Primary structure of a mouse mastocytoma proteoglycan core protein

Lena KJELLÉN,*|| Inger PETTERSSON,* Peter LILLHAGER,† Marie-Louise STEEN,† Ulf PETTERSSON,† Pirjo LEHTONEN,‡ Torbjörn KARLSSON,‡ Erkki RUOSLAHTI§ and Lars HELLMAN†¶

*Department of Veterinary Medical Chemistry, Swedish University of Agricultural Sciences, The Biomedical Center, Box 575, S-751 23 Uppsala, Sweden, †Department of Medical Genetics, University of Uppsala, The Biomedical Center, Box 589, S-751 23 Uppsala, Sweden, ‡Pharmacia, S-751 82 Uppsala 1, Sweden, and §Cancer Research Center, La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.

The complete nucleotide sequence of a mouse mastocytoma proteoglycan core protein mRNA was determined. The mRNA, estimated to contain 1.1 kb, encodes a protein with an M_r of 16715. A 21-amino acid-residue region of the protein is composed of alternating serine and glycine residues. Southern-blot analysis of mouse genomic DNA with cDNA containing sequences corresponding to the Ser-Gly repeat region revealed more than 15 gene fragments. Hybridization with a probe corresponding to the *N*-terminal portion of the core protein identified two fragments, and cDNA covering the *C*-terminal part of the core protein and the 3' untranslated part of the mRNA hybridized to a single fragment. Antibodies against the core protein, obtained after immunization of rabbits with a fusion protein, reacted with both chondroitin sulphate proteoglycans and heparin proteoglycans produced by the tumour. In immunoblotting of a microsomal fraction from the mastocytoma, the antiserum recognized a single protein (M_r 17000), which probably represents the core protein before glycosylation.

INTRODUCTION

Mast cells synthesize proteoglycans that are stored in intracellular granules together with histamine and neutral proteinases (Stevens, 1987). Characterization of mouse and rat mast cells has made it possible to distinguish between two classes of mast cells, referred to as connective-tissue-type and mucosal-type mast cells (for a review on the background for this classification see Stevens, 1986). Whereas the connective-tissue-type mast cells synthesize predominantly heparin proteoglycans (Katz *et al.*, 1986), the mucosal-type mast cells contain mainly chondroitin sulphate proteoglycans (Enerbäck *et al.*, 1985; Stevens *et al.*, 1986). Cells belonging to the latter group include bone-marrow-derived mast cells and the rat basophilic leukaemia (RBL-1) cell line (Serafin *et al.*, 1986; Seldin *et al.*, 1985). The RBL-1 cells have also been shown to synthesize hybrid proteoglycans where heparin and chondroitin sulphate side chains are attached to the same core protein (Seldin *et al.*, 1985).

The function of the mast-cell proteoglycans may be to serve as a matrix to which histamine and the neutral proteinases are attached by ionic linkage (Schwartz & Austen, 1984). However, the mast-cell heparin is also a potent anticoagulant, which inhibits several of the blood-coagulation enzymes by specifically interacting with antithrombin III (Lindahl *et al.*, 1980). It has been proposed that the mast-cell heparin has a role in the regulation of extravascular coagulation under conditions of inflammation (Lindahl & Pejler, 1987). The mouse mastocytoma used in the present work has been extensively utilized in studies on the biosynthesis of heparin (see Lindahl *et al.*, 1986; Lindahl & Kjellén, 1987). The

mastocytoma cells resemble the connective-tissue-type mast cells, and synthesize heparin proteoglycans and to a lesser degree also chondroitin sulphate proteoglycans when cultured *in vitro* (Jacobsson *et al.*, 1985).

The proteoglycans of both subpopulations of mast cells differ from proteoglycans of most other cell types in their resistance to proteolysis. Amino acid analyses of different mast-cell proteoglycans show that the core proteins, whether substituted with heparin or chondroitin sulphate, have a high content of serine and glycine (Robinson *et al.*, 1978; Metcalfe *et al.*, 1980; Stevens *et al.*, 1985; Seldin *et al.*, 1985). Robinson *et al.* (1978), who studied Pronase-treated rat skin heparin proteoglycans, postulated that the polysaccharide-substituted part of the core protein was composed of alternating serine and glycine residues.

A chondroitin sulphate proteoglycan core protein containing such a Ser-Gly repeat was identified by nucleotide sequence analysis of rat yolk-sac-tumour proteoglycan cDNA clones (Bourdon *et al.*, 1985, 1986). The repeat sequence consists of 49 amino acid residues, and the entire core protein contains 179 amino acid residues. Southern-blot analysis has indicated that a single gene in the rat codes for this proteoglycan. The rat mRNA coding for the proteoglycan is expressed only in parietal yolk-sac and in mast cells (Tantravahi *et al.*, 1986; Bourdon *et al.*, 1987b). On the basis of its high content of alternating serine and glycine residues this gene product has been named serglycin (Ruoslahti, 1988). Recently, the amino acid sequence of the RBL-1 serglycin, deduced from cDNA nucleotide sequence analysis, was reported (Avraham *et al.*, 1988). It is identical with the rat parietal-yolk-sac serglycin sequence, indi-

|| To whom correspondence should be addressed.

¶ Present address: Department of Immunology, University of Uppsala, The Biomedical Center, Box 582, S-751 23 Uppsala, Sweden.

cating that no alternative exon splicing has been used to modify the transcript of the different cell types. However, cell-type-specific differences in the length of the 5' untranslated regions of the mRNAs may be of importance for the regulation of transcription in the different cell types (Bourdon *et al.*, 1987b; Avraham *et al.*, 1988). A human analogue to rat serglycin, expressed in promyelocytic-leukaemia HL-60 cells, has also been analysed (Stevens *et al.*, 1988). The predicted amino acid sequence of this protein shows 48% sequence identity with rat serglycin (Avraham *et al.*, 1988).

In the present paper we report the complete nucleotide sequence of a serglycin mRNA from the mouse mastocytoma. Immunological data are presented that demonstrate that mouse serglycin is present as a core protein in mastocytoma chondroitin sulphate proteoglycans and heparin proteoglycans.

MATERIALS AND METHODS

Isolation of polyadenylated mRNA

RNA was extracted by the guanidinium isothiocyanate method (Chirgwin *et al.*, 1979) from a transplantable mastocytoma (Furth *et al.*, 1957) maintained as a solid tumour in (Laden $\times$ A₁)F₁ mice by intramuscular transplantation in the hind legs every 10–12 days. Healthy mice of the same strain were used as a source of kidney and liver RNA. Rat leukaemic basophils (RBL-1 cell line; Eccleston *et al.*, 1973) grown in spinner cultures were pelleted by centrifugation, frozen in liquid N₂ and stored at -70°C before RNA preparation by the same method. Polyadenylated RNA was isolated by chromatography on oligo(dT)-cellulose.

Construction of the bacteriophage λ gt 11 cDNA library

Double-stranded cDNA was synthesized as described by Huynh *et al.* (1985) and ligated to *Eco*RI-digested alkaline-phosphatase-treated bacteriophage λ gt 11 DNA, obtained from Promega Biotech, Madison, WI, U.S.A. (Proclone GT system). Bacteriophage DNA was packaged by use of the same system. The library comprised 0.7×10^5 independent recombinants.

Construction of hybridization probes

Insert DNA of (1) a rat yolk-sac chondroitin sulphate proteoglycan cDNA, pPG1 (Bourdon *et al.*, 1985), (2) the mastocytoma PGM cDNA and restriction fragments from it and (3) a rat IR2 immunocytoma IgE ϵ -chain and κ -chain cDNA (Hellman *et al.*, 1982, 1985) were ^{32}P -labelled by using [α - ^{32}P]dCTP (New England Nuclear) in the RPN.1601Y multiprime DNA labelling system, obtained from Amersham International.

Screening of the bacteriophage library

Nitrocellulose replicas of plaques were pre-hybridized in solution containing $6 \times \text{SSC}$ ($1 \times \text{SSC}$ is $0.15 \text{ M-NaCl}/0.015 \text{ M-sodium citrate buffer pH } 7.0$), $3 \times \text{Denhardt's solution}$ (Denhardt, 1966), 2 mM-EDTA , 0.5% SDS and $0.1 \text{ mg of yeast RNA/ml}$ for 3 h at 65°C before hybridization for 12–15 h in the same buffer containing ^{32}P -labelled probe. The filters were washed in $2 \times \text{SSC}/0.5\%$ SDS. The temperature chosen for hybridization/washing was 42°C and 65°C respectively when pPG1 and PGM insert DNA were used as probes.

Subcloning and DNA sequencing

cDNA inserts, isolated by agarose-gel or polyacrylamide-gel preparative electrophoresis (Maniatis *et al.*, 1982) after *Eco*RI restriction of recombinant bacteriophage DNA, were subcloned into pUC19 plasmids. Recombinant plasmids were isolated as previously described (Bywater *et al.*, 1983). DNA sequencing was performed according to the method of Maxam & Gilbert (1980).

Primer extension analysis

A DNA fragment of PGM (bp 91–134; see Fig. 1), isolated after cleavage of PGM cDNA with *Bst*NI, was 5'-end- ^{32}P -labelled with polynucleotide kinase (Biolabs, Beverly, MA, U.S.A.) and mixed with polyadenylated mastocytoma mRNA in $40 \text{ mM-Pipes buffer, pH } 6.4$, containing 1 mM-EDTA , 0.4 M-NaCl and 80% (v/v) formamide. After being heated at 80°C for 10 min, the hybridization mixture was incubated at 60°C for 1 h, at 55°C for 1 h and at 42°C for 2 h before precipitation with ethanol. The extension reaction was catalysed by Moloney-murine-leukaemia-virus reverse transcriptase (Gibco/BRL, Grand Island, NY, U.S.A.) and carried out at 37°C . After removal of the RNA template by alkaline hydrolysis, the extended material was extracted with phenol, precipitated and dissolved in loading buffer followed by separation on 6% sequencing gels (Maxam & Gilbert, 1980).

Northern-blot hybridization

RNA was denatured in $20 \text{ mM-Mops buffer, pH } 7.0$, containing 50% (v/v) formamide and $2.2 \text{ M-formaldehyde}$ at 70°C for 3 min before being loaded on 1.2% -(v/v)-formaldehyde-containing agarose gels. The fractionated RNA was transferred to nitrocellulose in $20 \times \text{SSC}$. The filters were hybridized at 42°C in solution containing 50% (v/v) formamide, $10 \text{ mM-Hepes buffer, pH } 7.4$, $1 \times \text{Denhardt's solution}$, $1 \text{ mg of yeast RNA/ml}$, $3 \times \text{SSC}$ and 0.1% SDS and washed in $0.1 \times \text{SSC}/0.1\%$ SDS at the indicated temperatures.

Southern-blot hybridization

Southern-blot hybridization (Southern, 1975) was performed as described by Maniatis *et al.* (1982). The nitrocellulose filters were washed at 65°C in $0.1 \times \text{SSC}/0.1\%$ SDS.

Production of fusion protein and immunization

The insert from the cDNA clone PGM was cloned in the bacterial expression vector pATH11 (Spindler *et al.*, 1984), kindly provided by Dr. T. J. Koerner, and transformed into HB101. The resulting fusion protein contains 333 amino acid residues from the *trpE* protein. The fusion protein is under control of the *trp* promoter, which is inducible by 3-(indol-3-yl)acrylic acid. After induction, the bacteria were lysed and the solubilized proteins were separated by SDS/10%-polyacrylamide-gel electrophoresis (Spindler *et al.*, 1984). The fusion protein was excised and homogenized in Freud's complete adjuvant before immunization of rabbits.

Purification and characterization of mastocytoma proteoglycans

Five tumour-bearing mice were injected intraperitoneally with 0.5 mCi of carrier-free [^{35}S]sulphate

(Amersham International) per animal. After 2 h the tumour tissue was dissected out and mixed with tumours from five uninjected mice. The tissue (4–6 g) was homogenized in 50 ml of 50 mM-Tris/HCl buffer, pH 8.0, containing 1% (v/v) Triton X-100 and proteinase inhibitors (1 mM-phenylmethanesulphonyl fluoride, 2 mM-*N*-ethylmaleimide, 2 mM-EDTA and 10 μ g of pepstatin/ml), and incubated for 30 min at 4 °C before centrifugation at 100 000 *g* for 1 h. (The same proteinase inhibitors, at the indicated concentrations, were added to all buffers in the subsequent purification steps.) The supernatant was collected and NaCl was added to a final concentration of 0.35 M together with 30 ml of DEAE-Sephacel, equilibrated in 50 mM-Tris/HCl buffer, pH 8.0, containing 0.1% Triton X-100 and 0.35 M-NaCl. After end-over-end incubation at 4 °C overnight, the mixture was poured into a column, which was extensively washed with the equilibration buffer and subsequently with 50 mM-sodium acetate buffer, pH 4.0, containing 0.1% Triton X-100 and 0.35 M-NaCl. The 35 S-labelled proteoglycans were eluted as a single broad peak in a linear gradient of 0.35–1.5 M-NaCl in 50 mM-sodium acetate buffer, pH 4.0, containing 0.1% Triton X-100. After concentration by ethanol precipitation (Paulsson *et al.*, 1983), the 35 S-labelled proteoglycans were separated from free polysaccharide chains by gel chromatography on a Sephacryl S 400 HR column (Pharmacia) eluted in 50 mM-Tris/HCl buffer, pH 8.0, containing 0.1% Triton X-100 and 3.0 M-NaCl. The purified proteoglycan fraction contained on an average 100 μ g of hexuronic acid, determined as described by Bitter & Muir (1962), corresponding to about 0.3 mg of polysaccharide. The heparin content of the 35 S-labelled proteoglycans was determined after HNO₂ deamination at pH 1.5 (Shively & Conrad, 1976) as previously described (Woods *et al.*, 1984). Antithrombin-Sepharose (Höök *et al.*, 1976) was kindly given by Dr. Ulf Lindahl, Uppsala, Sweden. The conditions for chromatography on antithrombin-Sepharose were as previously described (Thunberg *et al.*, 1982).

Enzymic degradation

Isolated 35 S-labelled proteoglycans (corresponding to approx. 65 μ g of hexuronic acid) and 125 I-labelled proteoglycans were digested with 0.025 unit of chondroitinase ABC (chondroitin ABC lyase, EC 4.2.2.4; obtained from Seikagaku, Tokyo, Japan) at 37 °C for 2 h in 50 mM-Tris/HCl buffer, pH 8.0, containing 50 mM-sodium acetate, 0.1% Triton X-100 and 0.05 mg of bovine serum albumin/ml. Digestion with heparinase (heparin lyase, EC 4.2.2.7; obtained from Seikagaku) was carried out at 37 °C for 2 h in 0.1 M-sodium acetate buffer, pH 7.0, containing 3 mM-CaCl₂ and 0.1% Triton X-100, with 0.006 unit of enzyme. For both enzyme preparations, 1 unit is defined as the amount of enzyme required to generate 1 μ mol of disaccharide/min from the appropriate substrate. Both incubations were performed in the presence of the proteinase inhibitors used in the purification of the proteoglycans, except that EDTA was omitted in the heparinase digestion.

125 I-labelling

Purified proteoglycans, *Staphylococcus aureus* Protein A and Pharmacia low-*M_r* calibration kit were radioiodinated by the chloramine-T method (Hunter, 1973) with the use of Iodobeads (Pierce Chemical Co.).

Electrophoresis, immuno-blotting and dot-blotting

Electrophoresis was performed on polyacrylamide gradient gels in SDS according to the method of Blobel & Dobberstein (1975). Immunoblot analysis of proteins electrophoretically transferred to nitrocellulose filters from SDS/polyacrylamide gels was performed as described by Burnette (1981). For immuno dot-blot analysis, samples were filtered through the membrane by gentle vacuum. Subsequently the nitrocellulose filter was baked for 1 h at 80 °C. After incubation of the filters with 1/50 dilutions of the appropriate serum, the specifically bound antibodies were allowed to react with 125 I-labelled Protein A. The antigens recognized were detected by autoradiography.

RESULTS

cDNA cloning

The cDNA insert from the rat yolk-sac proteoglycan (serglycin) cDNA clone pPG1 (Bourdon *et al.*, 1985) was used to identify cDNA clones in a bacteriophage λ gt 11 library, constructed with the use of polyadenylated mastocytoma RNA as template for first-strand synthesis. Two clones of equal size (0.9 kb) were shown to hybridize with a 1.1 kb-long mastocytoma transcript (results not shown). A hybridizing transcript of the same size was also detected when purified insert from the pPG1 clone was used as a probe. One of the mastocytoma cDNA clones, denoted PGM, was used for further screening of

Fig. 1. Structure of the mastocytoma core protein mRNA and the strategy used to determine its sequence

Two cDNA clones, PGM and 4C, and a primer extension product (see Fig. 2) were used to establish the sequence. The arrows indicate the sequence obtained from each fragment of PGM (a) and 4C (b) cDNA and the sequence information obtained from the extension product (c). The black and hatched areas of the mRNA represents the coding region and the poly(A) tail respectively. The nucleotides are numbered according to their position in the complete nucleotide sequence shown in Fig. 3.

the library, and 18 clones were identified. The size of the inserts in these clones varied from 0.5 kb to 0.9 kb. PGM and one of these clones, denoted 4C, were subcloned in pUC19 for further analysis.

Nucleotide sequence analysis

The complete nucleotide sequence of the insert of clone PGM and partial sequence of the clone 4C were established by the method of Maxam & Gilbert (1980) (see Figs. 1 and 3). The PGM cDNA is 915 nucleotide residues long and contains the coding region, except for the first two nucleotide residues, and the complete 3' non-coding region along with a short poly(A) tail. The 4C cDNA contains one additional base in the 5'-end, but none of the clones extends to the 5' untranslated region.

Primer extension analysis

To obtain the complete mRNA sequence a 44-nucleotide-residue fragment of PGM (bp 91–134) was 5'-end-labelled with polynucleotide kinase and used for primer extension of mastocytoma polyadenylated RNA. The extension products were analysed by electrophoresis on a 6% polyacrylamide gel under denaturing conditions.

Fig. 2. Analysis of polyadenylated RNA from the mouse mastocytoma by primer extension analysis

A 44-nucleotide-residue 5'-end-labelled DNA fragment of PGM (bp 91–134, see Fig. 1) was hybridized to the RNA and extended with reverse transcriptase. The extension products were analysed by electrophoresis on a 6% polyacrylamide gel. Markers consist of *Hind*III and *Hpa*II restriction fragments of pUC19 DNA, end-labelled with the Klenow fragment of DNA polymerase I, and are depicted in number of nucleotide residues.

Only one extension product, containing 135–140 nucleotide residues, was identified (Fig. 2). This band was excised from the gel, eluted and sequenced. The initiation codon could be identified in addition to a 41-nucleotide-residue 5' non-coding region (see Figs. 1 and 3).

Comparison of the structures of the mastocytoma and yolk-sac transcripts and their corresponding proteins

The 938-nucleotide-residue proteoglycan core protein mRNA, poly(A) tail not included, is composed of a 41-residue 5' non-coding region, a 456-residue coding region and a 441-residue 3' non-coding region. A polyadenylation signal (AATAAA) is present 25 residues upstream from the poly(A) site (Fig. 3). The coding region of the mastocytoma transcript corresponds to a protein with an M_r of 16715. A central region of the protein (amino acid residues 90–109) contains serine and glycine residues in repeating alternating sequence. Nine of the serine residues are contained in the tetrapeptide sequence Ser-Gly-Xaa-Gly, which has been shown to represent the recognition consensus amino acid sequence for the attachment of chondroitin sulphate to protein (Bourdon *et al.*, 1987a). Comparison of the mastocytoma and yolk-sac mRNA nucleotide sequences (results not shown), as well as the deduced amino acid sequences of the translated proteins, reveals an extensive sequence similarity (Fig. 3), identifying the mastocytoma protein as a mouse serglycin. Apart from the potential glycosaminoglycan-attachment region, which in rat serglycin contains 24 Ser-Gly repeats compared with ten in the deduced mastocytoma protein, the two polypeptides show 86% sequence identity (Fig. 3). Similarly to the yolk-sac protein (Bourdon *et al.*, 1986), the mastocytoma serglycin contains a potential signal-peptidase-cleavage site (von Heijne, 1984), suggesting the presence of a 25-amino acid-residue signal peptide.

Northern-blot analysis

Northern-blot analysis of mouse polyadenylated RNA from different tissue sources, with the PGM clone as a probe, showed that no or very low concentrations of hybridizing mRNA was present in kidney or liver (Fig. 4). In contrast, polyadenylated RNA isolated from RBL-1 cells contained a hybridizing transcript of size identical with that of the mastocytoma transcript, estimated to 1.1 kb (Fig. 4). An estimated transcript size of 1100 residues gives a calculated value of approx. 160 nucleotide residues for the poly(A) tail.

Southern-blot analysis

Purified fragments originating from different regions of the cDNA clone PGM were used as probes in a Southern-blot analysis of mouse genomic DNA (Fig. 5). Probes corresponding to the C-terminal part of the core protein and the 3' untranslated region of the mRNA (bp 372–958) (see Fig. 3) hybridized to one fragment in each restriction digest analysed. A 5' region probe (bp 44–270), devoid of sequence information for the Ser-Gly repeat region (see Fig. 3), hybridized with two fragments. The size of the fragment hybridizing with the 3' probe was identical with that of the larger fragment hybridizing with the 5' probe in three out of four restriction digests. Probes containing sequence information for the Ser-Gly repeat region (bp 44–371 and bp 270–958 respectively) hybridized to a large number of genomic fragments, as previously shown

1	ATTGCTAATCCAGAGGCTGAGTGGAGCCGAGCTGGTCAGG	ATG	CAG	GTT	CCC	GTC	GGC	AGC	AGG	CTT	GTC	CTG	GCT	CTC	GCC	
MOUSE		Met	---	Gln	Val	Pro	Val	Gly	Ser	Arg	Leu	Val	Leu	Ala	Leu	Ala
RAT		Arg							Thr							
84	TTC	GTC	CTG	GTT	TGG	GGA	TCT	TCA	GTG	CAA	GGT	TAT	CCT	GCT	CGG	AGA
	Phe	Val	Leu	Val	Trp	Gly	Ser	Ser	Val	Gln	Gly	Tyr	Pro	Ala	Arg	Arg
162	AAT	GGC	TTT	TTT	GCG	AAC	TGC	ATC	GAG	GAG	AAG	GGA	CCA	CAG	TTT	GAC
	Asn	Gly	Phe	Phe	Ala	Asn	Cys	Ile	Glu	Glu	Lys	Gly	Pro	Gln	Phe	Asp
	Asp		Ile											Arg		
237	CCT	CCC	ATG	AAT	AAT	CCT	GTT	TTG	ATG	GAA	GGA	CCC	TCA	AAA	GAT	TTC
	Pro	Pro	Met	Asn	Asn	Pro	Val	Leu	Met	Glu	Gly	Pro	Ser	Lys	Asp	Phe
				Thr	Asp					Arg		Phe	Pro	Asn		
312	GGT	TCG	GGC	TCC	GGC	TCT	GGC	TCC	GGC	TCT	GGC	TCG	GGT	TCC	GGC	TCC
	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser
369	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser	Gly	Ser
390	GAA	TAC	CAG	CCA	ACA	GAT	GAA	AGC	AAT	ATT	GTC	TAT	TTC	AAC	TAT	AAG
	Glu	Tyr	Gln	Pro	Thr	Asp	Glu	Ser	Asn	Ile	Val	Tyr	Phe	Asn	Tyr	Lys
									Asn							Gly
468	CAA	GAC	CAA	CCA	GAA	GAC	GAT	TTT	ATT	ATA	TGA	ATGTGACGGTCTCTGTCTCC	CCACCTCCATGTGGAACAATGTATT	CAGTATACTTAGTG		
	Gln	Asp	Gln	Pro	Glu	Asp	Asp	Phe	Ile	Ile	End					
560	TACCACGTTTAAATGACCAGTCTCAGGATAAAGAGTTTTACAGAAAATTTAAATGCCTGGAAAAGACTCTTGAATCCTGTTACCCCTTTCCTCATTAACTCG															
663	TAAGGAATTATGCTTTAATGCTGTTACCTATCTTGTGTTCTGGAAAATGCCTGCATTTATGTGTATTGAATCAACATTTAAGAAATTAACACACACCCCAT															
766	TATTATACAATAACTTTCAAAGCCATACTGGTTTTGAAAATTTAATTTGATAGCAAGTTGATGAACAATCTTTCATACCTAAAGTGTTTCAGGAACCCCACTC															
869	GCATTGTGAATTACAAATATATTCCTTTATGTGATTAAAAAGAAATAGTGAACACAGTGGCCTCCTAAAAA															

Fig. 3. Nucleotide sequence of the mouse serglycin mRNA and comparison of the predicted amino acid sequences of mouse and rat serglycins

Only amino acid residues in the rat sequence differing from those in mouse serglycin are shown. The broken lines indicate amino acid deletions in the sequences. The potential signal-peptidase-cleavage site is indicated by an arrow and the polyadenylation signal is underlined. The numbers on the right and the left indicate the amino acid residue and the nucleotide residue in the respective sequence.

for the rat genome (Tantravahi *et al.*, 1986; Bourdon *et al.*, 1987b).

Demonstration of mouse serglycin as a mastocytoma proteoglycan core protein

The insert of clone PGM was subcloned into the prokaryotic expression vector pATH11 and the resulting trpE fusion protein was purified by SDS/polyacrylamide-gel electrophoresis. The fusion protein was injected into rabbits and a polyclonal antiserum was produced. In immunoblotting of a microsomal fraction from the

Fig. 4. Northern-blot analysis of polyadenylated RNA from different tissue sources

Purified insert from the clone PGM was used as a probe. Washing after hybridization was carried out at 50 °C. Lane A, mouse mastocytoma RNA; lane B, mouse kidney RNA; lane C, mouse liver RNA; lane D, RBL-1 RNA. Insert DNA of a rat IgE ϵ -chain and κ -chain cDNA were hybridized to polyadenylated RNA from a rat IR162 immunocytoma, and the hybridizing transcripts of known sizes were used as size markers, depicted in number of nucleotide residues.

Fig. 5. Southern-blot analysis of restriction fragments of PGM cDNA

High- M_r genomic mouse liver DNA (10 μ g per lane) was cleaved with *Bam*HI (lanes 1), *Eco*RI (lanes 2), *Hind*III (lanes 3) and *Xba*I (lanes 4) and transferred to nitrocellulose after separation by electrophoresis on a 0.8 % agarose gel. The filter was probed with 32 P-labelled restriction fragments of PGM cDNA: (a) *Eco*RI-*Sty*I fragment (bp 44-371); (b) *Sty*I-*Eco*RI fragment (bp 372-958); (c) *Eco*RI-*Pss*I fragment (bp 44-270); (d) *Pss*I-*Eco*RI fragment (bp 271-958). *Hind*III restriction fragments of bacteriophage λ DNA, end-labelled with the Klenow fragment of DNA polymerase I, were used as markers (lanes 5), the sizes of which are depicted in kb.

Fig. 6. Immunoblotting of a mastocytoma microsomal fraction

A crude microsomal fraction from the mouse mastocytoma (250 μ g) (lanes B, C and E) and M_r marker proteins (lanes A and D) were reduced and subjected to SDS/10-15 % polyacrylamide-gel electrophoresis. After electrophoresis part of the gel was fixed and stained with Coomassie Brilliant Blue (lanes A and B). The other part of the gel (lanes C, D and E) was transferred to nitrocellulose, which was incubated with 1/50 dilutions of non-immune serum (lane C) and anti-(fusion protein) serum (lane E) followed by incubation with 125 I-labelled Protein A and autoradiography. Lane D, 125 I-labelled M_r marker proteins.

mastocytoma, this antiserum recognized a single component, with M_r 17000 (Fig. 6), which probably represents the core protein before glycosylation. When analysed in a dot-blot assay, purified 35 S-labelled proteoglycans, containing a mixture of heparin proteoglycans and chondroitin sulphate proteoglycans, were also found to react with the anti-(fusion protein) antibodies (results not shown). However, attempts to precipitate 35 S-labelled proteoglycans with the anti-(fusion protein) antibodies were unsuccessful (results not shown), suggesting that immobilization of the proteoglycans on nitrocellulose was essential for immunoreactivity. To determine whether serglycin was present in both heparin proteoglycans and chondroitin sulphate proteoglycans by using the dot-blot assay, it was necessary to obtain a complete separation of the two types of proteoglycans. Since intact proteoglycans could not be separated by either DEAE-Sephacel ion-exchange chromatography or gel chromatography (results not shown), chondroitinase-ABC-generated chondroitin sulphate core proteins and intact heparin proteoglycans respectively were instead purified by a combination of DEAE-Sephacel ion-exchange chromatography and affinity chromatography on antithrombin-Sepharose (Fig. 7). Proteoglycans isolated from the mastocytoma after labelling with [35 S]sulphate *in vivo* and proteoglycans from the same preparation labelled with 125 I were chromato-

graphed in parallel to obtain material for immuno dot-blot analysis and to monitor the position of the protein part of the proteoglycans respectively. [The 35 S radioactivity does not interfere with the detection of 125 I-labelled Protein A after immuno dot-blotting, as no positive signal was obtained when a non-immune serum was used instead of the anti-(fusion protein) serum (results not shown).] HNO_2 treatment of the 35 S-labelled proteoglycans followed by gel chromatography on Sephadex G-50 showed that 45 % of the radioactivity was present in heparin (results not shown). After DEAE-Sephacel chromatography of 35 S-labelled proteoglycans treated with chondroitinase ABC, approx. 50 % of the 35 S radioactivity was retained on the column, whereas the rest of the 35 S radioactivity, presumably chondroitin sulphate degradation products, was present in the non-DEAE-Sephacel-binding fractions (Fig. 7b). When the chondroitinase-ABC-treated 125 I-labelled proteoglycans were similarly analysed, the non-DEAE-Sephacel-binding fractions were shown to contain a 125 I-labelled component (Fig. 7b), whereas the non-digested 125 I-labelled material was quantitatively retained by the column (Fig. 7a). The non-DEAE-Sephacel-binding

Fig. 7. Purification and immuno dot-blot analyses of chondroitin sulphate proteoglycan core proteins and heparin proteoglycans

Mastocytoma ^{35}S -labelled proteoglycans (15000 c.p.m.) corresponding to approx. 65 μg of uronic acid ($\square$) and ^{125}I -labelled proteoglycans (30000 c.p.m.) ($\blacklozenge$) in a total volume of 600 μl respectively were applied to 0.5 ml columns of DEAE-Sephacel before (a) and after (b) chondroitinase ABC digestion. The arrows in (a) and (b) indicate change of buffer: first arrow, 50 mM-Tris/HCl buffer, pH 8.0, containing 0.35 M-NaCl and 0.1 % Triton X-100; second arrow, 50 mM-sodium acetate buffer, pH 4.0, containing 0.35 M-NaCl and 0.1 % Triton X-100; third arrow, 50 mM-sodium acetate buffer, pH 4.0, containing 1.5 M-NaCl and 0.1 % Triton X-100. Fractions 18–22 in (b) were pooled and dialysed against 50 mM-Tris/HCl buffer, pH 7.4, containing 0.05 M-NaCl and 0.1 % Triton X-100 before application on 0.6 ml antithrombin-Sephacel columns in a total volume of 1 ml (c). The arrows in (c) indicate change of buffer from 50 mM-Tris/HCl buffer, pH 7.4, containing 0.05 M-NaCl (first arrow) to 0.05 M-Tris/HCl buffer, pH 7.4, containing 3 M-NaCl (second arrow). One-tenth of the fractions obtained after chromatography of the ^{35}S -labelled samples were assayed for radioactivity, and ^{125}I radioactivity was measured on the entire fractions. The first fractions contain 600 μl (a and b) and 1000 μl (c) and all the following ones 200 μl . Fractions 3–7 (a), 3–7 (b), 1–5 (c) and 13–17 (c) of the ^{35}S -labelled samples were pooled (total volume 1 ml), and 250 μl , 125 μl and 62.5 μl samples were applied to nitrocellulose filters. The filters were incubated in 1/50 dilutions of anti-(fusion protein) serum and subsequently with ^{125}I -labelled Protein A before autoradiography.

fractions thus should contain deglycosylated chondroitin sulphate proteoglycan core proteins. This fraction (derived from the ^{35}S -labelled proteoglycans) gave a positive signal when tested for reactivity with the anti-(fusion protein) antibodies (Fig. 7b), demonstrating that serglycin functions as a mastocytoma chondroitin sulphate proteoglycan core protein. As expected, no immuno-reactive material was present in the corresponding fractions of the untreated ^{35}S -labelled proteoglycans (Fig. 7a). The DEAE-Sephacel-binding ^{35}S -labelled and ^{125}I -labelled macromolecules obtained after chromatography of the chondroitinase-ABC-treated proteoglycans (fractions 18–22 in Fig. 7b) were applied to Sepharose conjugated with antithrombin. A non-binding and an antithrombin-binding fraction could be recovered (Fig. 7c). Immuno dot-blot analysis of the ^{35}S -labelled fractions (Fig. 7c) demonstrated that both fractions reacted with the anti-(fusion protein) antibodies. The non-antithrombin-binding macromolecules were not further characterized (however, see the Discussion section), but the identity of the antithrombin-binding ^{125}I -labelled macromolecules as heparin-containing proteoglycans was established by rechromatography of this fraction on antithrombin-Sephacel after heparinase digestion; the enzyme treatment resulted in complete loss of affinity for the gel (results not shown). Furthermore, HNO_2 degradation of the antithrombin-binding ^{35}S -labelled macromolecules followed by gel chromatography on Sephadex G-50 showed that the ^{35}S radioactivity was present exclusively in heparin. Hence serglycin apparently occurs also in heparin-containing proteoglycans.

DISCUSSION

Nucleotide sequence data have been used to deduce the complete amino acid sequence of a mouse serglycin. The nucleotide sequence of the mouse transcript and its deduced amino acid sequence are closely related to the cDNA sequence and amino acid sequence of rat serglycin present as a core protein in rat yolk-sac chondroitin sulphate proteoglycans (Bourdon *et al.*, 1985, 1986) and synthesized by the mucosal-type mast-cell line RBL-1 (Avraham *et al.*, 1988).

Southern-blot analysis of rat genomic DNA with probes corresponding to either the 3' or the 5' part of serglycin mRNA has indicated that a single gene codes for this protein (Tantravahi *et al.*, 1986; Bourdon *et al.*, 1987b). Similar results were obtained for the mouse genome when a probe corresponding to the 3' part of mouse serglycin mRNA was used in genomic blot hybridization (Fig. 5). However, a 5' probe covering the N-terminal part of the cloned mouse protein identified two hybridizing gene fragments in all four restriction digests (Fig. 5). Although this result could indicate the presence of two serglycin genes in the mouse genome, it is also possible that the serglycin gene contains a large intron with cleavage sites for the restriction enzymes used to generate the genomic mouse fragments. The rat 5' probe, which only identified single gene fragments in the rat genome (Bourdon *et al.*, 1987b), did not contain sequence information for the signal peptide, which was present in the mouse 5' probe (see Figs. 1 and 3). If the structures of the rat and mouse genes are similar, this would indicate that the postulated large intron is located between the exon containing sequence information for

the signal peptide and exons coding for the rest of the protein. The large number of genomic fragments hybridizing with probes containing sequence information for the Ser-Gly repeat region (Fig. 5) indicates that the mouse genome, similarly to the rat and human genomes (Tantravahi *et al.*, 1986; Bourdon *et al.*, 1987b; Stevens *et al.*, 1988), contains multiple genes with sequences similar to this region.

Primer extension analysis demonstrated that the mouse mastocytoma serglycin mRNA, similarly to the rat serglycin mRNA produced by RBL-1 cells (Bourdon *et al.*, 1987b), contains a short 5' untranslated region. In contrast, mRNA isolated from a rat yolk-sac carcinoma cell line contained an additional 180 nucleotide residues upstream from the 5' cap site of RBL-1 mRNA (Bourdon *et al.*, 1987b). As discussed by Bourdon *et al.* (1987b), the expression of two differently sized serglycin mRNAs could result from the differential use of two promoters, and may suggest that the serglycin gene is controlled by different tissue-specific regulatory sequences.

On the basis of the fact that only gene coding for serglycin was present in the rat genome and that serglycin mRNA was expressed in both mucosal-type and connective-tissue-type mast cells, Tantravahi *et al.* (1986) postulated that serglycin is used as a core protein for both chondroitin sulphate and heparin synthesis. Our results support this notion. On DEAE-Sephacel ion-exchange chromatography of a chondroitinase ABC digest of mastocytoma proteoglycans, a non-DEAE-Sephacel-binding fraction that reacted with the anti-(fusion protein) antiserum could be isolated (Fig. 7b). Since no such material was found on DEAE-Sephacel ion-exchange chromatography of untreated proteoglycans (Fig. 7a), the chondroitinase-ABC-generated non-DEAE-Sephacel-binding fraction must represent chondroitin sulphate proteoglycan core proteins. However, most of the immunoreactive chondroitinase-ABC-treated macromolecules were retained on the DEAE-Sephacel column, suggesting that the majority of the chondroitin sulphate proteoglycan core proteins still had the ability to bind to the gel, conceivably via short oligosaccharide stubs generated by the enzyme treatment. The DEAE-Sephacel-binding macromolecules were therefore chromatographed on antithrombin-Sepharose to obtain heparin proteoglycans free from contaminating chondroitin sulphate proteoglycan core proteins (Fig. 7c). Chondroitin sulphate does not interact with antithrombin (Teien *et al.*, 1976), and affinity chromatography on antithrombin-Sepharose can be used to fractionate heparin into a non-binding fraction and two fractions with low affinity and high affinity for antithrombin (Höök *et al.*, 1976). (The elution conditions used in Fig. 7 will only allow for the separation of heparin with and without affinity for antithrombin respectively.) Immuno dot-blot analysis showed that roughly half of the immunoreactive macromolecules had affinity for antithrombin, thus strongly suggesting that serglycin also occurs in heparin proteoglycans (Fig. 7c). The non-antithrombin-binding fraction conceivably contains chondroitin sulphate proteoglycan core proteins in addition to heparin proteoglycans lacking affinity for antithrombin. However, it is also possible that a portion of the proteoglycans are hybrids carrying both heparin and chondroitin sulphate side chains on the same core protein. If so, the non-DEAE-Sephacel-binding fraction obtained after chondroitinase ABC digestion would

represent core proteins from a population of proteoglycans carrying exclusively chondroitin sulphate chains, whereas the DEAE-Sephacel-binding fraction would contain hybrid proteoglycans and possibly a population of proteoglycans with heparin chains only. In any case, our results demonstrate that serglycin is present in chondroitin sulphate proteoglycans and in heparin-containing proteoglycans. Whether or not the latter are hybrid proteoglycans remains to be determined. However, the fact that no heparin proteoglycan core proteins could be isolated after DEAE-Sephacel ion-exchange chromatography of proteoglycans treated with heparinase (results not shown) may indicate that very few 'pure' heparin proteoglycans are synthesized by the mastocytoma.

As mast cells appear to utilize serglycin for synthesis of both heparin proteoglycans and chondroitin sulphate proteoglycans, regulatory signals other than the primary structure of the newly translated core protein must direct the synthesis of glycosaminoglycans resulting in the production of either type of proteoglycan. Differences in proteolytic cleavage of the newly synthesized mast-cell core protein could conceivably affect subsequent glycosylation. Tentatively, a proteolytic cleavage of the core protein may be a necessary signal for the biosynthesis of one of the polysaccharides but not for the other. Such a signal would allow for the production of hybrid proteoglycans, which then should contain a proteolytically processed core protein. Alternatively, the type of proteoglycan synthesized may depend on the expression of the polysaccharide-synthesizing enzymes. The polysaccharide-protein linkage region, a galactosyl-galactosyl-xylosyl trisaccharide sequence, is shared by both kinds of proteoglycans. Polymerization of the polysaccharide chains is achieved by the stepwise transfer of D-glucuronic acid and N-acetyl-D-hexosamine. In heparin biosynthesis the hexosamine transferred is N-acetyl-D-glucosamine, whereas chondroitin sulphate contains N-acetyl-D-galactosamine (for reviews on glycosaminoglycan biosynthesis see Rodén, 1980; Lindahl *et al.*, 1986; Lindahl & Kjellén, 1987). The committing step will be the incorporation of the first hexosamine unit following the linkage region. It is possible that the enzymes responsible for the transfer of these particular hexosamine units are under regulatory control and that the abundance of these enzymes will determine the relative amounts of heparin proteoglycans and chondroitin sulphate proteoglycans synthesized.

We thank Å. Oldberg (University of Lund, Sweden) for helpful advice. This work was supported by grants from the Swedish Medical Research Council [(03X-)6525, (03P-)6958 and (13X-)2309], the Swedish National Board for Technical Development and the Juvenile Diabetes Foundation.

REFERENCES

- Avraham, S., Stevens, R. L., Gartner, M. C., Austen, K. F., Lalley, P. A. & Weis, J. H. (1988) *J. Biol. Chem.* **263**, 7292-7296
- Bitter, T. & Muir, H. M. (1962) *Anal. Biochem.* **4**, 330-334
- Blöbel, G. & Dobberstein, B. (1975) *J. Cell Biol.* **67**, 835-851

- Bourdon, M. A., Oldberg, Å., Pierschbacher, M. & Ruoslahti, E. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1321–1325
- Bourdon, M. A., Shiga, M. & Ruoslahti, E. (1986) *J. Biol. Chem.* **261**, 12534–12537
- Bourdon, M. A., Krusius, T., Campbell, S., Schwartz, N. B. & Ruoslahti, E. (1987a) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3194–3198
- Bourdon, M. A., Shiga, M. & Ruoslahti, E. (1987b) *Mol. Cell. Biol.* **7**, 33–40
- Burnette, W. N. (1981) *Anal. Biochem.* **112**, 195–203
- Bywater, M., Bywater, R. & Hellman, L. (1983) *Anal. Biochem.* **132**, 219–224
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
- Denhardt, D. T. (1966) *Biochem. Biophys. Res. Commun.* **23**, 641–652
- Eccleston, E., Leonard, B. J., Lowe, J. S. & Welford, H. J. (1973) *Nature (London)* **244**, 73–76
- Enerbäck, L., Kolset, S. O., Kusche, M., Hjerpe, A. & Lindahl, U. (1985) *Biochem. J.* **227**, 661–668
- Furth, J., Hagen, P. & Hirsch, E. I. (1957) *Proc. Soc. Exp. Biol. Med.* **95**, 824–828
- Hellman, L., Pettersson, U. & Bennich, H. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1264–1268
- Hellman, L., Steen, M.-L. & Pettersson, U. (1985) *Gene* **40**, 115–124
- Höök, M., Björk, I., Hopwood, J. & Lindahl, U. (1976) *FEBS Lett.* **66**, 90–93
- Hunter, W. M. (1973) in *Handbook of Experimental Immunology*, (Weis, D. E., ed.), pp. 17.1–17.36, Blackwell Scientific Publications, London
- Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach* (Glover, D., ed.), pp. 49–78, IRL Press, Oxford
- Jacobsson, K.-G., Riesenfeld, J. R. & Lindahl, U. (1985) *J. Biol. Chem.* **260**, 12154–12159
- Katz, H. R., Austen, K. F., Caterson, B. & Stevens, R. (1986) *J. Biol. Chem.* **261**, 13393–13396
- Lindahl, U. & Kjellén, L. (1987) in *Biology of Extracellular Matrix: Biology of Proteoglycans* (Wight, T. N. & Mecham, R. P., eds.), pp. 59–104, Academic Press, New York
- Lindahl, U. & Pejler, G. (1987) *Acta Med. Scand. Suppl.* **715**, 139–144
- Lindahl, U., Bäckström, G., Thunberg, L. & Leder, I. G. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6551–6555
- Lindahl, U., Feingold, D. S. & Rodén, L. (1986) *Trends Biochem. Sci.* **11**, 221–225
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560
- Metcalf, D. D., Smith, J. A., Austen, K. F. & Silbert, J. E. (1980) *J. Biol. Chem.* **255**, 11753–11758
- Paulsson, M., Sommarin, Y. & Heinegård, D. (1983) *Biochem. J.* **212**, 659–667
- Robinson, H. C., Horner, A. A., Höök, M., Ögren, S. & Lindahl, U. (1978) *J. Biol. Chem.* **253**, 6687–6693
- Rodén, L. (1980) in *The Biochemistry of Glycoproteins and Proteoglycans* (Lennarz, W. J., ed.), pp. 267–371, Plenum Press, New York
- Ruoslahti, E. (1988) *Annu. Rev. Cell Biol.* **4**, 229–255
- Schwartz, L. B. & Austen, K. F. (1984) *Prog. Allergy* **34**, 271–321
- Seldin, D. C., Austen, K. F. & Stevens, R. L. (1985) *J. Biol. Chem.* **260**, 11131–11139
- Serafin, W. E., Katz, H. R., Austen, K. F. & Stevens, R. L. (1986) *J. Biol. Chem.* **261**, 15017–15021
- Shively, J. E. & Conrad, H. E. (1976) *Biochemistry* **15**, 3932–3942
- Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517
- Spindler, K. R., Rosser, D. S. E. & Berk, A. J. (1984) *J. Virol.* **49**, 132–141
- Stevens, R. L. (1986) *Ciba Found. Symp.* **124**, 272–280
- Stevens, R. L. (1987) in *Biology of Extracellular Matrix: Biology of Proteoglycans* (Wight, T. N. & Mecham, R. P., eds.), pp. 367–387, Academic Press, New York
- Stevens, R. L., Otsu, K. & Austen, K. F. (1985) *J. Biol. Chem.* **260**, 14194–14200
- Stevens, R. L., Lee, T. D. G., Seldin, D. C., Austen, K. F., Befus, A. D. & Bienenstock, J. (1986) *J. Immunol.* **137**, 291–295
- Stevens, R. L., Avraham, S., Gartner, M. C., Bruns, G. A. P., Austen, K. F. & Weis, J. H. (1988) *J. Biol. Chem.* **263**, 7287–7291
- Tantravahi, R. V., Stevens, R. L., Austen, K. F. & Weis, J. H. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 9207–9210
- Teien, A. N., Abildgaard, U. & Höök, M. (1976) *Thromb. Res.* **8**, 859–864
- Thunberg, L., Bäckström, G. & Lindahl, U. (1982) *Carbohydr. Res.* **100**, 393–410
- von Heijne, G. (1984) *J. Mol. Biol.* **173**, 243–251
- Woods, A., Höök, M., Kjellén, L., Smith, C. G. & Rees, D. A. (1984) *J. Cell Biol.* **99**, 1743–1753

Received 19 September 1988/21 April 1989; accepted 9 May 1989