J. Biochem. 122, 17-24 (1997)

Molecular Cloning, Genomic Organization, Promoter Activity, and Tissue-Specific Expression of the Mouse Ryudocan Gene[1]

Shinobu Tsuzuki,* Tetsuhito Kojima,* Akira Katsumi,* Tomio Yamazaki,* Isamu Sugiura,* and Hidehiko Saito*,†,[2]

*First Department of Internal Medicine, Nagoya University School of Medicine, Nagoya 466; and †Aichi Blood Disease Research Foundation, Nagoya 463

Received for publication, December 27, 1996

Ryudocan, a ubiquitous heparan sulfate proteoglycan, is a member of the syndecan family of cell surface proteoglycans. The full-length cDNA encoding the murine ryudocan core protein has now been cloned and sequenced. The deduced primary structure of mouse ryudocan, including the three glycosaminoglycan attachment sites in the extracellular domain as well as the transmembrane and cytoplasmic regions, is highly similar to those of the rat, human, and chicken proteins. Northern analysis detected a 2.7-kb transcript in all mouse tissues examined, with the highest concentrations apparent in liver, kidney, and lung. The mouse ryudocan gene was shown to span approximately 19.7 kb of genomic DNA and to contain five exons, with an intron-exon organization identical to that of the human gene. The promoter region of the mouse gene contains various *cis*-acting elements, including a TATA-like box and a GC box as well as potential binding sites for the transcription factors NF-IL6, MyoD, GATA, C/EBP, AP-2, NF-κB, AP-1, and Sp1. Transient transfection experiments with a construct containing the 690 bp upstream of the transcription start site fused to a luciferase reporter gene showed functional promoter activity. Deletion analysis suggested that the proximal promoter region including the TATA-like box, the GC box, and other Sp1 binding sites was required for full transcriptional activity. These findings will be useful for the study of ryudocan gene regulation and the generation of mice with targeted disruption of the gene.

Key words: cDNA, gene organization, heparan sulfate proteoglycan, promoter function, ryudocan.

Ryudocan, originally isolated from endothelial cells as an anticoagulant heparan sulfate proteoglycan (HSPG), is now known to be a member (syndecan-4) of the syndecan family of cell surface HSPGs (*1*). We have recently demonstrated that human ryudocan is expressed in fibrous connective tissues, peripheral nerve tissues, and placental trophoblasts, and that the protein purified from human endothelial-like cells avidly binds, *via* its heparan sulfate side chain, basic fibroblast growth factor, a nerve growth factor midkine, and an anticoagulant tissue factor pathway inhibitor (*2*). These observations suggest that ryudocan may modulate the activities of soluble factors that show an affinity for heparan sulfate side chain in the cellular microenvironment.

Ryudocan is also implicated in cell-cell and cell-matrix binding. Thus, B lymphoid cells transfected with ryudocan cDNA aggregate in a manner dependent on the heparan sulfate side chain of the encoded protein (*3*). Ryudocan has been localized immunocytochemically to focal adhesions in anchorage-dependent cells (*4*). Given that cytoskeletal microfilament bundles terminate, and cells interact with the extracellular matrix, at focal adhesions, ryudocan may thus play a role in the cytoskeletal organization in relation to the extracellular microenvironment.

To shed light on the function of ryudocan and its transcriptional regulation, we have now determined the nucleotide sequence of both the cDNA and gene encoding the core protein of mouse ryudocan and investigated the functional activity of the gene promoter.

EXPERIMENTAL PROCEDURES

Cloning of Mouse Ryudocan cDNA from LTA Cells—Polyadenylated [poly(A)$^+$] RNA from mouse LTA cells, which have previously been used for HSPG analysis with regard to the specific interaction with antithrombin III (*5*), was reverse-transcribed with the use of an oligo(dT)$_{12-18}$ primer to generate first-strand cDNA. Mouse ryudocan

[1] This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan, and by Grants-in-Aid for Comprehensive Research on Aging and Health and for Survey and Research on Specific Diseases from the Ministry of Health and Welfare of Japan. The nucleotide sequences of the murine ryudocan cDNA and gene have been deposited in GenBank under the accession numbers D89571 and D89572, respectively.

[2] To whom correspondence should be addressed at: First Department of Internal Medicine, Nagoya University School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya 466. Phone: +81-52-744-2136, Fax: +81-52-744-2157

Abbreviations: GAG, glycosaminoglycan; GAPDH, glyceraldehyde-3-phosphatase dehydrogenase; HSPG, heparan sulfate proteoglycan; nt, nucleotides; PCR, polymerase chain reaction; poly(A)$^+$, polyadenylated; RACE, rapid amplification of cDNA ends.

cDNA was then amplified by the polymerase chain reaction (PCR) with degenerate primers (2S and 1A) that had been used for the cloning of rat ryudocan cDNA (*6*). A resulting 206-bp PCR product was isolated, subcloned into the *Xba*I site of pUC19, and shown to contain a 158-bp specific insert [mTV158; nucleotides (nt) 177 to 334] that was identified as a portion of mouse ryudocan cDNA by nucleotide sequencing. Plasmid DNAs were prepared by the method of alkaline lysis (*7*) and sequenced by the dideoxynucleotide chain-termination method (*8*) with Sequenase (United States Biochemical, Cleveland, OH, USA) and both dGTP and dITP.

To clone both ends of the mouse ryudocan cDNA, we applied the method of rapid amplification of cDNA ends (RACE) with modifications (*9*). Poly(A)$^+$ RNA from mouse LTA cells was reverse-transcribed to generate first-strand cDNA with an anchor primer containing $(dT)_{18}$ at the 3′-end (TET-ANC, TCCACTGGGGACTGATTCCGAT$_{18}$). After completion of the poly(A) tail of the first-strand cDNA by incubation with terminal deoxynucleotidyltransferase and dATP, second-strand cDNA was generated with Taq polymerase and TET-ANC as primer. Subsequently, the 5′-most region of mouse ryudocan cDNA was obtained by nested PCR with two primer sets: (i) 5′-CCACTGGGGACTGATTCCG (62ANC) and 5′-GACACGGATGCCAGGCTGTG (MTV131, nt 324 to 305), and (ii) 5′-ACTGGGGACTGATTCCGATTTT (64ANC) and 5′-GTGCATTCTCAGGGATGTGG (MTV114, nt 307 to 288). Both 62ANC and 64ANC were based on the anchor sequence of TET-ANC, whereas MTV131 and MTV114 were based on the mTV158 sequence. A resulting 327-bp PCR product (mRc5′end) was cloned into the *Eco*RV site of pBluescript II KS$^+$ (Stratagene, La Jolla, CA, USA) by the TA cloning method (*10*). A DNA fragment encompassing the remainder of the 5′-half of the mouse cDNA (mRc5′) was obtained by PCR with the first-strand cDNA described above and the specific sense primer MR1 (5′-CGACTGGTTTGCGCTGTTG, nt 1 to 19), derived from the sequence of mRc5′end, and the antisense primer H1471A (5′-GGCTCTTCTCTCATTTTCAAGAAA, nt 1458 to1435), based on the rat ryudocan cDNA (*6*).

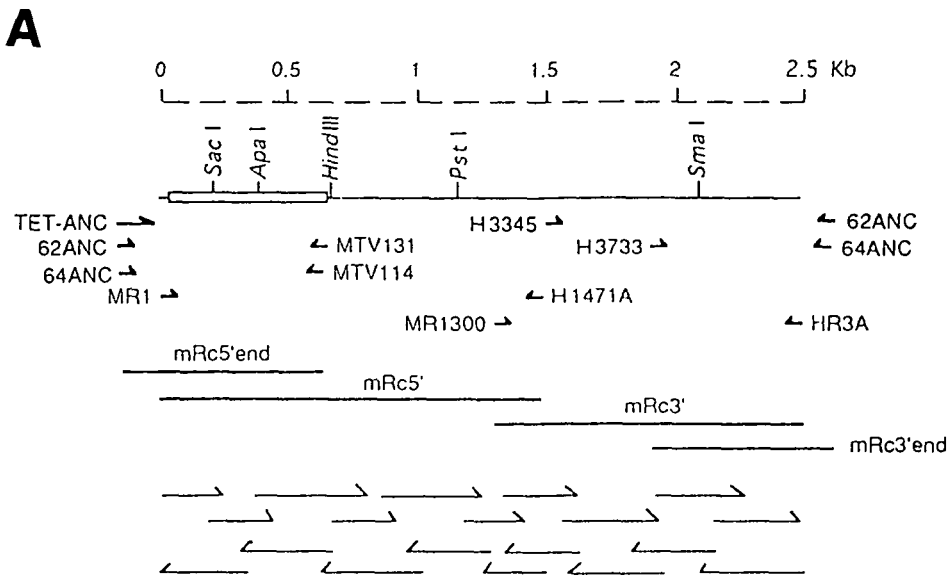
For the cloning of the 3′-region of the mouse ryudocan cDNA, we performed PCR with the specific sense primer MR1300 (5′-AAATGGTTCATTCCTTTATGGG, nt 1306 to 1327), derived from the sequence of mRc5′, and the antisense primer HR3A (5′-AGAAAGTACCAGGTTTTATTATCT, nt 2447 to 2424), based on the polyadenylation signal region of the human ryudocan cDNA (*11*). A resulting 1,142-bp PCR product (mRc3′) was cloned by TA cloning as described above. The 3′-most portion of the cDNA was generated by nested PCR with primer sets comprising (i) H3345 (5′-TATAGCTTCAGACGGGGCTGC, nt 1500 to1520), based on the rat ryudocan cDNA (*6*), and 62ANC, and (ii) H3733 (5′-CTGACTCTAACCTCACTGTGCC, nt 1891 to 1913), based on the rat ryudocan cDNA (*6*), and 64ANC. A resulting 590-bp product (mRc3′ end) was subcloned and sequenced as described above.

Northern Blot Analysis—A Northern blot containing 2 μg of poly(A)$^+$ RNA from various BALB/c mouse tissues (Clontech, Palo Alto, CA, USA) was subjected to hybridization with mouse ryudocan cDNA clones mRc5′ (nt 1 to 1458) and mRc3′ (nt 1306 to 2447). The probes were labeled with ^{32}P by use of a Megaprime DNA labeling kit (Amersham, Arlington Heights, IL, USA). The membrane was analyzed with the Fuji Bio-Imaging Analyzer AS2000 System (Fuji, Tokyo), and exposed to Hyperfilm ECL (Amersham) at −70°C with intesifying screens. The blot was subsequently rehybridized with a ^{32}P-labeled 778-bp *Pst*I-*Xba*I fragment of human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA (*11*). Total RNA (30 μg) from LTA cells or NS-1 cells was also subjected to Northern blot analysis, as above, to compare the amount of mouse ryudocan mRNA.

Isolation of Mouse Ryudocan Genomic Clones—Approximately 9×10^5 plaques of the Lambda FIXII 129SVJ mouse liver genomic phage library (Stratagene) were screened with a 690-bp [α-^{32}P]dCTP-labeled probe generated by *Hin*dIII-*Pst*I digestion of mRc5′ cloned in pBluescript II KS$^+$. Phage DNA from two partially overlapping independent positive clones (λMRG1 and λMRG2) was isolated and digested with *Sac*I, and the resulting DNA fragments were cloned into pBluescript II KS$^+$ for further characterization. All clones were fully sequenced by fluorescence-based cycle sequencing with a model 373A automated DNA sequencer (Applied BioScience, Foster City, CA, USA). Most of the genomic sequence was acquired from nested-deletion overlapping subclones prepared with exonuclease III and S1 nuclease (Erase-a-Base System; Promega, Madison, WI, USA), with the remainder obtained by use of synthetic oligonucleotides based on the preliminary sequence.

S1 Mapping of Transcription Start Site—Total RNA was isolated from LTA cells and analyzed by S1 nuclease protection as described (*12*). A gel-purified oligonucleotide primer, PS1 (5′-GCAAGCAGGCAGGCAGGCG), complementary to nt +27 to +45 of the mouse ryudocan gene sequence, was labeled at the 5′-end with T4 polynucleotide kinase and [γ-^{32}P]ATP, then annealed to a plasmid containing the 5′-most 1,020 bp (nt −690 to +330) of the murine genomic ryudocan DNA in pBluescript II KS$^+$. Complementary DNA was synthesized with Klenow fragment and deoxynucleoside triphosphates, then digested with *Pst*I (nt −222). The resulting ^{32}P-labeled single-stranded S1 probe (complementary to nt −222 to +45) was purified on an alkali denaturing gel. Total RNA (20 μg) from LTA cells together with 10^5 cpm of the S1 probe was precipitated with ethanol, resuspended in 20 μl of formamide hybridization buffer [80% (v/v) formamide, 40 mM Pipes (pH 6.4), 400 mM NaCl, 1 mM EDTA], denatured at 80°C for 10 min, annealed overnight at 30°C, and treated at 37°C for 30 min with S1 nuclease mapping buffer [280 mM NaCl, 50 mM sodium acetate buffer (pH 4.5), 4.5 mM $ZnSO_4$, single-stranded DNA (20 mg/ml), S1 nuclease (1,000 u/ml)]. The reaction was terminated by the addition of EDTA stop buffer and nucleic acid was precipitated with ethanol. The S1 nuclease-resistant DNA fragment was analyzed on a 7% polyacrylamide denaturing gel along with DNA sequence ladders produced with the PS1 primer.

Assay of Promoter Activity In Vivo—The *Sal*I-*Nco*I fragment of the 5′-flanking region (nt −690 to +25) of the mouse ryudocan gene was inserted between *Xho*I and *Nco*I sites of the PGL3-basic vector (Promega), upstream of the luciferase gene. The resulting plasmid (PGL3mR) was used to prepare 5′-deletion constructs by use of the restriction endonucleases *Pvu*II for pGL3mPv (nt −531 to +25), *Pst*I for pGL3mPs (nt −233 to +25), *Hin*dIII for pGL3mH



B

```
                                                                    *
1   CGACTGGTTTGCGCTGTTGAAGCCATGGCGCCTGCCTGCCTGCTTGCGCCGCTGCTGCTGCTGCTCCTCGGAGGCTTCCCCTTGGTCCCC
                                M  A  P  A  C  L  L  A  P  L  L  L  L  L  G  G  F  P  L  V  P     22

                    *           *           *           *           *           *           *
91  CGAGAGTCGATTCGAGAGACAGAGGTCATCGACCCCCAGGACCTCCTGGAAGGCAGATACTTCTCTGGAGCCCTCCCCGACGACGAAGAT
    G  E  S  I  R  E  T  E  V  I  D  P  Q  D  L  L  E  G  R  Y  F  S  G  A  L  P  D  D  E  D     52
    ↑                                                   Δ
                    *           *           *           *           *           *           *
181 GCTGGCGGCTCGGATGACTTTGAGCTCTCGGGTTCTGGAGATCTGGATGACACGGAGGAGCCCAGGCCCTTCCCTGAAGTGATTGAGCCC
    A  G  G  S  D  D  F  E  L  S  G  S  G  D  L  D  D  T  E  E  P  R  P  F  P  E  V  I  E  P     82
                        Δ     Δ
                    *           *           *           *           *           *           *
271 TTGGTGCCACTGGATAACCACATCCCTGAGAATGCACAGCCTGGCATCCGTGTCCCCTCAGAGCCCAAGGAACTGGAAGAGAATGAGGTC
    L  V  P  L  D  N  H  I  P  E  N  A  Q  P  G  I  R  V  P  S  E  P  K  E  L  E  E  N  E  V    112

                    *           *           *           *           *           *           *
361 ATTCCTAAAAGGGCCCCCTCCGACGTGGGGGATGACATGTCCAACAAAGTATCCATGTCCAGCACTGCCCAGGGCAGCAACATCTTTGAG
    I  P  K  R  A  P  S  D  V  G  D  D  M  S  N  K  V  S  M  S  S  T  A  Q  G  S  N  I  F  E    142

                    *           *           *           *           *           *           *
451 AGAACTGAGGTCTTGGCAGCTCTGATCGTGGGCGGCGTGGTAGGCATCCTCTTTGCCGTTTTCCTGATCCTGCTGCTGGTGTACCGCATG
    R  T  E  V  L  A  A  L  I  V  G  G  V  V  G  I  L  F  A  V  F  L  I  L  L  L  V  Y  R  M    172
                ═══════════════════════════════════════

                    *           *           *           *           *           *           *
541 AAGAAGAAGGACGAAGGCAGTTACGACTTGGGCAAGAAACCCATCTACAAAAAAGCCCCCACCAATGAGTTCTACGCATGAAGCTTCCTC
    K  K  K  D  E  G  S  Y  D  L  G  K  K  P  I  Y  K  K  A  P  T  N  E  F  Y  A  *           198

                    *           *           *           *           *           *           *
631 CCGCGAGCGCTGCTTGGACTTATTGGGGAGAGGAGTGGAGGGTTGTGGGTGGCGGGCGTTGGCAGAGAGCAGCAGGCACCTTAATGCTGA
721 CTTGTCAGTATCTCCATCTCTAGTCACCTTTCTGGTGTCAGAAGACATGTGATCTTCTACTGTGCTGCCTGAGAGAGAGAGAGAGAGA
811 GGGAGAGAGAGAGAGAGAGGCTGTGTCTGTGTGTCTGTGTCTCAGTTGCTCTGGCAGAAAATGGGGTTAAACTTGCCCTTTCTGA
901 AGGCAAGCCTAGAATTGGGTCTTTTGTTGTCATTGTTCCAAATTTCCAGAATAGAATATAGGACCAGTTTAGATCCTGTAGTAAACATGT
991 CCCATCTATGACTGCCTTGATTATAGAGGCAAGGGGTTACTGTGTGAATCCCGGCTCCCTTCCACATGCTGTACACCCTATCCATCTGTC
1081 AGGAGCTGGGGCAAGGAGCCAGAACCCTCTGCGCCTTGAGATGAGTCTGCCATGAGAACTTGCTCACGCTGCAGAGTCTCTGTGGCGTAC
1171 CTGGGGGCATTCTAAGTCCAGTGACTTTTGAAATTCAACCTTTTTTTAAAAAAAAAAAATCTAGGCAGGGCGGGGGTGAAGTGCTGAAAGC
1261 TCACACTGAAGTGTGTTTGTAGGAGATTCATAAAACTAAGTTTTTAAATGGTTCATTCCTTTATGGGAGCGTCTGCGCCTGGGCTGAGTG
1351 TGGGGGATACACAGATGTTCTATCTAGAACATTTCCATTGCAACAGCTAAGTTTGTGTTTCGTGGTTTTTTATGTTTTGTTTGTTTTTT
1441 GAAAATGAGAGAAGAGTTGGAGATATGATTTTTATGATTTTTTTTTGTTTTTTACTATTTATAGCTTCAGACAGGGCTGCTTTTCATCTT
1531 TCTATTGTCCCCCATGTTTTTTTTTTTTTTTTTTTTTAATGCTCCGTGTCCTTGTTTTTGACCTTGGCCCTTCCTGAAGTTGCTTTATC
1621 TTACGAAGTAGCTACAGGGTTCTAGCAGTTTCCAGAATATAATGTAGGGGGATAGCGGGATATTTTTGTTTGTGTTCTTGTAATATATTA
1711 TCCTTCCTTGGTTCTAGAAGAATAGATAAATATATTTTTTTAGGCTATAGAATGATATTACAGGTCCCATGTTGGCTGGATGGTTGCGTG
1801 AGTTTTCGTGTGGCTGAAGAGGCTGCTGCCCTCTTACCTTGCCCTGTTCCCGGTGCCTTCTTGGGAATGAGCTAGAGTAGTTTAGGGTCC
1891 CTGACTCGAACCCTCACTGTGCCTTCTGCCGGGGAGGAGGGCCTCCTGGGCCTGCTTCCTCCAGGGGCTCTAGATGCACAATCCAATACA
1981 GTGGCCTCGTGTGGCTGTATAACTCGGTTCACTTCAGACTAGTGATTGGAAGTCATTCCTCGGTTGCACTAGCCACACAAAATGCTCAG
2071 TCACCATCCCCCGGCACCGAGGGTTAAGCTGGTTTAGCAGGTCATTGCTCAGCTTTGCTTGTCTCCCTGGGCAAGGGAAGGGGACCAGG
2161 AGAAAGGTGATGAGGCCCCAGGGACTCCTGTCATCTGGTACTCTCCCTCTTACAGAGGAGAGAGCTAGTGGCTTAAAAGGCCTGGGGTGG
2251 TGGGGAGATCACACCACTCTCAACCATACTCTGTTCCTGTGAGGCCCCCATCACCACCTCCGTGCAGGGTCCTAGCTGGCTGGGTCCTCT
2341 TCTAGCCTTGGTGCCTGCTCCTTTCTGCACCATTCACACTGTTGTCTGTTACTGATCTTTTTTTTTTTTTTTTGATAAAAAGATAAT
2431 AAAACCTGGTACTTTCTAGATTGTCAAAAAAAAAAAAAAAAA
```

Fig. 1. **Restriction map, cloning strategy, and nucleotide and predicted amino acid sequences of mouse ryudocan cDNA.** (A) Restriction map and strategy for the isolation of mouse ryudocan cDNA. The open box in the restriction map represents the coding region of the cDNA. The bold half-arrows indicate the position and direction of PCR primers used for cloning of the cDNA. The four clones obtained are shown below the primers, and the thin half-arrows indicate the direction and length of the nucleotide sequences obtained from analysis of the various PCR products. (B) Nucleotide sequence of the cDNA and predicted amino acid sequence of the encoded protein. The arrow points to the predicted cleavage site of the signal peptide. The double underline shows the putative transmembrane domain. Open triangles indicate the three potential GAG attachment sites. The conventional polyadenylation signal sequence, AATAAA, is single-underlined. Numbers on the left denote the position of nucleotides in the cDNA, with every 10th residue marked with an asterisk, and numbers on the right indicate amino acids.

(nt −94 to +25), and *Bss*HII for pGL3mB (nt −18 to +25). LTA cells were cotransfected with 0.5 μg each of the indicated luciferase reporter constructs and a β-galactosidase reporter plasmid containing the SV40 early promoter and enhancer (pSV-LacZ, used to monitor transfection efficiency) with the use of Lipofectin reagent (Gibco BRL, Gaithersburg, MD, USA), as described previously (*13*). NS-1 cells were transfected with 10 μg of the plasmids by electroporation (960 μFD, 300 V) using Gene Pulser (Bio-Rad, Alfred Nobel Drive Hercules, CA, USA). Luciferase and β-galactosidase activities were measured as described previously (*13*).

RESULTS AND DISCUSSION

Cloning of Mouse Ryudocan cDNA and Analysis of the Deduced Protein Sequence—We isolated DNA fragments corresponding to the full-length mouse ryudocan cDNA from LTA cells, which express a HSPG anticoagulant activity (*5*), with the cloning strategy described in "EXPERIMENTAL PROCEDURES" (Fig. 1A). The mouse ryudocan cDNA comprises 2,455 bp and contains an open reading frame encoding a protein of 198 amino acids with a predicted molecular mass of 21.5 kDa. The predicted mouse protein shows 94.4, 80.3, and 57.1% sequence identity with the corresponding rat (*6*), human (*11*), and chicken (*14*) proteins, respectively (Fig. 2). The deduced amino acid sequence of mouse ryudocan contains a 23-residue signal peptide (*15*), a putative extracellular domain of 122 residues that includes three potential glycosaminoglycan (GAG) attachment sites, a membrane-spanning domain of 25 hydrophobic residues, and an intracellular region of 28 residues. These four domains respectively exhibit 82.6, 95.1, 96.0, and 100% amino acid identity with rat ryudocan, 47.8, 79.5, 92.0, and 100% identity with human ryudocan, and 34.8, 46.7, 84.0, and 100% identity with chicken ryudocan. The NH_2- and COOH-terminal sequences of the putative mature proteins are highly

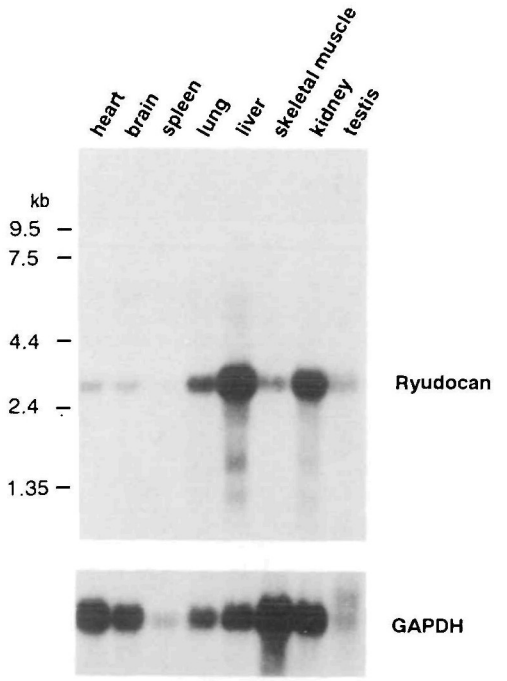


Fig. 3. **Tissue distribution of mouse ryudocan mRNA.** Poly-(A)$^+$ RNA from various mouse tissues was subjected to Northern blot analysis with ^{32}P-labeled probes for mouse ryudocan mRNA and, subsequently, GAPDH mRNA. The positions of molecular size standards (in kilobases) are shown on the left.

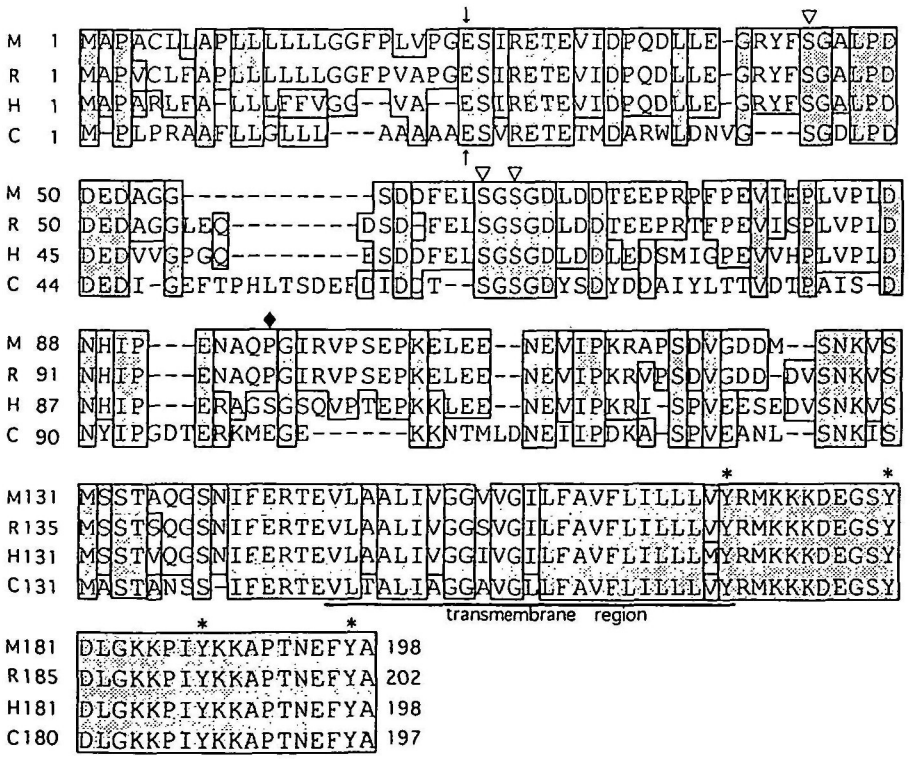


Fig. 2. **Homology among mouse (M), rat (R), human (H), and chicken (C) ryudocan protein sequences.** Numbers denote amino acid residues, and sequences were aligned manually. Residues identical in at least two species are boxed and those identical in all four species are also shaded. The underline indicates the transmembrane domain. The arrows point to the predicted cleavage site of the signal peptide. Open triangles show the potential GAG attachment sites conserved among all four species, with the additional site in the human sequence indicated by the closed diamond. The four conserved tyrosine residues in the transmembrane and intracellular domains are indicated by asterisks.

conserved, suggesting that these regions, including GAG attachment sites and four tyrosine groups in the transmembrane and cytoplasmic domains, might play important roles in ryudocan function (Fig. 2).

Distribution of Ryudocan mRNA in Mouse Tissues—Northern blot analysis revealed a 2.7-kb ryudocan mRNA in all mouse tissues examined (Fig. 3). Normalization of the amount of ryudocan mRNA with that of GAPDH mRNA showed that ryudocan gene transcripts were most abundant in liver, kidney, and lung, consistent with the tissue distribution of human ryudocan mRNA (*11*).

Structure of the Mouse Ryudocan Gene—The mouse ryudocan gene was isolated and sequenced as described in "EXPERIMENTAL PROCEDURES" (Fig. 4). The genomic clones encompassed ~19.7 kb, consisting of 690 bp of the 5′-flanking region, all five exons and introns, and a 122-bp 3′-flanking region. Sequences for all intron-exon splice junctions are consintent with the GT-AG rule and match the consensus sequences for intron splice donors and acceptors (*16*). The splice junctions between exons II and III and between exons IV and V are located within codons, whereas those between the remaining adjacent exons are located between codons. This exon-intron boundary organization is identical to that of the human ryudocan gene (*13*).

Exon I comprises 97 bp and encodes the entire 5′-untranslated region of the cDNA, the translation initiation site, the signal peptide, and the first two amino acids of the mature protein. Exon II (127 bp) encodes the three putative GAG attachment sites, which are conserved in rat (*6*), human (*11*), and chicken (*14*) ryudocan. Exon III (47 bp) and exon IV (196 bp) encode the middle and COOH-terminal regions of the extracellular domain. Exon V (1,988 bp)

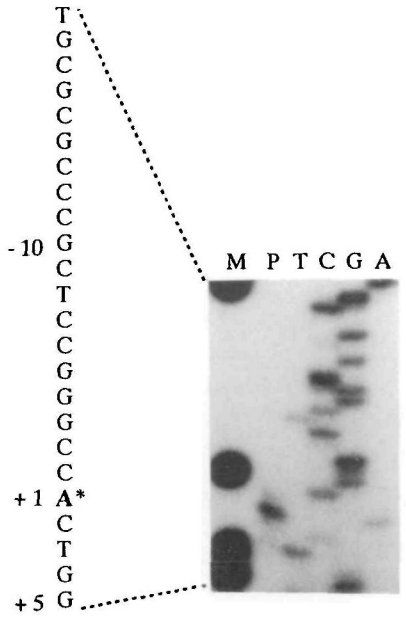
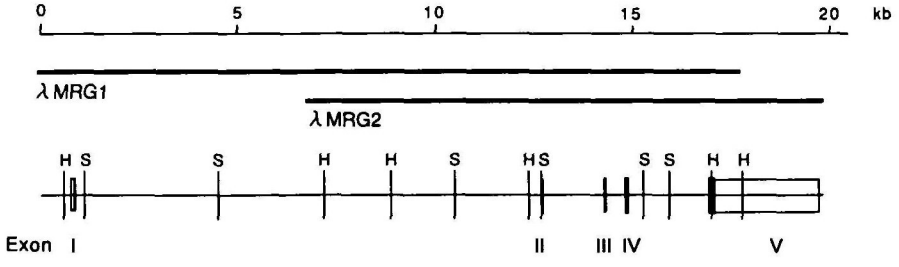


Fig. 5. **Determination of the transcription start site of the mouse ryudocan gene by S1 nuclease protection analysis.** A ^{32}P-labeled S1 probe complementary to nt −222 to +45 of the mouse ryudocan gene was annealed to total RNA (20 μg) from LTA cells and then digested with S1 nuclease. The protected DNA (P) and DNA sequence ladders (T, C, G, A) generated with the same oligonucleotide as that used to produce the S1 probe were analyzed on a polyacrylamide denaturing gel. The transcription start site is indicated by A*, and the numbers indicate nucleotide position. Dephosphorylated *Hin*fI fragments of $\phi \times 174$ were labeled with ^{32}P at the 5′-end and used as molecular size markers (lane M).

A



B

Exon (bp)		Splice donor	Intron (kb)	Splice acceptor	
I (97)	GAG TCG E S	gtatgt ·····	(11.8)	····· atctctcttcttag	ATT CGA I R
II (127)	GAT CTG G D L D	gtatgg ·····	(1.5)	····· ttctacctttgcag	AT GAC D
III (47)	CCC TTG P L	gtgagt ·····	(0.6)	····· cctttctctctcag	GTG CCA V P
IV (196)	TTG GCA G L A A	gtaagg ·····	(2.7)	····· tctctcctactgcag	CT CTG L
V (1,988)					

Fig. 4. ***Restriction map and exon-intron organization of the mouse ryudocan gene.*** (A) Restriction map of mouse ryudocan gene. The locations of λMRG1 and λMRG2 clones are shown below the scale. Boxes (I to V) indicate exons, with closed boxes representing the coding region and open boxes the 5′- and 3′-untranslated regions. The restriction sites of enzymes are indicated by H (*Hin*dIII) and S (*Sac*I). (B) Exon-intron boundaries of the mouse ryudocan gene. Exon sequences are indicated by uppercase letters and intron sequences by lowercase letters. Sizes of exons and introns are given in base pairs. The amino acid residues at the splice sites are shown below the exon sequences.

encodes the transmembrane and cytoplasmic domains as well as the long 3′-untranslated region.

Comparison of the nucleotide sequences of the ryudocan cDNA obtained from LTA cells (C3H/An mouse) and the ryudocan gene from the 129/SVJ mouse revealed several differences. The second nucleotide (G) of the cDNA was substituted by C in the corresponding gene sequence. A C→T substitution was also observed in the protein-coding region of the gene; however, this substitution did not affect the encoded amino acid (Ile[187], ATC→ATT). Differences in nucleotide sequence were also apparent between the 3′ untranslated region of the cDNA and exon V of the gene, as is the case in some of the cDNA sequence data of other proteins (*17*). These sequence differences might reflect gene polymorphisms or the difference in the strains of mice from which the analyzed cDNA and gene originated.

To determine the transcription start site of the mouse ryudocan gene, we performed S1 nuclease protection analysis (*12*) with total RNA from LTA cells and a 267-base S1 probe encompassing the first 47 bases of the cDNA as well as 220 bases of the 5′-flanking region (Fig. 5). The transcription start site was localized to 22 bp upstream of the first ATG, which was identical not to the 5′-end of the cDNA obtained by RACE but to the third nucleotide from the 5′-end. This discrepancy may be attributable to the difference in nucleotide sequence described above between

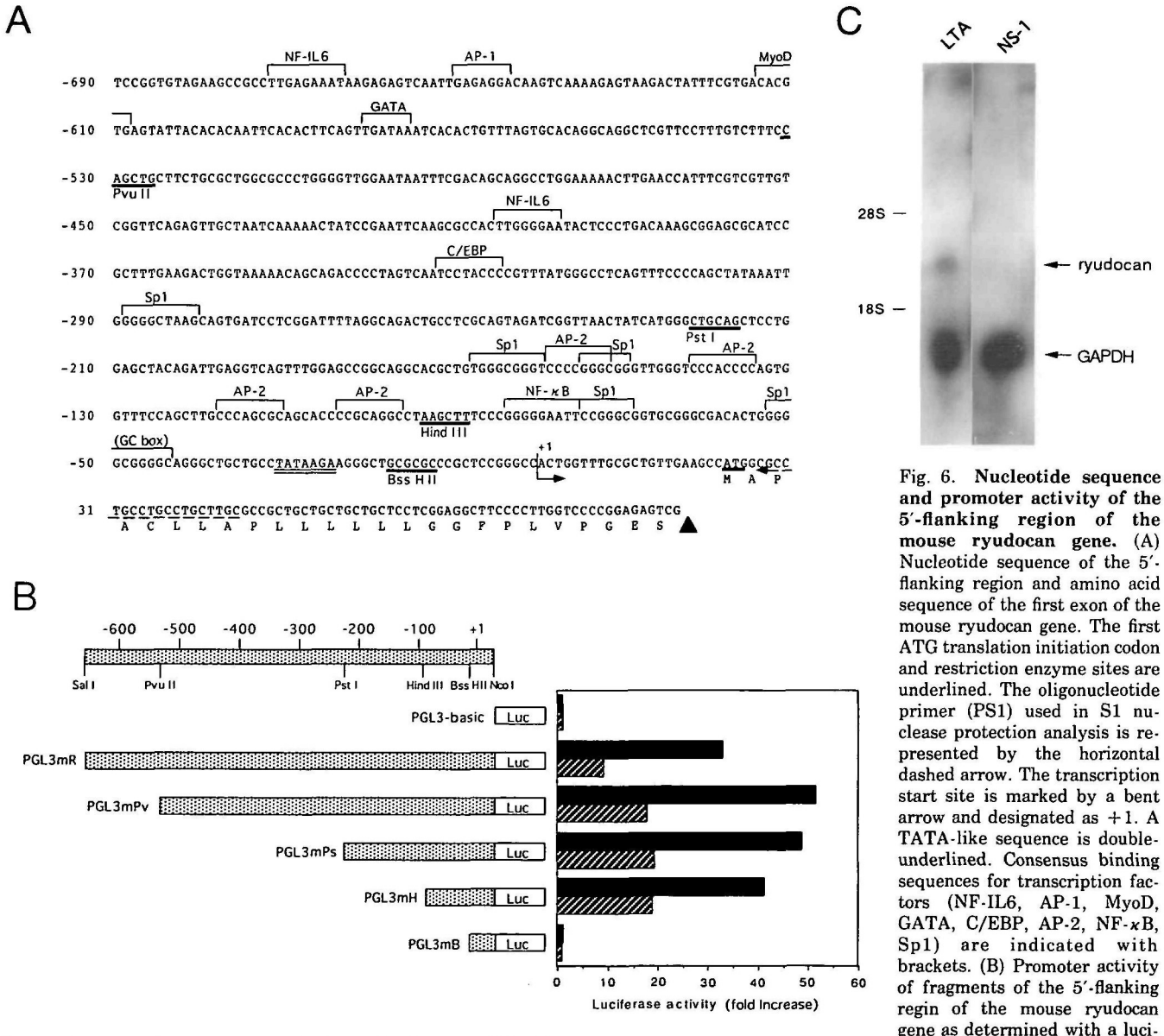


Fig. 6. **Nucleotide sequence and promoter activity of the 5′-flanking region of the mouse ryudocan gene.** (A) Nucleotide sequence of the 5′-flanking region and amino acid sequence of the first exon of the mouse ryudocan gene. The first ATG translation initiation codon and restriction enzyme sites are underlined. The oligonucleotide primer (PS1) used in S1 nuclease protection analysis is represented by the horizontal dashed arrow. The transcription start site is marked by a bent arrow and designated as +1. A TATA-like sequence is double-underlined. Consensus binding sequences for transcription factors (NF-IL6, AP-1, MyoD, GATA, C/EBP, AP-2, NF-κB, Sp1) are indicated with brackets. (B) Promoter activity of fragments of the 5′-flanking regin of the mouse ryudocan gene as determined with a luciferase (Luc) reporter gene in a transient transfection assay. The left part of the figure shows the restriction sites of the 5′-flanking region (*SalI-NcoI* fragment; nt −690 to +25) used for preparation of the deletion constructs as described in "EXPERIMENTAL PROCEDURES." Luciferase activity, corrected for differences in transfection efficiency on the basis of the activity of a cotransfected β-galactosidase construct, was determined in extracts of LTA cells (■) or NS-1 cells (▨) transfected with the indicated constructs. Data are expressed as fold increase relative to the luciferase activity obtained with the parent PGL3-basic vector, and are means of triplicates from a single experiment representative of two others. (C) Expression of mouse ryudocan mRNA in LTA cells and NS-1 cells. RNA from cells was subjected to Northern blot analysis with ^{32}P-labeled probes for mouse ryudocan mRNA and GAPDH mRNA. The positions of ribosomal RNA are shown on the left.

the 5′-end of the mRNA (*C*GA*CUGG...) (Fig. 1B) and the corresponding region of the gene (...*C*CA*CTGG...) (Fig. 6A). Thus, when the S1 probe complementary to the 129SVJ mouse gene was annealed to ryudocan mRNA from LTA cells (C3H/An mouse), the probe was protected from digestion with S1 nuclease in the complementary region (A*CTGG...), whereas the upstream region of the probe was cleaved. The 5′-end of the mRNA determined by S1 nuclease protection analysis was therefore "A*," whereas the 5′-end of the cDNA obtained by RACE was "*C*." Primer extension analysis was also performed to determine the transcription start site, but was not successful, presumably because of the highly GC-rich structure of the 5′-flanking region of the gene (*18*). The transcription start site determined by S1 nuclease protection analysis (A*) was denoted +1 in the numbering of the nucleotide sequence of the gene, whereas the 5′-end of the cDNA determined by RACE (*C*) was designated +1 in the numbering of the nucleotide sequence of the cDNA. The nucleotide sequence indicated by the sequence ladders in Fig. 5 appeared to differ with respect to certain G or C residues, presumably due to G/C compression (*8*), from the sequence shown in Fig. 6, which was confirmed by repeated analysis and is also depicted by letters on the left in Fig. 5.

Promoter Activity of the 5′-Flanking Region of the Mouse Ryudocan Gene in a Transient Transfection Assay—The 5′-flanking region (690 bp) of the mouse ryudocan gene contains several potential binding sites for transcription factors (Fig. 6A). A TATA-like sequence (TATAAGA) is located 31 bp upstream from the transcription start site. Five potential binding sites for the constitutive transcription factor Sp1, including a GC box, are also located in the 5′-flanking region, as are putative binding sites for NF-IL6, AP-1, MyoD, GATA, C/EBP, AP-2, and NF-κB.

To investigate the functional activity of this region as a promoter, we transfected LTA cells, which express mouse ryudocan, with constructs containing various portions upstream of a luciferase reporter gene and subsequently measured luciferase activity in cell lysates (*13*). The pGL3mR construct, encompassing nt −690 to +25 of the mouse ryudocan gene, yielded ~33 times as much luciferase activity as the parent pGL3-basic vector. The pGL3mPv (nt −531 to +25) and pGL3mPs (nt −223 to +25) constructs yielded similar amounts of luciferase activity, which were greater than that obtained with pGL3mR. This latter observation suggests that the NF-IL6, AP-1, MyoD, GATA, C/EBP, and 5′-most Sp1 binding sites are not responsible for the induced luciferase activity and that a silencer element might be present between nt −690 and −531. In contrast, pGL3mH (nt −94 to +25), which contained an NF-κB and two Sp1 sites but had lost two Sp1 sites as well as four AP-2 sites, yielded less luciferase activity than pGL3mR. The pGL3mB construct (nt −18 to +25), which contained no potential binding sites for transcription factors in the 5′-flanking region, showed no promoter activity. These results suggest that a TATA-like sequence and a GC box contribute to basal transcription of the mouse ryudocan gene, and that proximal Sp1 binding sites, other than the GC box, and AP-2 sites mediate increased levels of transcription. They also suggest that a TATA-like sequence and binding sites for the ubiquitous transcription factor Sp1 (*19*) underlie the constitutive transcription and the widespread expression of the ryudocan gene in different tissues (Fig. 3) (*11, 20*).

However, this is not the case for different cells. NS-1 cells, which are derived from mouse plasma cells, did not express ryudocan mRNA as revealed by Northern blot analysis (Fig. 6C), but they did induce luciferase activity driven by the 690-bp 5′-flanking region of mouse ryudocan gene, with full luciferase activity being induced by the most proximate portion of the 5′-flanking region including a TATA-like sequence, two Sp1 binding sites, and a NF-κB site (Fig. 6B). Although induced luciferase activity in NS-1 cells was approximately three times less than that in LTA cells, this difference might not explain the difference in the amount of ryudocan mRNA detected by Northern blot analysis in these two cells (Fig. 6C). Thus, cell-specific expression might not be attributable to the difference in the activity of transcription driven by the 690-bp 5′-flanking region of mouse ryudocan gene. One possible explanation of cell-specific expression is that silencing factors are crucial for the specificity, as is the case in expresion of the type II voltage-dependent sodium channel gene, which is restricted to neurons by a silencer element active in nonneuronal cells (*21, 22*). Because the 690-bp 5′-flanking region could drive transcription in cells not expressing ryudocan, as revealed by luciferase assay in NS-1 cells, a more distal portion of the 5′-flanking region might be required for the cell-specific expression, possibly *via* cell-specific silencers in cells which do not express ryudocan. Even if this is so, additional transcription factors, which are involved in cell-specific gene transcription, could further modulate the expression of ryudocan mRNA.

Whether the binding sites for NF-IL6, AP-1, MyoD, GATA, C/EBP, AP-2, and NF-κB, which are thought to contribute to tissue-specific or regulated gene transcription, are important for transcriptional control of the ryudocan gene requires further investigation. Expression of the related protein syndecan-1 varies markedly during organogenesis in the mouse (*23*). MyoD (*24*), GATA (*25, 26*), and C/EBP (*27*) play important roles in tissue-specific gene expression in skeletal muscle; endothelial and hematopoietic cells as well as fetal brain; and liver, adipose, and placental tissue, respectively. Skeletal muscle (*11*), endothelial cells, B lymphocytes with the exception of B stem cells, liver (*20*), and placenta (*2*) all express ryudocan. Thus, MyoD, GATA, and C/EBP may function in the developmentally regulated expression of ryudocan in these tissues. Similarly, NF-κB (*28*), originally recognized as a factor required for the expression of immunoglobulin κ chain, may regulate expression of ryudocan during B cell maturation. Alternatively, given that NF-κB (*28*) and NF-IL6 (*29*) contribute to the induction of gene expression by cytokines, they may mediate expression of ryudocan during inflammation.

We thank Dr. R.D. Rosenberg for providing LTA cells, Dr. H. Nagase for NS-1 cells, Dr. N. Emi for supplying the pSV-LacZ plasmid, and Mr. S. Suzuki, Ms. C. Wakamatsu, and Ms. M. Suzuki for technical assistance.

REFERENCES

1. Kojima, T., Leone, C.W., Marchildon, G.A., Marcum, J.A., and Rosenberg, R.D. (1992) Isolation and characterization of heparan sulfate proteoglycans produced by cloned rat microvascular

endothelial cells. *J. Biol. Chem.* **267**, 4859-4869
2. Kojima, T., Katsumi, A., Yamazaki, T., Muramatsu, T., Nagasaka, K., Ohsumi, K., and Saito, H. (1996) Human ryudocan from endothelium-like cells binds basic fibroblast growth factor, midkine, and tissue factor pathway inhibitor. *J. Biol. Chem.* **271**, 5914-5920
3. Stanley, M.J., Liebersbach, B.F., Liu, W., Anhalt, D.J., and Sanderson, R.D. (1995) Heparan sulfate-mediated cell aggregation. Syndecans-1 and -4 mediate intercellular adhesion following their transfection into human B lymphoid cells. *J. Biol. Chem.* **270**, 5077-5083
4. Woods, A. and Couchman, J.R. (1994) Syndecan-4 heparan sulfate proteoglycan is a selectively enriched and widespread focal adhesion component. *Mol. Biol. Cell* **5**, 183-192
5. de Agostini, A.L., Lau, H.K., Leone, C., Youssoufian, H., and Rosenberg, R.D. (1990) Cell mutants defective in synthesizing a heparan sulfate proteoglycan with regions of defined monosaccharide sequence. *Proc. Natl. Acad. Sci. USA* **87**, 9784-9788
6. Kojima, T., Shworak, N.W., and Rosenberg, R.D. (1992) Molecular cloning and expression of two distinct cDNA-encoding heparan sulfate proteoglycan core proteins from a rat endothelial cell line. *J. Biol. Chem.* **267**, 4870-4877
7. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*, unit 1.25, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
8. Slatko, B.E., Albright, L.M., and Tabor, S. (1994) DNA sequencing by the dideoxy method in *Current Protocols in Molecular Biology* (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds.) unit 7.4, Wiley, New York
9. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002
10. Marchuk, D., Drumm, M., Saulino, A., and Collins, F.S. (1991) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* **19**, 1154
11. Kojima, T., Inazawa, J., Takamatsu, J., Rosenberg, R.D., and Saito, H. (1993) Human ryudocan core protein: molecular cloning and characterization of the cDNA, and chromosomal localization of the gene. *Biochem. Biophys. Res. Commun.* **190**, 814-822
12. Greene, J.M. and Struhl, K. (1994) S1 analysis of messenger RNA using single-stranded DNA probes in *Current Protocols in Molecular Biology* (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds.) unit 4.6, Wiley, New York
13. Takagi, A., Kojima, T., Tsuzuki, S., Katsumi, A., Yamazaki, T., Sugiura, I., Hamaguchi, M., and Saito, H. (1996) Structural organization and promoter activity of the human ryudocan gene. *J. Biochem.* **119**, 979-984
14. Baciu, P.C., Acaster, C., and Goetinck, P.F. (1994) Molecular cloning and genomic organization of chicken syndecan-4. *J. Biol. Chem.* **269**, 696-703
15. von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* **14**, 4683-4690
16. Shapiro, M.B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryocytes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155-7174
17. Miyauchi, T., Jimma, F., Igakura, T., Yu, S., Ozawa, M., and Muramatsu, T. (1995) Structure of the mouse basigin gene, a unique member of the immunoglobulin superfamily. *J. Biochem.* **118**, 717-724
18. Triezenberg, S.J. (1994) Primer extension in *Current Protocols in Molecular Biology* (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K., eds.) unit 4.8, Wiley, New York
19. Saffer, J.D., Jackson, S.P., and Annarella, M.B. (1991) Developmental expression of Sp1 in the mouse. *Mol. Cell. Biol.* **11**, 2189-2199
20. Kim, C.W., Goldberger, O.A., Gallo, R.L., and Bernfield, M. (1994) Members of the syndecan family of heparan sulfate proteoglycans are expressed in distinct cell-, tissue-, and development-specific patterns. *Mol. Biol. Cell* **5**, 797-805
21. Mori, N., Schoenherr, C., Vandenbergh, D.J., and Anderson, D.J. (1992) A common silencer element in the SCG10 and type II Na^+ channel genes binds a factor present in nonneuronal cells but not in neurinal cells. *Neuron* **9**, 45-54
22. Chong, J.A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altshuller, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995) REST: A mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949-957
23. Sutherland, A.E., Sanderson, R.D., Mayes, M., Siebert, M., Calarco, P.G., Bernfield, M., and Damsky, C.H. (1991) Expression of syndecan, a putative low-affinity fibroblast growth factor, in the early mouse embryo. *Development* **113**, 339-351
24. Davis, R.L., Weintraub, H., and Lassar, A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987-1000
25. Leonard, M., Brice, M., Engel, J.D., and Papayannopoulou, T. (1993) Dynamics of GATA transcription factor expression during erythroid differentiation. *Blood* **82**, 1071-1079
26. Yamamoto, M., Ko, L.J., Leonard, M.W., Beug, H., Orkin, S.H., and Engel, J.D. (1990) Activity and tissue-specific expression of the transcriptional factor NF-E1 multigene family. *Genes Dev.* **4**, 1650-1662
27. Birkenmeier, E.H., Gwynn, B., Howard, S., Jeny, J., Gordon, J.I., Landschulz, W.H., and McKnight, S.L. (1989) Tissue-specific expression, developmental regulation, and genetic mapping of the gene encoding CCAAT/enhancer binding protein. *Genes Dev.* **3**, 1146-1156
28. Baeuerle, P.A. and Henkel, T. (1994) Function and activation of NF-kappa B in the immune system. *Annu. Rev. Immunol.* **12**, 141-179
29. Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirao, T., and Kishimoto, T. (1990) A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. *EMBO J.* **9**, 1897-1906