Identification and Expression in Mouse of Two Heparan Sulfate Glucosaminyl N-Deacetylase/N-Sulfotransferase Genes*

(Received for publication, February 26, 1997, and in revised form, December 11, 1997)

Marion Kusche-Gullberg‡§, Inger Eriksson¶, Dagmar Sandbäck Pikas‡, and Lena Kjellén¶

From the §Department of Medical Biochemistry and Microbiology, University of Uppsala, S 751 23 Uppsala, and the ¶Department of Veterinary Medical Chemistry, Swedish University of Agricultural Sciences, 75207 Uppsala, Sweden

The biosynthesis of heparan sulfate/heparin is a complex process that requires the coordinate action of a number of different enzymes. In close connection with polymerization of the polysaccharide chain, the modification reactions are initiated by N-deacetylation followed by N-sulfation of N-acetylgalcosamine units. These two reactions are carried out by a single protein. Proteins with such dual activities were first purified and cloned from rat liver and mouse mastocytoma. The mouse mastocytoma enzyme is encoded by an ~4-kilobase (kb) mRNA, whereas the rat liver transcript contains ~8 kb. In the present study, the primary structure of the enzyme encoded by the mouse 8-kb transcript is described. It is demonstrated that both the 4-and 8-kb transcripts have a wide tissue distribution and that they are encoded by separate genes. Characterization of the gene encoding the 4-kb transcript demonstrates that it spans a region of about 8 kb and that it contains at least 14 exons. The similarity of this gene and the previously characterized human gene for the 8-kb transcript is discussed.

Heparan sulfate (HS)† proteoglycans occur on cell surfaces and in the extracellular matrix of loose connective tissue and in basement membranes. In connective tissue-type mast cells, the polysaccharide is stored intracellularly in granules. The mast cell polysaccharide is traditionally referred to as heparin but is composed of the same structural units as HS. HS has been implicated in cell adhesion processes, cytokine action, regulation of enzyme activities and in the maintenance of the permeability of basement membranes (reviewed in Refs. 1–3). The extent of these reactions varies, giving rise to enormous structural heterogeneity. Compared with HS, heparin is more highly sulfated and has a higher content of N-sulfated glucosamine and iduronic acid (3). The first modifying enzyme, a combined N-deacetylase/N-sulfotransferase has a prime regulatory role in determining the overall structure and charge density of HS as N-deacetylation followed by N-sulfation are required for all subsequent modifications (1).

The enzyme purified from the mouse mastocytoma is encoded by an ~4-kb mRNA while the rat liver transcript contains ~8 kb. The major part of the protein sequences have ~70% identity. However, the most N-terminal 80 amino acids show a low level of structural homology, with approximately 30% identical amino acids. In addition, both the 5′ and the 3′ noncoding regions seem unrelated (9, 10, 14). In this study, we demonstrate that the two transcripts are widely distributed in mouse tissues and that they are encoded by different genes.

During biosynthesis, the HS-chains are polymerized by sequential addition of glucuronic acid and N-acetylgalcosamine units. Concomitantly, the polymer is modified through a series of reactions that include N-deacetylation and N-sulfation of glucosamine residues, epimerization of glucuronic acid to iduronic acid, and finally O-sulfation at various positions (3, 6, 7).

The enzyme purified from the mouse mastocytoma is encoded by an ~4-kb mRNA while the rat liver transcript contains ~8 kb. The major part of the protein sequences have ~70% identity. However, the most N-terminal 80 amino acids show a low level of structural homology, with approximately 30% identical amino acids. In addition, both the 5′ and the 3′ noncoding regions seem unrelated (9, 10, 14). In this study, we demonstrate that the two transcripts are widely distributed in mouse tissues and that they are encoded by different genes.

The structure of the gene encoding the 4-kb transcript is discussed. The gene encoding the 8-kb transcript, whereas probes B and D hybridize with the coding and 3′ untranslated region of the 8-kb transcript. Probe A is a HindIII cDNA fragment of the 4-kb mouse transcript (nt 1412–1900 in Fig. 1 of Ref. 10). Probe B was generated by RT-PCR of rat liver RNA using the same conditions as described for the 504-bp PCR-product (10), with a sense and an antisense primer corresponding to nt 1639–1660 and 2025–2004, respectively, in the rat liver cDNA-sequence (14). To the 5′-end of the primers was added a dCCGAATTTC extension to create an EcoRI site in the PCR-product. Probe C is a 356-bp cDNA-fragment, corresponding

*This work was supported by grants from the Swedish Medical Research Council (6525 and 12074). Konung Gustav V:s 80-årsfond, Magnus Bergvalls stiftelse, Stiftelsen Lars Hiertas minne, Polysaccharidforsking AB (Uppsala, Sweden), and the European Commission Contract No. B104-CT95-0026. 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.

†Heparan sulfate (HS) proteoglycans occur on cell surfaces and in the extracellular matrix of loose connective tissue and in basement membranes. In connective tissue-type mast cells, the polysaccharide is stored intracellularly in granules. The mast cell polysaccharide is traditionally referred to as heparin but is composed of the same structural units as HS. HS has been implicated in cell adhesion processes, cytokine action, regulation of enzyme activities and in the maintenance of the permeability of basement membranes (reviewed in Refs. 1–3). The extent of these reactions varies, giving rise to enormous structural heterogeneity. Compared with HS, heparin is more highly sulfated and has a higher content of N-sulfated glucosamine and iduronic acid (3). The first modifying enzyme, a combined N-deacetylase/N-sulfotransferase has a prime regulatory role in determining the overall structure and charge density of HS as N-deacetylation followed by N-sulfation are required for all subsequent modifications (1).

‡To whom correspondence should be addressed: Dept. of Medical Biochemistry and Microbiology, University of Uppsala, The Biomedical Center, Box 575, S-751 23, Uppsala, Sweden. Fax: 46-18-471 42 09; E-mail: Marion.Kusche@medkem.uu.se.

§The abbreviations used are: HS, heparan sulfate; bp, base pair(s); kb, kilobase pair(s); nt, nucleotide(s); RT-PCR, reverse transcriptase-polymerase chain reaction; SSPE, saline/sodium phosphate/EDTA.

¶The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM EBI Data Bank with accession number(s) AF049894.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBankTM EBI Data Bank with accession number(s) AF049894.

This paper is available on line at http://www.jbc.org
to the 3’ noncoding end of the 4 kb mouse mastocytoma transcript (nt 2951–3307; Ref. 10). Probe D is a Hinfl fragment of the 504-bp PCR product generated by RT-PCR of rat liver RNA (10), corresponding to nt 3396–3679 in Ref. 14.

RNA Purification—Hepatocytes were isolated after collagenase perfusion of a rat liver and purified by density gradient centrifugation in Percoll (17). Total RNA was isolated from these cells and from mouse mastocytoma (18) and mouse liver tissue using the LiCl/urea/SDS method (19).

Northern Blot Analysis—A single filter containing poly(A)-selected RNA from several adult mouse tissues (CLONTECH) was hybridized at 65 °C in 5× SSPE (1× SSPE is 0.15 M NaCl, 10 mM NaH2PO4, 1 mM EDTA, pH 7.4), 10× Denhardt’s solution, 2% SDS, 100 μg/ml salmon sperm DNA, at three occasions with probes labeled with [α-32P]dCTP (NEN Life Science Products) as described above. The probes used were a 356-bp cDNA-fragment, identical to the 356-bp cDNA fragment (Probe C) used in Southern blotting (see above), a full-length cDNA corresponding to an mRNA encoding a protease specific for connective tissue-type mast cells, kindly provided by Dr. Lars Hellman, University of Uppsala (MMCP-4; Ref. 20), and a 2-kb cDNA recognizing β-actin mRNA (CLONTECH).

A similar filter was hybridized, as described above, to a 216-bp fragment that had been amplified from mouse liver RNA by RT-PCR using primers corresponding to nt 2103–2123 (sense) and nt 2318–2298 (antisense), respectively, in the nucleotide sequence of the 8-kb rat liver transcript (14). The identity of the amplified product was established by nucleotide sequence analysis, using [α-32P]dATP and the Sequenase kit (U. S. Biochemical Corp.).

Cloning and cDNA Analysis of Mouse Liver 8-kb Transcript—An RT-PCR-based strategy was employed to clone the mouse homologue of the rat liver N-deacetylase/N-sulfotransferase. Primers were selected in regions corresponding to conserved sequences in the rat (14) and the human (15) transcripts. One pair of primers (1) corresponding to nt 1–18 (sense) and 1199–1182 (antisense), respectively, in the rat liver cDNA-sequence (14) were used to amplify the 5’ half of the cDNA. An overlapping clone extending to the 3’ untranslated region was obtained using two primers (2) corresponding to nt 1147–1163 (sense) and nt 2717–2700 (antisense), respectively, in Ref. 14. Random hexamer primed cDNA was synthesized from 1 μg of mouse liver total RNA using the GeneAmp RNA PCR kit (Perkin-Elmer). The PCR reaction for primer pairs 1 was performed under conditions of: 1 cycle of 94 °C for 1 min, 30 cycles each of 94 °C for 15 s, 50 °C for 15 s, 72 °C for 1 min, and a final extension at 72 °C for 2 min. The amplified 1199-bp fragment was subcloned into the pUC 119 vector and sequenced. Conditions used to amplify the second part of the cDNA using primer pairs 2 were: 1 cycle of 94 °C for 1 min, 35 cycles each of 94 °C for 20 s, 45 °C for 20 s, 72 °C for 2 min plus 1 s for each cycle, and a final extension at 72 °C for 5 min. The amplified product (1571 bp) was subcloned into the pCR II vector (Invitrogen) and sequenced. The sequence identities between the mouse liver and rat liver 8-kb transcript are 96% for both the 1199 and 1571-bp fragments (EBI/GenBank™ accession number AF049894). To obtain a full-length clone, a unique NdeI site at nt 1178 was inserted. The insert in the pCR II plasmid was excised with NdeI and EcoRI and ligated into the same site in the pUC vector containing the PCR product of primer pairs 1. The sequence around the NdeI site in the resulting full-length clone was verified by nucleotide sequence analysis as described above, using synthetic oligonucleotide primers. Sequences were analyzed with the aid of the Lasergene software package (DNASTAR Inc., Madison).

Isolation of Genomic Clones—A mouse xFIX II genomic library (Stratagene, La Jolla, CA) was screened using as probes a mixture of a 1.7-kb cDNA fragment (10), corresponding to a region entirely within the coding region of the 4-kb transcript, and “probe B” (see above), corresponding to 366 nt of the coding region of the 8-kb transcript. The probes were labeled with [α-32P]dCTP (NEN Life Science Products) as described above. About 1 × 106 plaques were screened, and approximately 60 clones were plaque purified. Of those, two different clones for the 4-kb transcript and 4 different clones for the 8-kb transcript were obtained. The genomic clones were further analyzed by restriction mapping and Southern blotting using as probes specific cDNA oligonucleotide primers (17 nt) that had previously been used for nucleotide sequence analysis of the mouse mastocytoma 4-kb cDNA (10). Genomic DNA fragments obtained from the phage DNA after digestion with SacI and/or BamHI were subcloned into the plasmid vector pUC 119 for characterization. The gene encoding the 4-kb transcript was further characterized, using other restriction fragments of the λ clones that were subcloned into pUC 119 and sequenced. The size of each intron was determined by nucleotide sequence analysis.

RNA Expression Analysis Using RT-PCR—One μg of total RNA from each tissue in a total volume of 20 μl was used for the generation of single-stranded cDNA in the GeneAmp RNA PCR kit. After dilution to 40 μl, 1 μl of each sample was used for PCR in a Rapidcycler (Idaho Technology), with Taq DNA polymerase (MBI-Fermenta) and TaqStart Antibody (CLONTECH). The sense primer corresponds to nt 1301–1317 in the sequence of the mouse mastocytoma 4-kb transcript (10), identical to nt 1593–1609 in the sequence of the rat liver 8-kb transcript (14). Also, the antisense primer corresponds to identical regions in the mouse 4-kb (nt 2299–2292; Ref. 10) and in the rat 8-kb transcript (nt 2594–2577; Ref. 14). The 10 μl reactions were done in heat-sealed glass capillaries in 50 mM Tris buffer, pH 8.3, containing 3 mM MgCl2 and 0.25 mg/ml bovine serum albumin. After an initial hold for 2 min at 95 °C, each cycle included denaturation at 95 °C for 0 s, annealing at 56 °C for 0 s, and extension at 72 °C for 30 s. After 20, 25, or 30 cycles, the capillaries were emptied into Eppendorf tubes, and to the tubes were added 10 μl of “restriction enzyme mix,” containing 4 units of EcoRI or HindIII, 50 ng of pUC 119, 20 μg/ml bovine serum albumin in the appropriate restriction enzyme buffer (MBI-Fermenta). After incubation for 2 h at 37 °C and addition of 4 μl of sample dye, 20 μl of the samples were electrophoresed in 1.5% agarose gels. Samples subjected to 30 PCR cycles and subsequently incubated without any restriction enzyme were also analyzed. The plasmid vector pUC 119 served as a control of complete restriction enzyme cleavage, which was checked after ethidium bromide staining of the gel. After blotting to Hybond-N+, the nylon membrane was pre-hybridized at 42 °C for 30 min in ExpressHyb hybridization solution (CLONTECH) and incubated in the same solution at 42 °C for 1 h with a 32P-labeled oligonucleotide IE 16 (DAACTATGGAAAATCCCG). This sequence is found in both transcripts from rat and mouse, and hybridizes with the intact rat and mouse 1-kb products as well as the 529-nt EcoRI and the 668-nt HindIII fragment generated from the 1-kb band corresponding to the rat and mouse 8-kb transcript, respectively. The membrane was finally washed with 2× SSC, 0.05% SDS twice for 10 min at 42 °C, before exposure to x-ray film.

RESULTS

Southern blots of genomic DNA isolated from mouse liver was hybridized with 32P-labeled cDNA-probes from the coding and from the 3’-untranslated regions of the 4-kb mouse and the 8-kb rat transcript, respectively (Fig. 1). While the coding regions of the 4-kb mouse and the 8-kb rat transcript are fairly similar (~70% identity at the amino acid level), the untranslated regions are unrelated (9, 10). The hybridization with
mouse DNA of the probe corresponding to the 3'-untranslated region of the 8-kb rat transcript (Fig. 1D) clearly demonstrates that also the mouse genome contains sequence information for this transcript. The cDNA probe corresponding to the coding region of the 8-kb transcript, hybridized to other genomic fragments (Fig. 1B), demonstrating the presence of restriction sites for \(XbaI\), \(HindIII\), \(EcoRI\), and \(BglII\) in this gene. Notably, the hybridization pattern was distinct from that obtained with the probes corresponding to the 4-kb transcript (Fig. 1, A and C), indicating that the two transcripts are encoded by different genes.

The expression of the 8-kb transcript in mouse liver was established using RT-PCR followed by nucleotide sequence analysis of the isolated clones. The nucleotide sequence of the coding region (GenBank™ accession number, AF049894) was highly homologous to that of the corresponding region in the rat 8-kb transcript (96% identity). On the protein level, only a few amino acids differed between the two species (Fig. 2). According to Northern blotting of different mouse tissues, the 8-kb transcript was widely distributed, with lung, liver, and kidney containing the highest amounts (Fig. 3).

**FIG. 2.** Alignment of the \(N\)-deacetylase/\(N\)-sulfotransferase amino acid sequences. Deduced amino acid sequence of mouse liver (ML) \(N\)-deacetylase/\(N\)-sulfotransferase 8-kb transcript and alignment with mouse mastocytoma (MM) 4-kb and rat liver (RL) 8-kb transcript homologues. The numbers indicate the amino acid residues of the proteins. All residues in the sequence corresponding to the 8-kb mouse liver transcript (ML 8 kb), as well as identical residues in homologs, are boxed. Positions where amino acids are identical in two of the sequences are shaded. Gaps introduced to optimize alignment, using the Clustal algorithm with PAM250 residue weight table (Lasergene software, DNASTar), are shown as dashes (—).

**FIG. 3.** Distribution of the 8-kb \(N\)-deacetylase/\(N\)-sulfotransferase mRNA in adult mouse tissues. A mouse multiple tissue Northern blot was hybridized with a 216-bp \(32^P\)-labeled probe recognizing the coding region of the 8 kb transcript (A) and a 2-kb cDNA recognizing \(\beta\)-actin (B). The filter was washed at 50 °C in 0.1× SSC, 0.1% SDS. Sizes of RNA-markers in kilobases are indicated.
to parts of the coding region of the mastocytoma and the rat liver cDNA transcripts, receptively (see "Experimental Procedures"), were mixed and used for screening of a 129/sv mouse genomic DNA library. Several positives clones were isolated and subsequently identified by Southern hybridization with cDNA fragments and oligonucleotides specific for the 4 and the 8-kb transcript, respectively (see "Experimental Procedures"). Two clones, 2:3 and 4:3 in Fig. 4, were derived from the gene encoding the 4-kb transcript, but these clones did not cover the 5'-part of the gene. An ~400-bp Sac I fragment from the most 5'-terminal region of the 4:3 genomic clone was therefore used as a probe in another screening of the library. This approach yielded two more clones (8:2 and 10:1). Together with clones 2:3 and 4:3, approximately 30 kb of genomic sequence were now covered (Fig. 4).

Four of the clones identified in the initial screening were fragments of the gene encoding the 8 kb transcript. Characterization of these clones are in progress and will be reported in a separate communication.

**Structure of the Gene Encoding the 4-kb Transcript—** A crude map of the gene was constructed based on hybridization of restriction fragments with oligonucleotides, previously used as specific primers in nucleotide sequence analysis of cDNA corresponding to the mouse mastocytoma 4-kb transcript (10). Genomic fragments generated by digestion of the clones with ScaI and/or BamHI were subcloned into pUC 119. The exon-intron organization and the size of exons and introns were determined by nucleotide sequence analyses of these fragments. The data available suggest that the mouse 4-kb N-deacetylase/N-sulfotransferase gene consists of 14 exons distributed within 8 kb of genomic DNA. However, it cannot be excluded that additional introns may exist in the extreme 5' and 3'-ends of the gene. The exons contain between 88 and 1322 nucleotides, and all of the donor and acceptor splicing

---

**FIG. 4. Schematic representation of the organization of the mouse gene encoding the 4-kb N-deacetylase/N-sulfotransferase transcript.** The four genomic clones used in determining the sequence of the N-deacetylase/N-sulfotransferase gene are shown (top). The clones covered together approximately 30 kb extending in the 3' direction (---). Below is a schematic view of the gene where exons are represented by boxes while horizontal lines depict introns. The portions of the exons that contain protein coding sequences are filled. The box representing the left-most exon is open to indicate that the 5'-boundary of this exon has not been localized.

**Exon-intron organization of the mouse 4-kb N-deacetylase/N-sulfotransferase gene**

The exons are numbered starting at the 5' end of the gene. Exon sequences are in capital letters; intron sequences are in lowercase letters. The amino acids corresponding to the exon sequence are shown immediately below the nucleotide sequence.

| Exon number | Exon size bp | 5' Splice donor | Intron length bp | 3' Splice acceptor | Amino acid interrupted |
|-------------|--------------|----------------|-----------------|--------------------|-----------------------|
| 1           | >200         | AGCCAAGgtatg   | 428             | tacagATTGTA        | UTR                   |
| 2           | 1322         | T GTT GAAGgtag | 201             | tccagGCT TTG TT    | Glu                   |
| 3           | 88           | CAT ACC Ggtag  | ~269            | cacagGG ACA GAA    | Gly                   |
| 4           | 155          | T GCT TTGgtag  | 189             | cccagGAG CAT GG    | Leu                   |
| 5           | 186          | C ATC ATGgtag  | ~461            | ctcagGTT CTC CC    | Met                   |
| 6           | 129          | T AAT CCGgtag  | 126             | tgcaGATG GT GT     | Pro                   |
| 7           | 183          | G TGG CAGgtag  | 87              | ttcagAAAT CCC TG   | Gln                   |
| 8           | 97           | AAG ACA Ggtaa  | ~1306           | tgcagGC ACC ACC    | Gly                   |
| 9           | 124          | TC GAC TGgtag  | ~121            | ttcagG TAC ATG     | Trp                   |
| 10          | 175          | G TAC CAGgtag  | ~200            | tgcagCAC CAG C     | Gln                   |
| 11          | 171          | T GGA CAGgtacg | ~88             | ttcagTTG CTG A     | Gln                   |
| 12          | 110          | CC CTC AGgtggg | 147             | agtagG TTT GAT     | Arg                   |
| 13          | 103          | C ATG GAAGtag  | 100             | atcagTCT GCC T     | Glu                   |
| 14          | ~614         | T GTG GGC-3'UTR| V G             |                    |                       |

---

*Intron sizes were determined by sequencing; "~" indicates that portions of the introns were sequenced on one strand only.

*The size of exon 1 that contains 5'-untranslated regions was not determined since the 5'-boundary of this exon has not been localized.

*UTR, untranslated region.
mouse "4 kb" ggtcaggtgtagagattgtacctacgtcgttcagcag  
mouse "8 kb" ctcctctctgctgctagctctaccccttctctctctcgaag  

ATCCCCCTGAGCAACCAACAAAGCAGATTCTGCTGACAGCAGAAAACCTCTGAACATCCC 
TTTCTCATCTGCAAGAAGGTAGATGATTCTTCAAGGMAAAGCAACAAGCTCTTCC  

CCCAATGCTGGCACTGGGGCAAGAAGACGATaattgcagctctgatctgcttagtaa 
cgccctctctatgctggcagcgcccggcctgctc  

**Fig. 5. Comparison of the nucleotide sequence of corresponding exons in the mouse genes encoding the 4- and 8-kb transcripts.** Exon sequences are in *capital letters*, while intron sequences are in *lowercase letters*. The *HindIII* site, present in the exon from the gene encoding the 8-kb transcript but absent from exon 8 of the gene encoding the 4-kb transcript, is *underlined*. y = t or c.

sites agree to the consensus sequences for eukaryotic genes (21). These data are summarized in Fig. 4 and in Table I. Exon 1 contains part of the 5′-untranslated region, whereas the remainder of the 5′-untranslated sequence is located in exon 2, which contains the translation start site. This exon, which is much larger than any of the other exons, also encodes both the short N-terminal cytoplasmatic domain and the transmembrane part of the protein in addition to a large portion of the luminal domain of the N-deacetylase/N-sulfotransferase. While most of the intron junctions are of the 0 type (splicing occurs between codons), introns 3 and 8 are type 1 introns (where splicing occurs after the first base of the codon). Introns 9 and 12, finally, are of type 2 (splicing occurs after the second base of the codon).

**Expression of the 4-kb N-Deacetylase/N-Sulfotransferase Transcript**—It has generally been assumed that the 4-kb transcript encodes the N-deacetylase/N-sulfotransferase involved in heparin biosynthesis and thus that this transcript is confined to mast cells. We previously noted a low expression of the 4-kb transcript in rat liver (10). Since mast cells in the liver capsule may have been the source of this transcript, we decided to look for the presence of the 4-kb transcript in purified hepatocytes using a PCR approach. In this method, a primer pair is used that will amplify a 1-kb band from cDNA generated by reverse transcription of either of the 4 or 8-kb transcripts. (The nucleotide sequence of the 4- and 8-kb transcripts is identical in the regions where the primers anneal.) After the PCR, the 1-kb band is incubated with a restriction enzyme that cleaves the band derived from the 8-kb transcript but leaves the band generated from the 4-kb transcript intact. For RNA derived from rat tissues, EcoRI is used, while *HindIII* cleaves the 1-kb band generated from the mouse 8-kb transcript (see Fig. 5 where exon 8 of the gene encoding the 4-kb transcript is compared with the corresponding exon in the mouse gene for the 8-kb transcript.). As shown in Fig. 6, most of the 1-kb PCR product obtained using rat hepatocyte cDNA as template is cleaved by EcoRI, but small amounts of the 4-kb transcript are present in these cells, represented by the weak intact 1-kb band (Fig. 6A). Mouse liver, in contrast, appears to contain similar amounts of the 4- and 8-kb transcripts (Fig. 6B), and mouse mastocytoma contains almost exclusively the 4-kb transcript (Fig. 6C). Small amounts of the *HindIII* cleavage product, not evident in the figure, can be detected after additional PCR cycles (data not shown). In Fig. 6, the large difference in expression of N-deacetylase/N-sulfotransferase mRNA in mastocytoma and liver is also evident.

The distribution of the 4-kb transcript in different adult mouse tissues was also studied by Northern blotting using a probe specific for the 4-kb transcript (Fig. 7A). All tissues examined expressed the transcript, testis, liver, and kidney containing the highest amounts of the mRNA (Fig. 7A). The distribution of an mRNA encoding a protease specific for connective tissue-type mast cells (MMCP-4; Ref. 20) was different, skeletal muscle containing considerable amounts of the transcript while lower levels could be detected in heart, spleen, and lung (Fig. 7B). From these results, it can be concluded that expression of the 4-kb N-deacetylase/N-sulfotransferase transcript is not restricted to mast cells.

**DISCUSSION**

Southern blotting of mouse genomic DNA (Fig. 1) showed that the two different *N-deacetylase/N-sulfotransferase* enzymes, encoded by the 4 and 8-kb transcript, respectively, originate from related but distinct genes. Nucleotide sequence analysis of cDNA corresponding to the 8 kb mRNA (Fig. 2), established that mice also express this transcript. The nucleotide sequence of the rat and human transcripts have previously been published (14, 15). The mouse protein is very similar to its rat counterpart with 99% identity on the protein level (Fig. 2). As demonstrated by Northern blotting, the 8-kb transcript is widely distributed and was detected in all tissues analyzed (Fig. 3).

The gene encoding the human 8-kb transcript has previously been characterized (16). Characterization of the mouse gene encoding the 4-kb transcript, which spans at least 7.4 kb and is split into 14 exons (Fig. 4), revealed a high resemblance between the two genes. The exons have the same size, except for exons 1, 2, and 14. In addition, the exon-intron phases from exon 2 and downstream are of the same type (eight type 0, two type 1 and two type 2 boundaries). The structural similarities between these genes strongly suggest that they have evolved from a common ancestral gene. However, some differences between the two genes may be noted. 1) The introns of the human gene encoding the 8-kb transcript are much larger (16). This gene was estimated to span approximately 35 kb of genomic DNA. 2) In the murine gene encoding the 8-kb transcript, the initiation codon is found in the large exon 2, which

---

2 The nucleotide sequence of the four different PCR products corresponding to the 4- and 8-kb transcripts in rat and mouse have been determined. A unique EcoRI site is found in the 1-kb band generated from the rat 8-kb transcript, whereas the mouse 8-kb mouse transcript contains a unique *HindIII* site.
contains 1005 nucleotides of coding sequence, while the corresponding region in the other gene (1008 nucleotides) is divided between the two first exons.\(^3\) In the human gene, the splice donor site of exon 8 is a GC rather than the expected GT.\(^3\) This rare variant was not found in the murine gene encoding the 4-kb transcript. 4) Exon 14 in the mouse gene has 123 nucleotides of coding region, whereas the corresponding exon in the human gene contains 117 nucleotides. While so far only two different genes encoding glucosaminyl N-deacetylase/N-sulfotransferases have been identified, we cannot exclude the presence of additional genes.

It was previously suggested that the enzyme encoded by the 4-kb transcript, first recognized in mouse mastocytoma (8), was restricted to connective tissue mast cells and to the biosynthesis of heparin (9, 10). However, the tissue distribution of the 4-kb N-deacetylase/N-sulfotransferase mRNA reported in this paper (Figs. 6 and 7), demonstrates that this transcript is widely distributed and that it is present in cells other than mast cells. Accordingly, the enzyme encoded by the 4-kb transcript appears to take part both in heparin and heparan sulfate biosynthesis. The high level of expression of the 4-kb transcript in mastocytoma compared with its expression in e.g. liver (Fig. 6), may be the reason why this transcript was not previously recognized in other cells (9). As indicated by overexpression of the enzymes in vitro (12, 22), the two enzymes may have different catalytic properties; transfection of cDNA corresponding to the 4-kb transcript in the human kidney cell-line 293 resulted in a dramatic increase in N-sulfation of the heparan sulfate produced by the cells (22), while overexpression in COS cells of cDNA for the 8-kb mRNA did not cause any significant change in the overall N-sulfation of the polysaccharide (12).

The enzyme encoded by the 4-kb transcript thus appears to participate in the production of HS with higher N-sulfate content. Tentatively, different locations of the 4- and 8-kb transcripts in a certain tissue may result in a local variation of HS structure. Such local variation of HS structure was recently reported for kidney, using a monoclonal antibody recognizing N-unsubstituted glucosamine residues, found in HS in the glomerular basement membrane but not in the basement membranes of the tubules (23). These results imply that the fine structure of heparan sulfate may be carefully regulated. Expression of the 4- or 8-kb transcript may be one of the means for the cell to control the structure of the heparan sulfate produced, which in turn will be important for the ability of the proteoglycan for functionally important interactions with effector molecules such as cytokines, enzymes, and enzyme inhibitors (4, 6).

Acknowledgment—We thank Britt-Marie Fogelholm for excellent technical assistance.
