slight similarity each other (about 43%; T. Shinomura and K. Kimata, unpublished work). The first exon sequence of mouse PG-Lb showed about 70% identity with chicken PG-Lb. Both sequences contained a possible TATA box sequence, but not a CCAAT element.

We have detected various nuclear-factor binding motifs (*cis*-elements) in the 5′ flanking region and the first intron (Table 2). In relation to expression of PG-Lb in testis, it is noteworthy that Sox 5- and Sox 9-binding sequences are detected in this region [27,28]. Both of these factors are members of a large family of proteins which harbour a DNA-binding domain similar to that of sex-determining region Y (SRY), testis determining gene, in mammals [31]. Recently, Sox 9 was found as a potent activator of the chondrocyte specific expression enhancer of the proα1(II) collagen gene [32]. Therefore the tissue-specific expression of PG-Lb might be closely related to an expression of a set of different Sox genes. As described above, Cart-1 homoeoprotein might also be involved in regulating the expression of PG-Lb, judging from their similar expression patterns. Although the sequence of the *cis*-element recognized by Cart-1 homoeoprotein is still not clear, the Cart-1 homoeodomain shows substantial sequence identity with S8, Pax-3 and MHox. It is noteworthy that S8 homoeoprotein binding sequences [33] are also found in the 5′ flanking region of mouse PG-Lb gene. The nuclear-factor-binding motifs

described above might be important for the tissue-specific expression of PG-Lb.

To characterize the PG-Lb gene further, we determined the location in the mouse genome. The PG-Lb gene was labelled at the distal region of mouse chromosome 10. Both the decorin and the lumican gene had previously been mapped on mouse chromosome 10 [8,18,34]. In contrast, the biglycan gene had been mapped on mouse choromosome X [35]. It is noteworthy that PG-Lb, decorin and lumican genes are located on the same chromosome.

Epiphycan, which is a member of small proteoglycans with a leucine-rich repeat core protein, has recently been isolated from bovine fetal epiphyseal cartilage, and a cDNA study has revealed that this proteoglycan appears to be the bovine equivalent of chicken and mouse PG-Lb [36], which suggests the wide distribution of PG-Lb, with important biological functions, in the Animal Kingdom.

Overall, further study on the regulation of PG-Lb gene expression may help us to understand not only its unique expression mechanism in cartilage, but also its function in cartilage differentiation and morphogenesis.

This work was supported by grants-in-aid from the Ministry of Education, Culture, and Science, Japan, and by special research funds from Seikagaku Corp. and from Sumitomo Pharmaceuticals Co. Ltd.

REFERENCES

1 Shinomura, T., Kimata, K., Oike, Y., Noro, A., Hirose, N., Tanabe, K. and Suzuki, S. (1983) J. Biol. Chem **258**, 9314–9322
2 Shinomura, T. and Kimata, K. (1992) J. Biol. Chem. **267**, 1265–1270
3 Day, A. A., Ramis, C. I., Fisher, L. W., Gehron Robey, P., Termine, J. D. and Young, M. F. (1986) Nucleic Acid Res. **14**, 9861–9876
4 Fisher, L. W., Termine, J. D. and Young, M. F. (1989) J. Biol. Chem **264**, 4571–4576
5 Kurita, K., Shinomura, T., Ujita, M., Zako, M., Kida, D., Iwata, H. and Kimata, K. (1996) Biochem. J. **318**, 909–914
6 Hardingham, T. E. and Fosang, A. J. (1992) FASEB J. **6**, 861–870
7 Kresse, H., Hausser, H. and Schönherr, E. (1993) Experientia **49**, 403–416
8 Iozzo, R. V. and Murdoch, A. D. (1996) FASEB J. **10**, 598–614
9 Oldberg, Å., Antonsson, P., Lindblom, K. and Heinegård, D. (1989) EMBO J. **8**, 2601–2604
10 Hedbom, E. and Heinegård, D. (1989) J. Biol. Chem. **264**, 6898–6905
11 Grover, J., Chen, X.-N., Korenberg, J. R. and Roughley, P. J. (1995) J. Biol. Chem. **270**, 21942–21949
12 Corpuz, L. M., Funderburgh, J. L., Funderburgh, M. L., Bottomley, G. S., Prakash, S. and Conrad, G. W. (1996) J. Biol. Chem. **271**, 9759–9763
13 Shinomura, T., Kimata, K., Oike, Y., Maeda, N., Yano, S. and Suzuki, S. (1984) Dev. Biol. **103**, 211–220
14 Bianco, P., Fisher, L. W., Young, M. F., Termine, J. D. and Robey, P. G. (1990) J. Histochem. Cytochem **38**, 1549–1563
15 Hedlund, H., Mengarelli-Widholm, S., Heinegard, D., Reinholt, F. P. and Svensson, O. (1994) Matrix Biol. **14**, 227–232
16 Fisher, L. W., Heegaard, A.-M., Vetter, U., Vogel, W., Termine, J. D. and Young, M. F. (1991) J. Biol. Chem. **266**, 14371–14377
17 Danielson, K. G., Fazzio, A., Cohen, I., Cannizzaro, L. A., Eichstetter, I. and Iozzo, R. V. (1993) Genomics **15**, 146–160
18 Scholzen, T., Solursh, M., Suzuki, S., Reiter, R., Morgan, J. L., Buchberg, A. M., Siracusa, L. D. and Iozzo, R. V. (1994) J. Biol. Chem **269**, 28270–28281
19 Wegrowski, Y., Pillarisetti, J., Danielson, K. G., Suzuki, S. and Iozzo, R. V. (1995) Genomics **30**, 8–17
20 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
21 Frohman, M. A., Dush, M. K. and Martin, G. R. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 8998–9002
22 Edwards, J. B. D. M., Delort, J. and Mallet, J. (1991) Nucleic Acids Res. **19**, 5227–5232
23 Trask, B. J. (1991) Trends Genet. **7**, 149–154
24 Lewin, B. (1980) Cell **22**, 324
25 Rothberg, J. M., Jacobs, J. R., Goodman, C. S. and Artavanis-Tsakonas, S. (1990) Gene Dev. **4**, 2169–87
26 Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science **209**, 1406–1414
27 Connor, F., Cary, P. D., Read, C. M., Preston, N. S., Driscoll, P. C., Denny, P., Crane-Robinson, C. and Ashworth, A. (1994) Nucleic Acids Res. **22**, 3339–3346
28 Südbeck, P., Lienhard Schmitz, M., Baeuerle, P. A. and Scherer, G. (1996) Nat. Genet. **13**, 230–232
29 Antonsson, P., Heinegård, D. and Oldberg, Å. (1993) Biochim. Biophys. Acta **1174**, 204–206
30 Zhao, G.-Q., Zhou, X., Eberspaecher, H., Solursh, M. and de Crombrugghe, B. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 8633–8637
31 Harley, V. R., Lovell-Badge, R. and Goodfellow, P. N. (1994) Nucleic Acids Res. **22**, 1500–1501
32 Lefebvre, V., Huang, W., Harley, V. R., Goodfellow, P. N. and de Crombrugghe, B. (1997) Mol. Cell. Biol. **17**, 2336–2346
33 de Jong, R., van der Heijden, J. and Meijlink, F. (1993) Nucleic Acids Res. **21**, 4711–4720
34 Funderburgh, J. L., Funderburgh, M. L., Hevelone, N. D., Stech, M. E., Justice, M. J., Liu, C.-Y., Kao, W. W.-Y. and Conrad, G. W. (1995) Invest. Ophthalmol. Vis. Sci. **36**, 2296–2303
35 Chatterjee, A., Faust, C. J. and Herman, G. E. (1993) Mammal. Genome **4**, 33–36
36 Johnson, H. J., Rosenberg, L., Choi, H. U., Garza, S., Höök, M. and Neame, P. J. (1997) J. Biol. Chem., **272**, 18709–18717

22 August 1997/16 January 1998; accepted 17 February 1998