The Bovine Mimecan Gene

MOLECULAR CLONING AND CHARACTERIZATION OF TWO MAJOR RNA TRANSCRIPTS GENERATED BY ALTERNATIVE USE OF TWO SPLICE ACCEPTOR SITES IN THE THIRD EXON*

(Received for publication, February 24, 1999)

Elena S. Tasheva‡, Martha L. Funderburgh, James McReynolds, James L. Funderburgh, and Gary W. Conrad

From the Division of Biology, Kansas State University, Manhattan, Kansas 66506-4901

Mimecan is a proteoglycan expressed by many connective tissues. It was originally isolated in a truncated form as a bone-associated glycoprotein, osteoglycin, and was considered an osteoinductive factor. Recently, we demonstrated that the full-length translation product of the cDNA encoding mimecan is a corneal keratan sulfate proteoglycan present in other tissues without keratan sulfate chains. We also described multiple mimecan mRNA transcripts generated by differential splicing and alternative polyadenylation. In this study, we isolated genomic clones and determined the genomic organization of the bovine mimecan gene. The gene is spread over >33 kilobases of continuous DNA sequence and contains eight exons. The newly discovered first exon, identified by 5′-rapid amplification of cDNA ends, consists of a 5′-untranslated region and is enriched in C+G nucleotides. Two transcription initiation sites starting at the first and at the second exons were determined by primer extension. Molecular characterization shows that alternatively spliced RNA isoforms are generated by the use of two distinct splice acceptor sites in the third exon situated 278 base pairs apart. We determined a partial genomic structure of the human mimecan gene and demonstrated two alternatively spliced RNA transcripts that are generated likewise. Despite the diversity of mimecan transcripts, the primary structure of the core protein is encoded from exons 3 to 8 and remains unchanged, indicating its functional importance. Using ribonuclease protection assay, we analyzed the patterns of spliced RNA expressed in cultured bovine keratocytes. We demonstrated that their expression is differentially modulated in a temporal manner by basic fibroblast growth factor.

Mimecan belongs to a family of leucine-rich proteoglycans that are secreted into the extracellular matrix. Initially, it was isolated as a glycoprotein of 12 kDa from bovine bone (1, 2) and as a 25-kDa KSPG1 from bovine cornea (3, 4). The protein of 12 kDa was originally named osteoinductive factor; subsequently, after the determination that co-purifying bone morphogenetic proteins 2 and 3 are the source of its growth stimulatory activity, the protein was renamed osteoglycin (5, 6). At the same time, a report from this laboratory described three unique core proteins separated from purified corneal KSPGs, and these were later identified as lumican (7), keratocan (8), and KSPG25, or mimecan (9). In addition to cornea, by use of an antibody to KSPG25, the protein was detected in a number of other bovine tissues. Comparison of the cDNA sequences showed that the entire sequence for osteoglycin occurs in KSPG25. As a result of these findings, the name mimecan was given to the gene and its full-length translation product (a 34-kDa glycoprotein) in recognition of its elusive nature, its presence in many tissues as a non-sulfated glycoprotein, and its tissue-specific glycosylation and thus conversion to KSPG in cornea (9). In addition to the tissue-specific post-translational modification of this protein, recently we demonstrated multiple mimecan mRNA transcripts resulting from differential splicing and alternative polyadenylation (10). The diversity of mimecan mRNA isoforms was found in the non-coding regions. The primary structure of the encoded protein remained unchanged, indicating its functional importance. Indeed, mimecan appears to be conserved among species (2, 9, 11, 12). Thus, in situ proteolytic cleavage of the full-length protein was proposed to account for the multiple size protein products detectable in different tissues: a 12-kDa osteoglycin in bone and a 25-kDa KSPG25 in cornea (9). At present, the physiological function of mimecan is unknown. However, the protein shares a common core structure with a family of related proteins known as SL-RPs. These are characterized by repeats of leucine-rich motifs bounded by two cysteine clusters and a conserved pattern of glycosylation. Numerous examples illustrate the ability of these proteins to bind growth factors and/or growth factor receptors and therefore to modulate cell proliferation and differentiation (reviewed in Refs. 13 and 14). These characteristics suggest that mimecan may play its specific role in modulation of cellular growth and matrix assembly, similarly to other closely related PGs, such as decorin (15, 16, 17) and lumican (18, 19, 20). As yet, there is no information on the structure of the mimecan gene. In this study, we describe the isolation of genomic clones spanning the bovine mimecan gene. We report here the structure of the gene and describe a new exon in the 5′-untranslated region. We show that mimecan transcripts from bovine cornea are heterogeneous in their 5′-regions. Two RNA transcripts are generated by alternative transcription initiation sites and an additional two result from alternative use of two distinct splice acceptor sites in the third

cDNA ends; RPA, ribonuclease protection assay; SLRP, small leucine-rich proteoglycan.

* This work was supported by National Institutes of Health Grants EY00952 (to G. W. C.) and EY09368 (to J. L. F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF105150 (bovine mimecan gene) and AF112465 (human mimecan genomic clone).

‡ To whom correspondence should be addressed: Div. of Biology, Ackert Hall, Kansas State University, Manhattan, KS 66506-4901. Tel.: 785-532-6553; Fax: 785-532-6653; E-mail: est@ksu.edu.

1 The abbreviations used are: KSPG, keratan sulfate proteoglycan; bp, base pair(s); b-FGF, basic fibroblast growth factor; EST, expressed sequence tag; GSP, gene-specific primer; kb, kilobase(s); PCR, polymerase chain reaction; PG, proteoglycan; RACE, rapid amplification of reverse transcription-PCR.

This paper is available on line at http://www.jbc.org
were positive with the middle and the 3'-end probes, and one was positive with the 5'-end and intron probes isolated in the initial experiments. The positive clones were further characterized by restriction enzyme mapping, and the overlapping restriction fragments were subcloned into pBluescript (Stratagene) and sequenced.

DNA Sequencing and Analysis—Plasmid DNA was purified using Qiagen kits. Nucleotide sequence was determined by the dideoxy chain termination method (23) using M13 Forward and Reverse primers. Sequences were determined on both strands of overlapping clones. Alignment of nucleotide sequences for determination of the intron/exon boundaries was performed using the program contained in GeneWorks 2.5.1 (IntelliGenetics Corp.). Comparisons with EMBL and NCBI data bases were performed using Basic Local Alignment Search Tool (BLAST) (24) and Signal Scan Data bases (25).

Analysis of Mimecan mRNAs by Reverse Transcription-PCR and 5'-RACE—Two human eyes (from 57- and 92-year-old donors) were obtained from the Missouri Lions Eye Research Foundation, Columbia, MO. The corneas were excised and incubated in Dulbecco's modified Eagle's medium/F-12 serum-free culture medium containing 0.2% trypsin. The corneas were excised and incubated in Dulbecco's modified Eagle's medium/F-12 serum-free culture medium containing 0.2% trypsin. RNA (1 μg) was reverse-transcribed using 10 μM of the bovine GSP12 or human GSP3, with SuperScript II reverse transcriptase (Life Technologies, Inc.) and 0.5 μM dNTPs for 50 min at 42 °C in buffer supplied by the manufacturer. Reactions were terminated by addition of 1 μl of RNase H, followed by incubation at 37 °C for 20 min and heat inactivation (95 °C, 5 min). The cDNA fragments were then amplified by PCR with bovine GSP1 and GSP9, or human GSP1 and GSP2. The sequence and positions of all GSPs used in this study are shown in Table I. All 50-μl PCR amplification reactions contained 10 mM Tris, pH 9, 50 mM KCl, 0.1% Triton 100, 1.5 mM MgCl₂, Taq polymerase (5 units), 0.2 μM dNTPs, and 100 ng of each primer pair and were carried out for 30 cycles (95 °C, 1 min; 55 °C, 1 min; 72 °C, 2 min). Resulting PCR products were analyzed by electrophoresis on 1.5% agarose gels with DNA visualized by ethidium bromide staining. The PCR products were eluted from the gel, subcloned into pGem-T vectors (Promega Corp.) and sequenced.

Transcriptional Start Point(s) Determination—Rapid amplification of cDNA ends using 5'-RACE system version 2.0 (Life Technologies) was applied to determine the 5'-ends of bovine mimecan cDNA (27, 28). Reverse transcription was performed with bovine GSP12. Anchor oligonucleotide provided in the kit was ligated to the 3'-end of the cDNA, and the ligation product was used as template for PCR reactions with a

### EXPERIMENTAL PROCEDURES

**Isolation of Intronic Fragments for the Bovine Mimecan Gene by PCR**—Five uncultured bovine genomic "libraries" were constructed using the Universal Genome Walker Kit (CLONTECH) and used as templates for PCR amplifications. Briefly, the method involves the digestion of genomic DNA with different restriction enzymes that leave blunt ends and subsequent ligation of each pool of DNA fragments to the Genome- Walker adaptor to create uncultured libraries. Nested Adaptor Primers that anneal to the adaptor at one end and nested GSPs that anneal to the mimecan sequence at the other end were then used for the primary and secondary PCR reactions. In addition to these libraries, undigested bovine and human genomic DNAs (CLONTECH) were also used as templates for amplification of intron fragments. Genomic DNA (500 ng) was amplified in 50-μl reactions that contained 500 μM dNTPs, 1.75 mM MgCl₂, 300 mM amounts of each primer pair, and 1 μl of enzyme mix containing thermostable Taq and proofreading Pwo DNA polymerases (Expand Long Template PCR System; Roche Molecular Biochemicals). Amplifications were carried out for 25 cycles, with denaturation at 92 °C for 10 s, annealing at 58 °C for 30 s, and extension at 68 °C for 18 min. The resulting PCR products were resolved by 1% agarose gel electrophoresis. The PCR products from several independent PCR amplifications were subcloned into pGEM-T vector (Promega) for identification by sequence analysis.

**Isolation of the Bovine Mimecan Genomic Clones**—A bovine genomic DNA library in A DASH II (Stratagene) was used for screening. Approximately 10⁶ plaques were screened by standard plaque hybridization (21) with a [α-32P]dCTP-labeled mimecan cDNA probe (9). Three genomic clones remained positive on tertiary screening. These were applied to determine the 5'-end of the cDNA, and the ligation product was used as template for PCR reactions with a

<table>
<thead>
<tr>
<th>Primers for bovine mimecan gene</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sense</strong></td>
<td></td>
</tr>
<tr>
<td>GSP1</td>
<td>5'-CTAATGACGATCCATCTTAGGAAACCC-3'</td>
</tr>
<tr>
<td>GSP2</td>
<td>5'-AAATCAGTATCCAGAATGAGTGGG-3'</td>
</tr>
<tr>
<td>GSP3</td>
<td>5'-CAGGGTTCATGAGGAGCTTCTCC-3'</td>
</tr>
<tr>
<td>GSP4</td>
<td>5'-GTTTGCTGTGCCCTCCCTGCTATACCTG-3'</td>
</tr>
<tr>
<td><strong>Antisense</strong></td>
<td></td>
</tr>
<tr>
<td>GSP5</td>
<td>5'-GCTAGTCAAGCAAAAGGTACGGAATTAGCTTCTTAG-3'</td>
</tr>
<tr>
<td>GSP6</td>
<td>5'-CTAAAGCACTGTTGCTGACTGTGGG-3'</td>
</tr>
<tr>
<td>GSP7</td>
<td>5'-GGATATTGACAGCCTTTATTTAAAGG-3'</td>
</tr>
<tr>
<td>GSP8</td>
<td>5'-TTCGAAATGTTGCTGCTGCTCTATACAGAGG-3'</td>
</tr>
<tr>
<td>GSP9</td>
<td>5'-CCATCCAGGTATTTATCCTCATAATCTTGG-3'</td>
</tr>
<tr>
<td>GSP10</td>
<td>5'-CCAAAATGCATTGTAAATGCGG-3'</td>
</tr>
<tr>
<td>GSP11</td>
<td>5'-CATCATTTTTCTTTCGGGGG-3'</td>
</tr>
<tr>
<td>GSP12</td>
<td>5'-CGAGCGATGTTGCACTTACTATTTTTTCTTCTTGGGGG-3'</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primers for human mimecan gene</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sense</strong></td>
<td></td>
</tr>
<tr>
<td>HGSP1</td>
<td>5'-TCTCATTACACCTCCCTCCATCGG-3'</td>
</tr>
<tr>
<td>HGSP2</td>
<td>5'-CCATCCAGGTATTTATCCTCATAATCTTGG-3'</td>
</tr>
<tr>
<td>HGSP3</td>
<td>5'-GAGTCCTTCTTAAGTTAGTATGCTG-3'</td>
</tr>
</tbody>
</table>

### Table I

**PCR primers used in this study**

<table>
<thead>
<tr>
<th>Primer sequence</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSP1</td>
<td>5'-CTAATGACGATCCATCTTAGGAAACCC-3'</td>
</tr>
<tr>
<td>GSP2</td>
<td>5'-AAATCAGTATCCAGAATGAGTGGG-3'</td>
</tr>
<tr>
<td>GSP3</td>
<td>5'-CAGGGTTCATGAGGAGCTTCTCC-3'</td>
</tr>
<tr>
<td>GSP4</td>
<td>5'-GTTTGCTGTGCCCTCCCTGCTATACCTG-3'</td>
</tr>
<tr>
<td>GSP5</td>
<td>5'-GCTAGTCAAGCAAAAGGTACGGAATTAGCTTCTTAG-3'</td>
</tr>
<tr>
<td>GSP6</td>
<td>5'-CTAAAGCACTGTTGCTGACTGTGGG-3'</td>
</tr>
<tr>
<td>GSP7</td>
<td>5'-GGATATTGACAGCCTTTATTTAAAGG-3'</td>
</tr>
<tr>
<td>GSP8</td>
<td>5'-TTCGAAATGTTGCTGCTGCTCTATACAGAGG-3'</td>
</tr>
<tr>
<td>GSP9</td>
<td>5'-CCATCCAGGTATTTATCCTCATAATCTTGG-3'</td>
</tr>
<tr>
<td>GSP10</td>
<td>5'-CCAAAATGCATTGTAAATGCGG-3'</td>
</tr>
<tr>
<td>GSP11</td>
<td>5'-CATCATTTTTCTTTCGGGGG-3'</td>
</tr>
<tr>
<td>GSP12</td>
<td>5'-CGAGCGATGTTGCACTTACTATTTTTTCTTCTTGGGGG-3'</td>
</tr>
</tbody>
</table>

Footnotes:

a From GenBank EST data base accession no. N66025.
b From GenBank EST data base accession no. AA045385.
c From GenBank EST data base accession no. W61418.
nested GSP5 and a primer complementary to the anchor sequence. The amplified products were cloned into pGEM-T and sequenced.

The transcriptional start site of the bovine mimecan gene also was analyzed by primer extension. GSP8 and GSP12 were end-labeled with [γ-32P]ATP (3000 Ci/m mole, NEN Life Science Products). Twenty-five μg of RNA was mixed with 5 X 10^4 cpm of labeled oligonucleotide, dried, and redissolved in 30 μl of hybridization solution containing 80% deionized formamide (26). Hybridization was carried out at 90 °C overnight. Primer extension was performed with 40 units of avian myeloblastosis virus reverse transcriptase (Promega) for 90 min at 42 °C. The extended products were analyzed on 8% sequencing gels.

Genomic DNA Hybridization—Ten μg of genomic DNA was digested with an appropriate restriction endonuclease, fractionated in a 0.8% agarose gel and transferred onto nylon membrane (GeneScreen Plus, NEN Life Science Products) as described previously (29, 30). The membranes were prehybridized for 1 h at 65 °C in a solution containing 0.2% polyvinylpyrrolidone, 0.2% Ficoll, 0.2% bovine serum albumin, 0.05 μl Tris-HCl, pH 7.5, 1 μl NaCl, 0.1% sodium pyrophosphate, and 1% SDS. The cDNA probes were generated by PCR from pBScDNA clone (9), [α-32P]dCTP-labeled using Prime-a-Gene labeling system (Promega), and used at 2 X 10^6 cpm/ml of the above solution. Hybridization was carried out overnight at 65 °C. Following hybridization, the filter membranes were washed and exposed to X-Omat AR film (Kodak Scientific Imaging System) at −70 °C for 1–3 days.

Analysis of the Expression of Spliced Mimecan mRNAs in Cultured Bovine Keratocytes—Bovine corneal keratocytes were isolated from the central portions of fresh corneas as described previously (30, 31). Cells were incubated at 2 X 10^6 cells/cm² in 75 cm² T flasks (Costar) into Dulbecco’s modified Eagle’s medium/F-12 medium with antibiotics (100 μg/ml each penicillin and streptomycin, 50 μg/ml gentamycin, and 2.5 μg/ml amphotericin B). Incubation continued in the above medium containing 0.1% platelet-poor horse serum (Sigma catalog no. P5552) and ITS (5 μg/ml insulin, 5 μg/ml selenous acid, 5 μg/ml transferrin; Collaborative Research Inc.) and b-FGF (Sigma) at 10 ng/ml was added to the culture. Incubation was continued for up to 6 days. At the indicated time periods, cells were washed twice with phosphate-buffered saline and total RNA was isolated. RPAs were performed on 1 μg of RNA using the following plasmid clones for the synthesis of antisense RNA probes: the 662-bp dIII restriction fragment from bovine mimecan cDNA (positions 298–960) cloned into pGEM-1 vector; the BglIII-NruI DNA fragment from bovine lumican (positions 548–1099) cloned into the BamHI-SmaI sites of pGEM-1; and a 350-bp DNA fragment obtained by PCR with primers complementary to bovine glyceraldehyde-3-phosphate dehydrogenase cDNA (GenBank accession no. U85042) and cloned into a pGEM-T vector (9, 10). RPAs were performed as described (9, 10). Radioactive bands were evaluated by autoradiography and laser densitometry using an ImageQuant PhosphoImager (Molecular Dynamics). The relative amounts of RNA were calculated as the ratio of the intensity of the protected band derived from the corresponding message to that of the protected band derived from the message for the internal standard (glyceraldehyde-3-phosphate dehydrogenase), after correcting for the variations in the length of all protected bands. These ratios were then plotted using the program contained in DeltaGraph® Professional.

RESULTS AND DISCUSSION

Identification of Two Splice Acceptor Sites in a 527-bp Exon of Bovine Mimecan Gene—The initial strategy used to determine the structure of the bovine mimecan gene was the cloning of genomic fragments obtained from genomic DNA by PCR. Having previously demonstrated a 278-bp differentially spliced region of mimecan cDNA, between nucleotides 62 and 340 (10), we chose to look for introns on both sides of this region. Using primer pair GSP1 and GSP5, designed to amplify the upstream intron, a PCR product of 765 bp was obtained from genomic DNA. The fragment was 658 bp larger than the distance between the two primers on cDNA, indicating that indeed the upstream intron was amplified. The results of these experiments are shown in Fig. 1. Similar results were obtained with additional primer pairs (GSP1-GSP6 and GSP1-GSP7). Sequence analysis of these PCR products and alignment to bovine mimecan cDNA located the upstream intron exactly as predicted, between nucleotides 62 and 63. Similarly, other primer pairs (GSP2 and GSP11, GSP3 and GSP11, GSP4 and GSP11) designed to amplify the downstream intron gave PCR products whose sizes were larger than the distance between primers on cDNA indicating that the downstream intron was amplified. Much to our surprise, sequence analysis of these PCR products and alignment to mimecan cDNA showed that the downstream intron of 1.41 kb is located between nucleotides 527 and 528 of mimecan cDNA, whereas the expected position was between nucleotides 340 and 341. Additional PCR with primers GSP4 and GSP9 gave a product of 322 bp, a size corresponding to the distance between primers on cDNA and therefore confirming the lack of an intron between nucleotides 340 and 341. At least three possibilities could explain these results. First, secondary DNA structures that make it difficult to obtain certain regions directly from genomic DNA; second,mimecan-related sequences, such as pseudogenes, may have been amplified; and third, a splice site within the 527-bp exon may exist. To test for the first possibility, as template for PCR amplifications, we used genomic DNA that was digested with restriction enzymes, i.e.
mains unchanged. A similar case is described in the human mimecan genes (33). In contrast, the RNA isoforms generated from bovine mimecan is expressed in a tissue-specific or developmental manner (32, 34). These express two RNA transcripts, but the translated domains are distinct splice donor sites (34). These are alternatively used to generate multiple transcripts, but the translated domains require different splicing mechanisms.

The extremely conserved terminal (AG) and the branch site (A) sequences can be found at both positions in bovine and human mimecan genes: at the intron/exon junction, as well as within the 527-bp exon, exactly 278 bp downstream of its 5′-splice site and the branch site consensus sequences in nuclear pre-mRNA introns (38, 40) are boxed by a line. The extremely conserved terminal (AG) and the branch site (A) nucleotides are underlined. N, any nucleotide; R, purine; Y, pyrimidine; P, pyrimidine.

There are examples of variant forms that share common 5′- and 3′-exons but differ in their internal “mini” exons. Thus, in the homoeotic gene Ultrabithorax of the fly, Drosophila melanogaster, two separate splice donor sites at the end of the common 5′ exon are situated 27 bp apart (32). In most cases described so far, these multiple transcripts encode different protein isoforms that are expressed in a tissue-specific or developmental manner (32, 33). In contrast, the RNA isoforms generated from bovine mimecan encode a protein whose primary structure remains unchanged. A similar case is described in the human α2(IV) collagen gene, in which a leader exon contains three distinct splice donor sites (34). These are alternatively used to generate multiple transcripts, but the translated domains remain unchanged, as in the bovine mimecan gene. Untranslated regions are likely to be involved in the regulation of protein expression. For example, the usage of alternative 5′-untranslated regions has been shown to discriminate the translation of mRNA coding for human insulin-like growth factor II (35). However, the biological significance of multiple mimecan transcripts remains to be determined.

Identification of Differentially Spliced Human Mimecan mRNAs—We next attempted to determine whether this alternative splice site is similarly used in the human mimecan gene. Comparison of the bovine mimecan cDNA with homologous human sequences available in the GenBank data bases resulted in the identification of several overlapping EST clones. A homology >85% was found between the 5′-region of bovine mimecan cDNA and four clones corresponding to the 5′-end of human mimecan cDNA (accession nos. W21015, T97846, R06469, and N50625). Interestingly, all four clones contained the shorter mRNA transcript where the nucleotides corresponding to the 278-bp region, subjected to differential splicing in bovine mimecan gene, were missing. Because the results of this search suggested a similar alternative splicing mechanism in the human mimecan gene, we performed experiments to detect it. First, we used human genomic DNA to PCR amplify the upstream intron. With HGSP1 and HGSP2, an amplicon of 1338 bp was obtained, cloned, and sequenced. This sequence has been deposited in GenBank under accession no. AF112465. As expected, this amplicon contained the upstream intron, as well as the sequence corresponding to the 278 bp that were missing in the EST clones. In the human mimecan gene, this sequence also appeared to be a part of a larger exon corresponding to the 527-bp exon of the bovine mimecan gene. Next, we demonstrated that the two spliced RNAs are transcribed from the human mimecan gene. For these experiments, we used total RNA from human corneal keratocytes and skeletal muscle cells. We reasoned that, if the tissue-specific pattern of expression of the human mimecan gene were similar to that of its bovine counterpart, then we should detect two transcripts in RNA from cornea and mainly the full-length message in muscle, as shown previously for bovine mimecan messages (10). As predicted, two amplicons were obtained from corneal cDNA, whereas only the larger band was visible in the reactions with muscle cDNA (Fig. 3). The authenticity of these PCR products was verified by sequence analysis. These results were consistent with the data obtained from the bovine mRNA samples and demonstrated that an alternative 3′-splice acceptor site within the 527-bp exon was utilized in corneal keratocytes to generate two RNA transcripts from the human mimecan gene as well.

Isolation of Genomic Clones for the Bovine Mimecan Gene—To obtain the complete structure of the gene and to confirm our data on alternative splicing within the 527-bp exon, we screened a phage bovine genomic library using a mimecan cDNA probe. Three positive clones were obtained and characterized by restriction enzyme mapping and Southern blotting. The three clones represented a 21,794-kb genomic DNA encoding bovine mimecan, including a 7.2-kb 5′-flanking region. The schematic of the structure of the bovine mimecan gene is shown in Fig. 4. The overlapping restriction fragments spanning this continuous sequence were subcloned into pBlue-
script and sequenced in both directions. The entire sequence has been deposited in GenBank nucleotide sequence database with the accession number AF105150. The number of intervening introns and their positions were determined by alignment of genomic sequences with published cDNA sequences. Analysis of this alignment revealed that the known cDNA is interrupted by six introns. The nucleotide sequences of the 527-bp exon and the two introns determined in phage clone were identical to the sequences obtained in the initial experiments.

Southern Blot Analysis—To determine whether bovine mimecan is encoded by a single gene, we performed Southern blot analysis of bovine genomic DNA digested with the restriction enzymes indicated in Fig. 5. The pattern of hybridization using bovine mimecan cDNA as a labeled probe and the lengths of the restriction fragments that hybridized to this probe corresponded to those calculated from the gene sequence. Thus, in total, three lines of evidence establish that the differentially spliced mRNA transcripts are generated by the use of an alternative 3'-splice acceptor site within the exon of 527 bp and that the bovine mimecan is encoded by a single gene.

Identification of the Transcription Initiation Site(s)—Two primer extension experiments were carried out with two antisense oligonucleotides located downstream of the translation initiation codon ATG, and these are shown in Fig. 6. Primer GSP8 starts 42 nucleotides downstream of the translation initiation site. Primer GSP12 was designed to cross two exons and starts 187 nucleotides downstream of the first ATG codon. With this experimental design, the extension of a genomic DNA, a frequent contaminant of mRNA preparations, should be eliminated. The difference in size between the extension products obtained with each of the two primers should match the distance of the region between them. In addition, because both primers were located downstream of the position where the differential splicing occurred, each one of them should allow us to detect two extended products, corresponding to the two spliced mRNA messages with a 278-bp expected distance between them, i.e. equal to the size of the differentially spliced region. As shown in Fig. 6, both primers gave the two extension products with sizes that differed from each other almost exactly as predicted, indicating a transcription start site 390 bp upstream of the ATG translational initiation codon. Surprisingly, both primers also rendered two additional extension products with sizes 160 bp longer than expected when compared with the primer positions on bovine mimecan cDNA. These results indicated that additional nucleotide sequence, previously undetected, was present at the 5'-end of bovine mimecan gene. The putative new exon contained the major site for transcription initiation, as judged by the intensity of the extended products. To clarify the results from primer extensions and to obtain the nucleotide sequence of the new 5'-end, we used 5'-RACE technique. Reverse transcription was conducted using GSP12, the primer used in primer extension experiments. PCR amplification was done with GSP5, which starts 103 bp downstream of the reported 5’-end of mimecan cDNA, and the anchor primer provided in the kit. An amplicon of about 270 bp was obtained in first PCR amplification. A smaller band of about 120 bp was detected only in secondary PCR amplification. The longer PCR product was cloned, and four clones were sequenced. In all four clones, the sequence data obtained were identical. Their 103-bp 3’-end sequence overlapped and was
100% identical to the 5’-end sequence of reported mimecan cDNA, indicating that these clones contained the bovine mimecan 5’-extended product. The sequence is shown in Fig. 6A. An additional 13 nucleotides upstream to the known cDNA end overlapped with the genomic DNA sequence, and therefore we extended the size of the corresponding exon and labeled it exon 2. The remaining 160 nucleotides of the RACE clone were not present in the 7-kb genomic sequence upstream to exon 2, thus indicating that a large intron is separating these two exons. To ascertain further that exon 1 is indeed a part of the bovine mimecan gene, we performed RPA on corneal RNA, using as probe the antisense RNA transcribed from the RACE clone. Two protected bands corresponding to the two transcription start sites were detected (data not shown). The results from RPA were consistent with the results from 5’-RACE and primer extension experiments and confirmed that the new sequence (exon 1) is a part of bovine mimecan mRNA. As previously reported, Northern blot analyses utilizing total RNA from bovine tissues, showed three mimecan transcripts of 2.4, 2.5, and 2.6 kb (10). The length of the longest transcript on Northern blot corresponds well to the length of the cDNA sequence obtained here, including the new 160 bp at the 5’-end and previously reported additional 398 bp at the 3’-end of bovine mimecan cDNA (10). It is interesting to note that even larger transcripts: 3.6 kb in human osteosarcoma cell lines (2), 3.7 kb in various mouse tissues (11), and 3.5 kb in rat aortic RNA (12), have been detected on Northern blots by other researchers. We performed a GenBank data base search to determine whether the nucleotide sequence of the newly detected exon was present in other published sequences. A number of sequences showed homology to mimecan exon 1. The level of about 30% similarity was found with human parathyroid hormone-like hormone mRNA, bovine SCO-spondin protein (subcommissural organ) mRNA, and pig-ferredoxin mRNA. In addition, sequences producing significant matches were found in the promoter regions of human epidermal growth factor receptor P1 and P2 genes, β-tubulin β2 gene, asparagine synthetase, and PRBP (protamine-1 RNA-binding protein 1) genes as well as mouse c-Ki-ras gene. These data indicated that the new exon contained sequences conserved across species while their location in promoter regions of other genes suggested important regulatory functions. We attempted to determine the approximate size of the first intron. The 274-bp RACE product was used as a probe on Southern blots with genomic DNA. Unfortunately, but as expected from the nucleotide sequence of exon 1 (CG content reaches 80%), this probe gave strong cross-hybridization to microsatellite DNA, so that the pattern of visible bands on autoradiographs and on ethidium bromide-
stained agarose gels of electrophoretically separated DNA was identical. The results from Southern blot demonstrated that this region was not suitable for use as a probe for rescreening the genomic library in order to obtain genomic clones of this region. Similarly, primers based on the exon 1 sequence were not suitable for PCR amplification and direct determination of exon/intron boundaries. Therefore, we chose a genome walking strategy using as template the DNA from our uncloned libraries. A 13.5-kb region of intron 1 was obtained and characterized. However, so far this contains no region homologous to exon 1. This may not be surprising when considering two facts: first, that the sizes of the largest introns (>13.2 kb in decorin gene (15) and >15 kb in mouse PG-Lb (36)) in related PGs have been determined approximately, without isolation of overlapping genomic clones; and second, that some eukaryotic genes have very large introns (64 kb in the case with human thyroglobulin gene (37)). Therefore, we cannot predict the size of the first intron in bovine mimecan gene. However, because an AC dinucleotide at the 3’-terminus of intron 1 can be aligned, we believe that exon 1 is a continuous part of the mimecan gene. The AT-AC introns have been reported only in a limited number of genes (38). These include tissue-specific regulated genes, such as the gene encoding cartilage matrix protein (39). In addition, the nucleotide sequence of human EST clone N56025 reveals that the corresponding intron/exon junction in human mimecan gene is almost identical to the sequence of this region in bovine mimecan gene. In both, an AC dinucleotide can be aligned at the intron 1 terminus, indicating that this different class of intron is also conserved in these two species. Interestingly, the human EST clone aligns with the bovine intron 1 sequence an additional 11 nucleotides upstream of exon 2. We emphasize, however, that these 11 nucleotides were not present in the bovine cDNA obtained by 5’-RACE and therefore they are a part of the untranscribed region, i.e. intron 1 in the bovine mimecan gene. Human genomic DNA amplification could explain the presence of these nucleotides in the EST clone sequences. However, at this time we cannot rule out the possibility of a trans-splicing event involved in the generation of the longer mimecan message. The origin and location of exon 1 of bovine mimecan gene remain under investigation in our laboratory.

**Bovine Mimecan Gene Structure**—The exon-intron organization of the gene is summarized in Table II. The gene is spread over >33 kb of continuous DNA sequence. Since the position of first exon remains to be determined, it is likely that the gene is larger in size. Exons 1 and 2 are composed of 5’-untranslated sequences. Exon 3 is the 527-bp exon described in the initial experiments and contains the two alternatively used splice acceptor sites. The starting codon is located 75 bp downstream of the second splice acceptor site. Thus, this exon contains 353 bp of 5’-untranslated sequence and 174 bp of coding sequence that includes the putative signal peptide and the N-terminal amino acids of the mature protein. For comparison, the N-terminal amino acid of corneal KS PG25 is located in the fourth exon and the N-terminal amino acid of bone osteoglycin is located in the sixth exon. Exons 5–8 encode the seven leucine-rich repeats found in bovine mimecan. Exon 5 contains the cysteine-rich cluster and the first leucine-rich repeat. Three repeats are encoded in exon 6, one in exon 7, and one in exon 8. The leucine-rich repeat 6 is split in the middle by an intron of 812 bp separating exons 7 and 8. The last exon contains the C terminus and the 3’-untranslated region, including the two polyadenylation signals previously demonstrated to be alternatively used. The first potential site for N-glycosylation and the proposed site for keratan sulfate attachment is located in exon 7. The other two potential N-glycosylation sites are located in exon 8. Introns are diverse in size, ranging from 739 bp (intron 2) to over 17 kb (intron 1). With the exception of intron 1, all acceptor and donor splice sites conform to the GT-AG rule (40).

Inspection of the 7-kb DNA sequence upstream of the second transcription initiation site in exon 2 revealed a TATA box 300 nucleotides upstream of the transcription initiation site, a distance much longer than the usual 25–30 nucleotides for active TATA boxes. However, it contains initiator elements, CG boxes, and binding sites for a number of transcription factors that may direct transcription initiation as in other genes (41, 42). Interestingly, 18 binding sites for transcription factor Pit-1 (43) can be identified in the mimecan intron 1 sequence so far determined. Clusters of binding sites for Pit-1 have been previously detected in promoter regions of bovine and human keratan genes (26, 30). Because of the unique expression of keratan in the cornea, the presence of similar binding motifs in the region flanking the second transcription initiation site suggests regulatory function. However, the functional motifs in the promoter region of the bovine mimecan gene remain to be determined in future studies.

From the data presented so far, the 5’-untranslated region of the bovine mimecan gene displays polymorphism; there are two transcription initiation sites and alternative splicing by use of two distinct splice acceptor sites in the third exon. In conjunction with our previous report of alternative use of two polyadenylation sites, these data provide proof for eight different mimecan cDNAs expressed in the same tissue and yet maintaining the same amino acid coding sequence. The most abundant mRNA isoform includes the exon 1 described in this report, as judged by the intensity of the bands in primer extension and RACE experiments. The same RNA isoform is spliced to delete the 278-bp sequence of exon 3, as judged by the intensity of the bands in RPA experiments. Three low abundance bovine cornea mRNA isoforms are the full-length tran-
scripts from exon 1 and exon 2 and the spliced variant starting from exon 2. Functional importance of these RNA transcripts awaits development of an expression system.

Comparison of the exon/intron organization of the bovine mimecan gene reported here with the partial genomic structure of the human mimecan gene also reported here and the single described exon of the mouse mimecan gene (11) shows that the exon/intron boundaries are highly conserved among these species. Analyses of the evolutionary relationships between the SLRP core proteins show that mimecan is most closely related to PG-Lb (13, 14). Recently, the gene for mouse PG-Lb has been reported to consist of seven exons spread over 50 kb of genomic DNA sequence (36). When comparing the genes of the bovine mimecan and mouse PG-Lb, we note with interest that the last four exons of the bovine mimecan gene and the mouse PG-Lb gene have identical sizes, although the intron lengths differ. The major difference between the two genes is in their 5′-untranslated regions. This is in agreement with the concept that PG-Lb/epiphycan (44) and mimecan have a common an-cestor. Most likely, the primordial gene has been duplicated and shuffled to different locations, acquiring introns and different sequences at their upstream non-coding regions. In this regard, the different tissue-specific expression of mimecan and PG-Lb gene can be attributed to the different regulatory sequences at their 5′ non-coding regions.

Expression of Spliced Mimecan mRNAs in Cultured Bovine Keratocytes upon Treatment with b-FGF—Because all eight mimecan transcripts were detected in bovine corneal kerato-cytes, we asked why multiple transcripts were present at the same time, in the same tissue, coding for the same unchanged protein. We hypothesized that each one of these classes of mRNA molecules may play an unique role toward maintaining transparency and integrity of the cornea. This could be accomplished by either RNA-RNA or RNA-protein interactions. As an initial step toward answering the question, we examined the expression of two mimecan isoforms in cultured bovine kerato- cytes. In view of the recent finding that b-FGF down-regulates mimecan expression in vascular smooth muscle cells (12), we also examined the effect of this growth factor on mimecan differentially spliced isoforms expression. Mimecan was as-sayed with the probe that allowed us to detect two of the spliced transcripts simultaneously (the transcript containing the 278 bp of exon 3 and the transcript lacking the 278 bp of exon 3) and, thus, to detect possible differences in their expression. In addition to mimecan, lumican RNA levels also were deter-mined. Lumican, similarly to mimecan, is a KSPG in the cornea, whereas in other connective tissues both proteins are present as glycoproteins lacking the keratan sulfate chains. The levels of expression of gliceraldehyde-3-phosphate dehydrogenase mRNA served as control throughout these experiments and were used as a standard for normalizing the intensity of the signals of other RNA samples on autoradiographs. The culture conditions were those previously shown to allow bovine keratocytes to express corneal keratan sulfates in a stable manner (31). The maximum effect of b-FGF was ob-tained at 10 ng/ml in preliminary experiments, and this concentra-tion was used. An autoradiograph of a representative RPA is shown in Fig. 7A, and a plot of mean values of relative amounts of RNAs, obtained from three independent experi-ments, is shown in Fig. 7B. In vivo, as shown with RNA from freshly isolated corneal keratocytes, the shorter mimecan tran-script was the major product. In culture, the expression of the shorter message was slightly up-regulated but remained the major mimecan transcript in RNA from cells cultured for up to 6 days in the absence of b-FGF. The longer mimecan transcript was low in RNA from fresh keratocytes and remained low in RNA from 6 days of culture in the absence of b-FGF. The lumican expression in the absence of b-FGF was similarly not significantly changed. Because the cells under the described culture conditions maintain their differentiated phenotype and continue to secrete KSPG in the medium, the high levels of shorter mimecan isoform suggest a functional importance of this transcript for corneal physiology. In addition, the expres-sion pattern of mimecan isoforms and lumican can be used as marker for differentiated corneal keratocytes. Six hours after adding b-FGF to the culture medium, there were detectable changes in mimecan isoforms expression. The level of longer transcript increased, in contrast to the level of the shorter message, which was down-regulated. At 12 h through 6 days, the patterns of expression of both isoforms were similar with highest levels at day 6. The expression of lumican mRNA was similar to that of the shorter mimecan RNA transcript; after initially low expression at 6 h, highest levels were found at day 6. These results demonstrate that shorter mimecan transcript and lumican are coordinately regulated in corneal keratocytes by b-FGF, with low levels in its presence after 6 h and high levels in its presence after 6 days of incubation. Although further studies are needed to determine the significance of these findings, our data that b-FGF differentially affects the levels in its presence after 6 days of incubation. Although further studies are needed to determine the significance of these findings, our data that b-FGF differentially affects the expression of two spliced mimecan mRNA isoforms in cultured corneal keratocytes support the idea of possible individual role for each of these molecules. In addition, the results presented here may be of relevance in future studies on determination of the role of PGs in mediating the complex biological actions of the FGF family of growth factors (45–47).

Conclusions—We have described the structure of the bovine mimecan gene and elucidated the molecular basis for hetero-
geneity of bovine mimecan mRNA transcripts. We have also described the partial structure of the human mimecan gene and demonstrated that two mRNA isoforms are generated by identical mechanisms in both species. It seems likely that the transcript resulting from initiation in the first exon and splicing out the 278 bp of the third exon is significant for corneal physiology, because this splice form is expressed at high levels in bovine cornea. The information obtained in this study will allow further functional characterization of the individual elements involved in the regulation of bovine mimecan gene expression.

Acknowledgments—We thank Prof. Beth Montelone for helpful discussions and critical reading of this manuscript. We gratefully acknowledge the assistance of Dr. Ron Walkenbach and Tina Livesay, Missouri Lions Eye Research Foundation, in providing the human eyes used in this study.

REFERENCES

The Bovine Mimecan Gene: MOLECULAR CLONING AND CHARACTERIZATION OF TWO MAJOR RNA TRANSCRIPTS GENERATED BY ALTERNATIVE USE OF TWO SPLICE ACCEPTOR SITES IN THE THIRD EXON

Elena S. Tasheva, Martha L. Funderburgh, James McReynolds, James L. Funderburgh and Gary W. Conrad

doi: 10.1074/jbc.274.26.18693

Access the most updated version of this article at http://www.jbc.org/content/274/26/18693

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 46 references, 22 of which can be accessed free at http://www.jbc.org/content/274/26/18693.full.html#ref-list-1