



Published in final edited form as:

*J Biol Chem.* 2002 December 20; 277(51): 49175–49185. doi:10.1074/jbc.M205131200.

## Cloning and Characterization of Two Extracellular Heparin-degrading Endosulfatases in Mice and Humans\*

Megumi Morimoto-Tomita<sup>‡,§</sup>, Kenji Uchimura<sup>‡,§</sup>, Zena Werb<sup>‡</sup>, Stefan Hemmerich<sup>¶</sup>, and Steven D. Rosen<sup>‡,||</sup>

<sup>‡</sup> Department of Anatomy, University of California, San Francisco, California 94143-0452

<sup>¶</sup> Thios Pharmaceuticals Inc., Oakland, California 94609

### Abstract

Here we report the cloning of a full-length cDNA encoding the human ortholog (HSulf-1) of the developmentally regulated putative sulfatases QSulf-1 (Dhoot, G. K., Gustafsson, M. K., Ai, X., Sun, W., Standiford, D. M., and Emerson, C. P., Jr. (2001) *Science* 293, 1663–1666) and RSulfFP1 (Ohto, T., Uchida, H., Yamazaki, H., Keino-Masu, K., Matsui, A., and Masu, M. (2002) *Genes Cells* 7, 173–185) as well as a cDNA encoding a closely related protein, designated HSulf-2. We have also obtained cDNAs for the mouse orthologs of both Sulfos. We demonstrate that the proteins encoded by both classes of cDNAs are endoproteolytically processed in the secretory pathway and are released into conditioned medium of transfected CHO cells. We demonstrate that the mammalian Sulfos exhibit arylsulfatase activity with a pH optimum in the neutral range; moreover, they can remove sulfate from the C-6 position of glucosamine within specific subregions of intact heparin. Taken together, our results establish that the mammalian Sulfos are extracellular endosulfatases with strong potential for modulating the interactions of heparan sulfate proteoglycans in the extracellular microenvironment.

Sulfatases are enzymes that hydrolyze sulfate esters (reviewed in Ref. 1). The majority of the known sulfatases are lysosomal enzymes. Enzymes in this class, working in concert with glycosidases, function in the sequential degradation of sulfated macromolecules. Proteoglycans are the major class of sulfated macromolecules; hence, many of the lysosomal enzymes are specific for the sulfate modifications on glycosaminoglycan chains. Steryl sulfatase, another intracellular sulfatase that hydrolyzes 3- $\beta$ -hydroxysteroid sulfates, is membrane-bound and localized to microsomes (1). Three other sulfatases have a nonlysosomal localization (2,3). Recently, Dhoot *et al.* (4) identified a gene, called QSulf-1, which is expressed in several regions of the quail embryo. QSulf-1 was inferred to be a sulfatase, because it possesses a domain closely related to the human lysosomal glucosamine-6-sulfatase (HG6S) (5), an enzyme involved in the degradation of heparan sulfate chains. Antisense experiments strongly implicate QSulf-1 in muscle specification within epiaxial somite progenitors. QSulf-1 is apparently targeted to the secretory pathway in that it is expressed on the cell surface of transfected cells. A close relative of QSulf-1 was subsequently discovered in the rat embryo (6). It was named RSulfFP1 due to its prominent expression of transcripts in the floor plate of

\*This work was supported by California Breast Cancer Program Grant 71B-0036; National Institutes of Health (NIH) Grant R37GM23547 (to S. D. R.); NCI, NIH, Grant CA72006 (to Z. W.); and postdoctoral fellowships from the Japanese Society for the Promotion of Science (to M. M.-T. and K. U.).

||To whom correspondence should be addressed. Tel.: 415-476-1579; Fax: 415-476-4845; sdr@itsa.ucsf.edu.

<sup>§</sup>These authors contributed equally to this work.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY101175 (HSulf-1), AY101176 (HSulf-2), AY101178 (MSulf-1), and AY101177 (MSulf-2).

the developing nervous system. In transfected cells, the protein is localized to the Golgi apparatus and the cell surface. Neither QSulf-1 nor RSulfFP1 was shown explicitly to have sulfatase activity.

Here we report the cloning of a full-length cDNA encoding the human ortholog (HSulf-1) of QSulf-1 and RSulfFP1 as well as a cDNA encoding a closely related protein, designated HSulf-2, and cDNAs for the mouse orthologs of both Sulfs. We demonstrate that the mammalian Sulfs are endoproteolytically processed and secreted into the extracellular space of transfected cells, where they exhibit both arylsulfatase activity and highly specific endoglucosamine-6-sulfatase activity against intact heparin.

## EXPERIMENTAL PROCEDURES

### Cloning of the Sulfs

For the cloning of HSulf-1, we BLAST screened the GenBank™ databases for genomic sequences related to the KIAA1077 cDNA, a partial human cDNA that is closely related to QSulf-1 (4). We assembled a 257-kb genomic contig from genomic clones (GenBank™ accession numbers AC013746.10 and AC091047.5) that contained the entire Sulf-1 gene. Using the GeneScan exon prediction program (available on the World Wide Web at genes.mit.edu/GEN-SCAN.html) in conjunction with EST data displayed in the University of California Santa Cruz human genome browser (available on the World Wide Web at genome.ucsc.edu), we identified a 232-bp sequence within this contig (putative first exon) encoding a signal sequence related to that of the QSulf-1 protein. We derived a forward primer (5'-AAGTATTCTTGCTGTGCTCTGGTTTTGGC-3') from this putative first coding exon and a reverse primer from the 5' end of the KIAA1077 cDNA (5'-AGAAGCGAGAATTCTTGATTAATCCAAGCC-3'). A PCR-based pool selection technique was applied to a human prostate "Cap-Finder" cDNA library (BD-Clontech, Palo Alto, CA), leading to the identification of a 4669-bp cDNA. The cDNA was fully sequenced on both strands, confirming the sequence derived by genomic analysis. Using the HSulf-1 cDNA sequence in a BLAST search, we identified a highly related partial cDNA sequence (KIAA1247; accession no. AB033073.1) from humans in the GenBank™ data base. We assembled a 311-kb contig from genomic sequences (AL121777.39, AL354813.31, and AL034418.5) and identified a putative first coding exon as described above. A full-length cDNA was predicted by exon scanning. We then used a forward primer 5'-TGCAGGGAACCTTCAAAGGACT-3' derived from the first coding exon and a reverse primer 5'-GAAGAAAGCTGTCCGGTAGCCA-3' derived from the 5' end of the KIAA1247 cDNA to clone a full-length (4279 bp) human Sulf-2 cDNA from a human lung cDNA library (Origene, Rockville, MD) as described above. A full-length cDNA was obtained and sequenced on both strands. The coding sequence exactly matched the sequence predicted from the exons identified in the genomic contig. In order to clone the mouse Sulf-1 ortholog, we screened the GenBank™ mouse EST data base with the predicted HSulf-1 protein sequence and identified EST BB558850 encoding the N terminus of a potential mouse Sulf-1. We then used forward

<sup>1</sup>The abbreviations used are: HG6S, human glucosamine-6-sulfatase; 4-MUS, 4-methylumbelliferyl sulfate; CHO, Chinese hamster ovary; CM, conditioned medium; CMK, decanoyl-Arg-Val-Lys-Arg-chloromethylketone; EST, expressed sequence tag; HPLC, high performance liquid chromatography; HSPG, heparan sulfate proteoglycan; IdoA, L-iduronic acid; Ni-NTA, nickel-nitrilotriacetic acid; ΔdiHS-NS, 2-deoxy-2-sulfamido-4-O-(4-deoxy-L-threo-hexenepyransyluronic acid)-D-glucose; ΔdiHS-6S, 2-acetamide-2-deoxy-4-O-(4-deoxy-L-threo-hexenepyransyluronic acid)-6-O-sulfo-D-glucose; ΔdiHS-(N,6)diS, 2-deoxy-2-sulfamido-4-O-(4-deoxy-2-O-sulfo-L-threo-hexenepyransyluronic acid)-6-O-sulfo-D-glucose; ΔdiHS-(N,2)diS, 2-deoxy-2-sulfamido-4-O-(4-deoxy-2-O-sulfo-L-threo-hexenepyransyluronic acid)-6-O-sulfo-D-glucose; ΔdiHS-(N,6,2)triS, 2-deoxy-2-sulfamido-4-O-(4-deoxy-2-O-sulfo-L-threo-hexenepyransyluronic acid)-6-O-sulfo-D-glucose; ΔDi-0S, 2-acetamide-2-deoxy-3-O-(D-gluco-4-enepyransyluronic acid)-D-galactose; ΔDi-6S, 2-acetamide-2-deoxy-3-O-(D-gluco-4-enepyransyluronic acid)-6-O-sulfo-D-galactose; ΔDi-4S, 2-acetamide-2-deoxy-3-O-(D-gluco-4-enepyransyluronic acid)-4-O-sulfo-D-galactose; ΔDi-diSD(2,6), 2-acetamide-2-deoxy-3-O-(2-O-sulfo-D-gluco-4-enepyransyluronic acid)-6-O-sulfo-D-galactose; ΔDi-diSE(4,6), 2-acetamide-2-deoxy-3-O-(D-gluco-4-enepyransyluronic acid)-4,6-bis-O-sulfo-D-galactose; contig, group of overlapping clones; PBS, phosphate-buffered saline; UTR, untranslated region.

and reverse primers (5'-TGCAGGGAACCTTCAAAGGACT-3' and 5'-TTCCTGCTGTATCCTTCCTCGG-3'), both derived from this EST, to clone a full-length (4671-bp) cDNA from a mouse 19-day embryo cDNA library (Origene) by pool selection, as described above. The 870-amino acid protein encoded by this cDNA is 93% identical to human Sulf-1, establishing that it is the mouse ortholog. We used this MSulf-1 cDNA to screen for matching genomic sequences on the Ensembl Mouse Genome Server (available on the World Wide Web at [www.ensembl.org/Mus\\_musculus](http://www.ensembl.org/Mus_musculus)) and identified a genomic sequence from mouse chromosome 1A3 (nt 12,800,000–12,915,000) that contained the entire 140-kb mouse Sulf-1 gene. To clone the mouse Sulf-2 ortholog, we screened the GenBank™ mouse EST data base with the predicted HSulf-2 protein sequence and thus identified EST AW763130 encoding the N terminus of a potential mouse Sulf-2. We obtained this EST from the IMAGE collection (clone 3155559), expanded the plasmid, and sequenced its insert. It contains a full-length 3603-bp cDNA, which encodes a mouse Sulf-2 ortholog of 875 amino acids (95% identical to HSulf-2). Using this cDNA to screen for matching genomic sequences, we identified a genomic sequence from mouse chromosome 2 (AL589873.11) containing the entire 82-kb MSulf-2 gene. The locus of this sequence was then mapped to mouse chromosome 2H3 by screening against the mouse genome assembly on the Ensembl server. We established that the mouse and human loci for the respective Sulfs are syntenic, using markers in the mouse genome informatics data base (available on the World Wide Web at [www.informatics.jax.org/searches/oxfordgrid\\_form.shtml](http://www.informatics.jax.org/searches/oxfordgrid_form.shtml)). The syntenic markers used were as follows. For Sulf-1, marker EYA-1: in humans, GenBank™ accession number AJ000098; in mice, AF097544. For Sulf-2, marker TDE-1: in humans, AF112227; in mice, AF181684.

### Sequence Analysis

Hydropathy analysis was done by the method of Kyte and Doolittle (7). The sulfatase domains and coiled-coil regions were identified by analysis with SMART (Simple Modular Architecture Research Tool), which is available on the World Wide Web at [smart.embl-heidelberg.de](http://smart.embl-heidelberg.de) (8). Adjustments in the boundaries of the sulfatase domains were made by comparing the sequences with those of QSulf-1 (AAK98515) and HG6S (P15586). Adjustments in the coiled-coil boundaries were made using the prediction program (9) located on the World Wide Web at [npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_lupas.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_lupas.html). The criteria for the two sulfatase signatures are described at [us.expasy.org/cgi-bin/get-prodoc-entry?PDOC00117](http://us.expasy.org/cgi-bin/get-prodoc-entry?PDOC00117). The signal sequences were predicted with the Signal P program (10) at [www.cbs.dtu.dk/services/SignalP/](http://www.cbs.dtu.dk/services/SignalP/). Alignment of proteins was performed with ClustalW (11) at [pbil.ibcp.fr/NPSA/npsa\\_clustalw.html](http://pbil.ibcp.fr/NPSA/npsa_clustalw.html). The dendrogram analysis was performed with Dendromaker ([www.cib.nig.ac.jp/dda/timanish/dendromaker/home.html](http://www.cib.nig.ac.jp/dda/timanish/dendromaker/home.html)). The choice of furin cleavage sites (12) was based on the consensus sequence (K/R)X(K/R)X(K/R)R-P1' in which at least two of the three (K/R) residues are present. At the P1' position, an amino acid with a hydrophobic aliphatic chain is not suitable.

### Analysis of mRNA Expression

To determine the expression of mRNAs for HSulf-1 and HSulf-2, PCR was applied to a panel of cDNAs prepared from 24 normal human tissues (Origene Inc.). Poly(A)<sup>+</sup> RNA was used to generate first strand cDNAs. The cDNAs were normalized across tissues to contain an equal concentration of actin transcripts. 371-bp HSulf-1 and 314-bp HSulf-2 cDNA products were amplified using the following PCR primers: for HSulf-1, forward 5'-CTCACAGTCCGGCAGAGCACGCGGAAC-3', reverse 5'-CACGGCGTTGCTGCTATCTGCCAGCATCC-3'; for HSulf-2, forward 5'-GAAAAGAGGCAGATTCACGTCGTTTCCAG-3', reverse 5'-ATCTGGTGCTTCTTTTGGGATGCGGGAG-3'. The conditions for denaturation, annealing, and extension of the template cDNA were as follows: 94 °C for 30 s; 64 °C for HSulf-1 and 55 °C for HSulf-2 for 30 s; and 72 °C for 1 min for 35 cycles, respectively. For

each source of cDNA, PCR was performed at four different cDNA concentrations (1×, 10×, 100×, and 1000×) using TITANIUM *Taq*DNA Polymerase (Clontech). The PCR products were then separated by electrophoresis on 2% agarose gels and visualized with ethidium bromide.

### Expression of Sulf Proteins in CHO Cells

DNA fragments that code for full open reading frames were amplified by PCR using HSulf-1 or HSulf-2 cDNA as template. The forward and reverse primers were as follows (boldface residues are restriction sites for cloning; see below): forward 5'-CAGCATCTCGAGACAATGAAGTATTCTTG-3' and reverse 5'-CAGGATGGATCCCCTTCCCATCCATCCCA-3' for HSulf-1; forward 5'-ATCTTACTCGAGAAGATGGGCCCCCGA-3' and reverse 5'-CACGATAAGCTTCCCTCCAGCCTTCCC-3' for HSulf-2. The conditions for denaturation, annealing, and extension of the template cDNA were as follows: 94 °C for 30 s; 60 °C for HSulf-1 and 64 °C for HSulf-2, all for 30 s; and 72 °C for 2 min for 35 cycles, respectively. These PCR products of HSulf-1 and HSulf-2 cDNAs were digested with *Xho*I and *Bam*HI, *Xho*I, and *Hind*III restriction enzymes, respectively, and subcloned into the corresponding sites of pcDNA3.1/Myc-His(-) (Invitrogen). This 5.5-kb vector is designed for overproduction of recombinant proteins with tandem C-terminal tags consisting of a polyhistidine metal-binding tag and the Myc epitope. Chinese hamster ovary (CHO) cells were grown in 10-cm dishes and transfected with 5 µg of pcDNA3.1/Myc-His(-)-HSulf-1 or -HSulf-2 using LipofectAMINE and Plus reagent (Invitrogen) according to the manufacturer's instructions. DNA was mixed with Plus reagent and incubated for 15 min at room temperature. The complexed DNA was combined with LipofectAMINE reagent, diluted in OptiMEM (Invitrogen), and incubated for 15 min at room temperature. The complexes were added to cells in culture dishes and incubated at 37 °C at 5% CO<sub>2</sub> for 5 h. After incubation, medium was replaced with OptiMEM. Cells were allowed to grow for an additional 48 h before the conditioned medium (CM) was collected.

For Western blotting, CM was concentrated on a Centricon30 microconcentrator (Millipore Corp.), separated by electrophoresis on reducing SDS-8% polyacrylamide gels (ISC BioExpress, Kaysville, UT), and blotted to ProBlott™ (Applied Biosystems, Foster City, CA). The membranes were blocked for 1 h with 5% nonfat milk and then incubated overnight with an anti-Myc antibody (Invitrogen) at a concentration of 0.22 µg/ml in 5% nonfat milk. When the anti-tetra-His antibody (Qiagen, Valencia City, CA) was used as the primary antibody, the membranes were blocked with 3% bovine serum albumin and incubated for 2 h with the antibody at 0.1 µg/ml in 3% bovine serum albumin. The membranes were washed and incubated with horseradish peroxidase goat anti-mouse IgG1 (0.4 µg/ml) (Caltag, Burlingame, CA) for 1 h before ECL detection reagents (Amersham Biosciences). To determine the amount of His fusion proteins in CM, a 6× His protein ladder from Qiagen was used as a standard, and protein amounts were determined by standard scanning techniques.

For analysis of the cellular distribution of the Sulfs, CM and the cells that conditioned the medium were collected separately. One aliquot of the cells was washed extensively and directly extracted with 1% Triton X-100 in phosphate-buffered saline (PBS) (with Ca<sup>2+</sup> and Mg<sup>2+</sup>) containing a mixture of protease inhibitors (Roche Molecular Biochemicals) at 4 °C overnight. The soluble extract was collected by centrifugation at 20,000 × *g* for 20 min and termed the "total cell lysate." An equal aliquot of washed cells was subjected to five cycles of freezing and thawing in 1 M NaCl in with protease inhibitors present. The supernatant constituted the "aqueous extract." The residue collected after centrifugation (20,000 × *g*, 20 min) was treated overnight with 1% Triton X-100 in PBS plus protease inhibitors to yield the "detergent extract."

## Arylsulfatase Assays of Expressed Sulf Proteins

100-fold concentrated CM derived from each transfection of CHO cells was dialyzed into 50 mM HEPES, pH 8.0. The His-tagged fusion proteins were bound to Ni-NTA resin (Qiagen) by rotation at 4 °C overnight and then washed with 50 mM HEPES, pH 8.0, three times. 250 mM imidazole in 50 mM HEPES, pH 8.0, was used to elute the His-tagged fusion proteins from Ni-NTA. The resins with bound material or the eluates were mixed with 10 mM 4-methylumbelliferyl sulfate (4-MUS) and 10 mM lead acetate in a total volume of 100  $\mu$ l. The reaction mixtures were incubated at 37 °C for various times. Reactions were terminated by the addition of 100  $\mu$ l of 0.5 M Na<sub>2</sub>CO<sub>3</sub>/NaHCO<sub>3</sub>, pH 10.7, to each 20- $\mu$ l aliquot of the reaction mixture. The fluorescence of 4-methylumbelliferone was measured on a multiwell plate reader, CytoFluorII (PerSeptive Biosystems, Framingham, MA), with an excitation wavelength of 360 nm and an emission wavelength of 460 nm. A substrate dose-response curve was performed over the concentration range of 1–10 mM 4-MUS.  $V_{\max}$  values were approximated from these data employing the estimate of the Sulf protein concentrations in the CM (see above).

To determine the pH dependence of the activity, His-tagged fusion proteins bound to Ni-NTA resin were washed with H<sub>2</sub>O three times and mixed with 10 mM 4-MUS, 10 mM lead acetate, and 50 mM HEPES (pH 8.0 or 7.0) or 50 mM sodium acetate buffer (pH 5.0 or 6.0). Human glucosamine-6-sulfatase was purchased from Glyko Inc. (Novato, CA) and was tested at a concentration of 1 milliunit.

In order to block the potential *N*-formylglycine modification of Cys<sup>87</sup> of HSulf-1 and Cys<sup>88</sup> of HSulf-2, cysteines 87 and 88 of HSulf-1 (designated as HSulf-1  $\Delta$ CC) and cysteines 88 and 89 of HSulf-2 (designated as HSulf-2  $\Delta$ CC) were mutated to alanines using the QuikChange™ XL site-directed mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions.

## Endoglucosamine-6-sulfatase Assays of the Sulfs

The procedures of Yoshida *et al.* (13) were adapted for these assays. CM from transfected CHO cells (wild-type Sulfs, vector control, or the double cysteine mutants) was prepared as described above. The standard reaction mixture contained 5  $\mu$ mol of Tris-HCl, pH 7.5, 1  $\mu$  mol of MgCl<sub>2</sub>, 10  $\mu$  g of porcine intestinal heparin (Sigma), and 20  $\mu$ l of the 100-fold concentrated CM in a total volume of 100  $\mu$ l. After incubation at 37 °C for 8 h, the reaction was stopped by heating at 100 °C for 2 min. A mixture of 1 milliunit of heparinase I (EC 4.2.2.7; Sigma), 0.25 milliunits of heparinase II (Sigma), and 0.1 milliunits of heparinase III (EC 4.2.2.8; Sigma) in 1.5  $\mu$ l of 50 mM Tris-HCl, pH 7.5, was added to the reaction mixture and incubated at 37 °C for 3 h. The digestion was stopped by heating at 100 °C for 2 min, and the mixture was filtered by centrifugation in an Ultra-free-MC filter (Millipore). The disaccharides of the digested heparin were then analyzed by HPLC on a Partisil-10 strong anion exchange column (Whatman, Fairfield, NJ) run at 41 °C. Disaccharides were eluted from the column by increasing the ionic strength as follows: time 0–5 min, 12 mM KH<sub>2</sub>PO<sub>4</sub>; time 5–40 min, gradient from 12 to 600 mM; time 40–45 min, 600 mM. Absorbance at 232 nm was monitored, and components were identified by comparison with authentic unsaturated disaccharide markers from Sigma (*i.e.*  $\Delta$ DiHS-0S,  $\Delta$ DiHS-6S,  $\Delta$ DiHS-NS,  $\Delta$ DiHS-(N,6)diS,  $\Delta$ DiHS-(N,2)diS, and  $\Delta$ DiHS-(N,6,2)triS). The endoglucosamine-6-sulfatase activities of human Sulf-1 and Sulf-2 against heparin proceeded linearly up to 10 h under these conditions. In other assays, 10  $\mu$ g of chondroitin 6-sulfate was employed as a substrate in the standard assay. After incubation with CM, 15 milliunits of chondroitinase ABC (EC 4.2.2.4; Sigma) was used to fragment the chondroitin sulfate. The standards used were as follows:  $\Delta$ Di-0S,  $\Delta$ Di-6S,  $\Delta$ Di-4S,  $\Delta$ Di-diSD (2,6), and  $\Delta$ Di-diSE (4,6). All were from Oxford GlycoSystems Inc.



## RESULTS

### Identification of Full-length cDNAs Encoding Mammalian Sulfs

Dhoot *et al.* (4) reported a partial cDNA (KIAA1077, GenBank™ accession number AB029000) encoding an apparent human ortholog of QSulf-1, which they designated HSulf-1. As described under “Experimental Procedures,” this sequence allowed us to assemble a 257-kb genomic contig that contained the entire HSulf-1 gene. Exon analysis, used in conjunction with EST information in the NCBI data base, led to the prediction of a full-length cDNA. We then cloned a 4669-bp full-length cDNA from a human prostate library and confirmed the predicted sequence obtained above. The cDNA contained an open reading frame of 2616 bp, predicting a protein of 871 amino acids (Fig. 1) (GenBank™ accession number AY101175). Based on this cDNA, we identified a highly related sequence (KIAA1247) in the NCBI data base. Using a similar approach to that for HSulf-1, we derived a genomic contig and then isolated a corresponding cDNA sequence from a human lung library. The 4286-bp sequence contained an open reading frame of 2613 bp (GenBank™ accession number AY101176), predicting a protein of 870 amino acids (Fig. 2). This protein is termed HSulf-2. We also identified cDNAs corresponding to apparent mouse orthologs, referred to as MSulf-1 (AY101178) and MSulf-2 (AY101177), as described under “Experimental Procedures.” Proteins of 870 and 875 amino acids are predicted from these cDNAs.

We then identified contigs containing the entire Sulf-1 and Sulf-2 genes from humans and mice. These mapped to single copy loci on chromosomes 8q13.2–13.3 and 20q13.12 in humans and on chromosomes 1A3 and 2H3 in mice, respectively. Using markers specified in the mouse genome informatics data base (14), we established that the loci of human and mouse Sulf-1 and those of Sulf-2 are respectively syntenic. The organization of these genes is depicted in Fig. 3 and described in further detail in Table I. The coding sequence of all four genes is split into 18–19 exons dispersed over 50–150 kb. All of the introns in the four Sulf genes feature canonical mammalian GT-AG splice junctions (15). The last exon is always the longest and encodes the last 9 residues of the gene product as well as the entire 3'-UTR of the mature cDNA. Finally, 5'-UTRs of varying lengths are encoded by a few noncoding exons relatively far 5' of the first coding exon.

Fig. 4A presents a dendrogram for the Sulf family proteins. Included in this comparison are QSulf-1 (4), RSulfFP1 (6), the mouse and human Sulfs reported in the present study, a *Caenorhabditis elegans* ortholog (CeSulf-1, AAA83618), and a *Drosophila melanogaster* Sulf ortholog (DmSulf-1, AAF55296). The analysis demonstrates that QSulf-1 is more homologous to Sulf-1 than to Sulf-2 in both humans and mice. By ClustalW alignment (11), QSulf-1 is 86% identical to HSulf-1 and 64% identical to HSulf-2. The species orthologs are highly homologous. Between mice and humans, the Sulf-1s are 93% identical, and the Sulf-2s are 94% identical. Comparison of the two Sulfs within the same species (mouse or human) shows an identity of 63–65%. RSulfFP1 shows 96% identity to MSulf-1, indicating that it is the rat ortholog of Sulf-1.

The structural organization of the mammalian Sulfs is very similar to that already described for both QSulf-1 (4) and RSulfFP1 (6) (Table II). Cleavable signal sequences, 22–27 amino acids in length, are predicted, and 10 or 11 potential *N*-linked sites are present. After the signal sequence, each protein exhibits a sulfatase domain of 372 amino acids (Table II), which is assigned based on comparison with the PFAM (protein family of alignments) data base of protein domain families (8). Among the lysosomal sulfatases, this domain is most closely related to the sulfatase domain of HG6S (accession number P15586), with a sequence identity of 44% versus HSulf-1 and 46% versus HSulf-2 (Fig. 4B). The overall homology falls off with the other cloned sulfatases (13–15% sequence identity), but, as noted for QSulf-1 (4), there is considerable conservation with the active site regions of these enzymes (Fig. 4C). Within the

amino-terminal region of the putative sulfatase domain is a cysteine (denoted by a *star* in Fig. 4, *B* and *C*), which is highly conserved in eukaryotic sulfatases. This residue is modified posttranslationally to *N*-formylglycine and, as such, is a part of the catalytic site of these enzymes (16).

Immediately following the sulfatase domain of the Sulfs is a hydrophilic region of 300–320 amino acids, containing a high content of charged amino acids, ~27% of which are basic and ~13% of which are acidic. At the C-terminus of each of the newly cloned Sulfs is a region of 108 residues, which shows significant homology to the C-terminal region of HG6S. In this region, there is complete conservation of 30 of 108 residues between HG6S and the two human Sulfs (Fig. 4*B*). Interestingly, this region also bears significant homology to a GlcNAc transferase from *Arabidopsis thaliana* (AAL60196). Over a 74-amino acid segment (626–699) of this plant protein, 8 of the aforementioned 30 amino acids are conserved. These comparisons suggest that the C-terminal regions of the Sulfs and the lysosomal G6S may be involved in recognition of glucosamine/GlcNAc components of substrates.

Inspection of the sequences reveals additional features that are shared by the mammalian Sulfs. Each protein exhibits a predicted coiled-coil structural unit of 34–35 residues. This feature is found in a corresponding position within the hydrophilic region of each protein (Table II), including QSulf-1 (not shown). Short coiled-coils serve as multimerization elements for a large number of both intracellular and extracellular proteins (9). The Sulf sequences also contain several consensus cleavage sites for furin, a trans-Golgi network endoprotease (12,17). These sites, which are mostly found in the hydrophilic regions, are highly conserved between the mouse and human orthologs of Sulf-1 and Sulf-2. As demonstrated below, processing of the secreted forms of the mammalian Sulfs appears to involve furin-mediated cleavage events.

### Expression of Sulf-1 and Sulf-2 mRNAs in Adult Tissues

The mRNA expression of HSulf-1 and HSulf-2 was examined by performing PCR on an array of normalized cDNAs derived from human tissues. Expression of the Sulfs was detected in several, but not all, human tissues with the highest levels in testes, stomach, skeletal muscle, lung, and kidney for Sulf-1 and in ovary, skeletal muscle, stomach, brain, uterus, heart, kidney, and placenta for Sulf-2 (Fig. 5).

### Expression and Proteolytic Processing of Sulf Proteins in CHO Cells

Sulf-1 and Sulf-2 cDNAs (mouse and human) were subcloned into the pcDNA3.1/Myc-His expression vector in order to generate recombinant proteins with a tandem arrangement of a His and a Myc tag at their C termini. Initially, we transfected COS-7 with cDNAs for Sulf-1 and Sulf-2 (mouse and human). Taking advantage of the Myc tag for Western blotting, we detected a 132-kDa band for each Sulf in detergent lysates of the transfected cells but failed to observe reactivity in the conditioned medium of these cells (data not shown). We also detected expression of the tagged proteins on the cell surface of transfected COS cells by immunofluorescence. These results parallel the findings with respect to QSulf-1 (4) and RSulfFP1 (6). However, when we used CHO cells for transfection, analysis of CM by Western blotting (anti-Myc antibodies) revealed a series of bands (Fig. 6). The results are shown for HSulf-1 and HSulf-2, but very similar patterns were observed for the mouse orthologs (data not shown). In each case, the highest molecular mass species had an apparent molecular mass of 132 kDa (Fig. 6*B*). Based on primary amino acid sequence, the calculated molecular masses of the tagged Sulf proteins were 100 kDa after cleavage of the signal sequences. The extra mass is attributable to *N*-glycosylation, since *N*-glycanase treatment of either HSulf-1 or HSulf-2 reduced the molecular mass of the 132-kDa species to ~100 kDa (data not shown). Interestingly, HG6S is also substantially glycosylated with 13 potential *N*-linked sites, of which at least 10 are used (5,18).

In addition to the 132-kDa species, HSulf-1 (Fig. 6A) showed two additional C-terminal tagged species with apparent molecular masses of 63 and 58 kDa. In other experiments, we also observed additional components of 46 and 44 kDa (Fig. 6B). Similarly, HSulf-2 yielded fragments of 64 and 60 kDa in addition to the 132-kDa species (Fig. 6A). Identical results were obtained when the anti-His antibody was used for Western blotting (data not shown). The quantities of His-tagged proteins in CM of transfected cells (48 h of collection) was considerably greater than that in total cell lysates (Fig. 6B). Furthermore, the proportion of lower molecular components was greatly enhanced in CM. As expected from the absence of transmembrane sequences in the primary sequences, solubilization of the His-tagged proteins from a crude membrane fraction did not require the presence of detergent in the extraction buffer (Fig. 6B).

This pattern of bands in CM was consistent with proteolytic processing. We considered the possibility that a proprotein convertase of the furin family, such as the broadly expressed endoprotease furin/PACE, might be responsible for this fragmentation, since there are several potential furin cleavage sites within the primary sequences of the Sulf-1 and Sulf-2 (Table II) and since CHO cells are known to express this enzyme (19). We therefore tested the effects of decanoyl-Arg-Val-Lys-Arg-chloromethylketone (CMK), an inhibitor of furin-type convertases (20), on the pattern of tagged species released into the CM of transfected CHO cells. As shown in Fig. 6C, CMK in a dose-dependent manner inhibited the appearance of the lower molecular weight forms in absolute amount (Fig. 6C) and as a percentage of total immunoreactive material (Fig. 6D). The inhibitory effects of CMK appeared to be more pronounced on Sulf-2 than Sulf-1. In a second approach, we co-transfected CHO cells with a cDNA for either HSulf-1 or HSulf-2 plus a cDNA encoding the Pittsburgh mutant of  $\alpha 1$ -antitrypsin, a reactive site variant that inhibits furin (19). As a control, cells were co-transfected with a cDNA for wild-type  $\alpha 1$ -antitrypsin, which does not inhibit furin. The co-expression of the active inhibitor effectively suppressed the appearance of the lower molecular forms for both Sulfs. Quantification revealed that the inhibition of the lower molecular weight species was about 75% in both cases (Fig. 6D). These results implicate furin (or a related serine protease) in the processing of both Sulfs within transfected CHO cells and indicate that the 132-kDa species are the precursors for the respective lower weight components.

### Arylsulfatase Activity of Expressed Proteins

The synthetic fluorogenic compound 4-MUS serves as a substrate for most sulfatases, both in the lysosomal and nonlysosomal classes (1). CHO cells were transfected with a cDNA for HSulf-1 or HSulf-2 or a vector control. Conditioned medium was collected and incubated with nickel resin (Ni-NTA) to bind the His-tagged recombinant proteins. The bead-bound material was assayed for activity on 4-MUS. For both enzymes, hydrolysis depended on time of reaction (Fig. 7A), concentration of substrate (Fig. 7B), and amount of CM (Fig. 7C). From the substrate dose-response curves of Fig. 7B and based on estimates of Sulf protein levels in CM, we approximated  $V_{\max}$  for the two Sulfs to be in the range of 1000–2000 nmol/min/mg of Sulf protein (see “Experimental Procedures”). Human glucosamine-6-sulfatase has a considerably lower  $V_{\max}$  against 4-MUS (100 nmol/min/mg), whereas the  $V_{\max}$  of human *N*-acetylgalactosamine-4-sulfatase (arylsulfatase B) is much higher (48,000) (18).

As described above, a conserved cysteine in the sulfatase domain of eukaryotic sulfatases is essential for their catalytic activities. We mutated the corresponding residues in Sulf-1 (Cys<sup>87</sup>) and Sulf-2 (Cys<sup>88</sup>) and the adjacent cysteines to alanines. Whereas the levels of the mutant proteins in CM were equivalent to those of the wild-type (not shown), arylsulfatase activity was completely lost for both proteins (Fig. 7C).

Since our foregoing analysis indicated that the Sulfs are targeted for secretion, we expected that their sulfatase activity would show optimal activity at neutral pH. As demonstrated in Fig.



7D, maximal activity was observed at pH values of 7 and 8. In contrast, the lysosomal enzyme, HG6S, assayed under the same conditions, showed measurable activity at pH 5 but none at pH 8, consistent with the acidic milieu of the lysosome. This finding dictated the use of neutral pH values (7.5–8) for the other assays (Figs. 7, A–C) and for the endosulfatase experiments described below.

### Endoglucosamine-6-sulfatase Activity of Expressed Proteins

The Sulfs were predicted to be active on heparin/heparan sulfate based on their sequence homology to glucosamine-6-sulfatase and the functional analysis of Dhoot *et al.* (4). Moreover, we suspected that the Sulfs might be endosulfatases, in contrast to glucosamine-6-sulfatase, which, like other lysosomal sulfatases, is an exosulfatase. To evaluate these possibilities, we treated intact heparin (10  $\mu$ g) with conditioned medium from Sulf-transfected CHO cells. To analyze activity on specific sulfation modifications, we digested the treated heparin with a mixture of bacterial heparinases and analyzed the disaccharide products by HPLC using standards (13). Both HSulf-1 CM and HSulf-2 CM produced an ~80% reduction in the amount of trisulfated units ( $\Delta$ DiHS-(N,6,2)triS) corresponding to IdoA2S-GlcNS6S and a parallel increase in that of disulfated units ( $\Delta$ DiHS-(N,2)diS) corresponding to IdoA2S-GlcNS (Fig. 8, A–C). There were no effects on the disulfated units  $\Delta$ DiHS-(N,6)diS, monosulfated, or nonsulfated units. Specific activities were calculated based on the volume of CM (Table III). The endosulfatase activities of Sulf-1 and Sulf-2 CM were increased 9- and 10-fold, respectively, relative to CM from mock-transfected cells (Table III). Mutation of the critical cysteines within the two Sulfs resulted in the complete loss of these activities (Table III). Further selectivity of the endosulfatase activity was indicated when we employed chondroitin 6-sulfate as a substrate. As shown in Fig. 8, D–F, *N*-acetylgalactosamine 6-sulfate residues in chondroitin 6-sulfate did not serve as substrates for the Sulfs.

## DISCUSSION

The present study was prompted by the discovery of QSulf-1 in the quail embryo (4). The novelty of this putative sulfatase is its presence on the cell surface of transfected cells and its apparent role in regulating an extracellular signaling pathway (*i.e.* the Wnt pathway) (see below). Also suggestive of an extracellular function is that RSulfFP1, the rat orthologue of QSulf-1, is strongly expressed in the floor plate of the developing nervous system (6), a region that plays a critical role in the patterning of the nervous system through its secretion of morphogens.

In the present paper, we report the cloning and expression of Sulf-1 and Sulf-2 in both humans and mice. MSulf-1 and HSulf-1 are clearly species orthologs of QSulf-1 and RSulfFP1, based on their very close sequence homology (94%). Sulf-2 is the product of a separate gene and represents a novel member of the family, which diverges somewhat in sequence identity (64%) but is very similar in length and overall structural organization to Sulf-1.

Several lines of our findings establish that the mouse and human Sulfs are targeted to the secretory pathway. First, in transfected COS cells, we observed expression on the cell surface, as was reported for QSulf-1 (4) and RSulfFP1 (6). Second, we found release of the expressed proteins into CM of the transfected cells. A pattern of several bands was observed, apparently arising through processing of the 132-kDa proteins. The predicted sequences of the Sulfs contain several consensus sites for furin cleavage, and these sites are strongly conserved between mice and humans. Furin is strongly implicated in processing of these proteins based on our findings with two classes of furin inhibitors (CMK and the Pittsburgh mutant of the  $\alpha$ 1-antitrypsin inhibitor). Furin is localized within the trans-Golgi network and, in some cases, on the cell surface and serves as a major processing enzyme in the secretory pathway. Interestingly, HG6S, which is targeted to the lysosomal compartment, contains no furin

cleavage sites in the sequence of the mature protein, although this enzyme is processed by other proteases including a signal peptidase (5). An important topic for future study is the subunit organization of the Sulfs in terms of their proteolytic fragments as well as the possible role of the highly conserved coiled-coil domains in multimerization of the subunits. Finally, the pH optimum for the arylsulfatase activity of the Sulfs was in the neutral range, consistent with an extracellular function for the enzymes.

The major class of sulfatases are lysosomal enzymes, which are involved in the degradation of sulfated substrates in an acidic environment. These enzymes are exosulfatases in that their activity is confined to sulfate esters on the nonreducing termini of saccharide units (1). The degradation of glycosaminoglycan chains involves the action of several of these sulfatases working coordinately with exoglycosidases, which remove one sugar unit at a time from a chain. One of the lysosomal sulfatases is G6S, which acts on glucosamine 6-sulfate and GlcNAc 6-sulfate in the degradation of HSPGs and possibly keratan sulfate. Deficiency of this enzyme in humans leads to the lysosomal accumulation of heparan sulfate chains and the accumulation GlcNAc 6-sulfate in the urine (5).

Most of the known sulfatases, including G6S, exhibit arylsulfatase activity, albeit to varying extents. The previous studies of QSulf-1 and RSulfFP1, which did not achieve expression of the proteins in a soluble aqueous form, did not report such an activity. In the present study, we have taken advantage of the fact that we were able to express the Sulfs in CM of transfected CHO cells and therefore could test for CM-associated arylsulfatase activity. Considerable activity, in fact greater than that reported for G6S (18), was found with 4-MUS as the substrate. Moreover, for each Sulf this activity required a highly conserved cysteine residue found within eukaryotic sulfatases (1).

Dhoot *et al.* (4) implicated QSulf-1 in the Wnt pathway of muscle specification. These investigators found that transfection of QSulf-1 into a muscle progenitor cell line increased responsiveness of these cells to Wnt in a co-culture model. Wnt is known to bind to heparan sulfate proteoglycans and to be regulated through these interactions (21,22). Dhoot *et al.* (4) speculated that QSulf-1 acts upon HSPGs to mobilize sequestered Wnt. A critical sulfatase activity was inferred, since mutation of the conserved cysteine (Cys<sup>89</sup>) blocked the potentiating effects of QSulf-1 on Wnt signaling. In the present study, we have directly shown that the Sulfs are active against intact heparin. Whereas heparan sulfate glycosaminoglycan chains contain highly sulfated regions (S domains) that are separated by regions of low sulfation, heparin chains contain a high density of sulfation throughout with an abundance of the trisulfated disaccharide, IdoA2S-GlcNS6S (23). Treatment of intact heparin at neutral pH with either HSulf-1 or HSulf-2 resulted in ~80% removal of sulfate from the 6-position of glucosamine within the trisulfated trisaccharides; however, there was no effect on (N,6)disulfated glucosamine within disaccharide units when the neighboring IdoA lacked 2-*O*-sulfation. These results establish that the Sulfs are endosulfatases with high selectivity for glucosamine 6-sulfate in the appropriate context within heparin. It is predicted that the Sulfs will show corresponding endosulfatase activities against S domains of heparan sulfate glycosaminoglycan chains.

A large number of growth factors, cytokines, and differentiation factors, as well as various classes of cell surface receptors, extracellular matrix proteins, and enzymes are known to bind to HSPGs with a demonstrated involvement of glucosamine 6-sulfate in many cases (24–26). These interactions are central to cell adhesion, the barrier functions of basement membrane and the extracellular matrix, and the growth and differentiation of cells. Given the expression of the Sulfs in a number of adult tissues, as shown herein, and during development (4,6), the Sulfs merit attention as to their roles in modulating HSPG interactions during normal and pathological processes.

## Acknowledgments

We thank Drs. Yibing Yan and Mark Sternlicht for advice on the use of furin inhibitors. We thank Diana Palmeri for assistance in deriving the original cDNA sequences and Claudia Fieger for advice on cloning the cDNAs into expression vectors.

## References

1. Parenti G, Meroni G, Ballabio A. *Curr Opin Genet Dev* 1997;7:386–391. [PubMed: 9229115]
2. Franco B, Meroni G, Parenti G, Levilliers J, Bernard L, Gebbia M, Cox L, Maroteaux P, Sheffield L, Rappold GA. *Cell* 1995;81:15–25. [PubMed: 7720070]
3. Puca AA, Zollo M, Repetto M, Andolfi G, Guffanti A, Simon G, Ballabio A, Franco B. *Genomics* 1997;42:192–199. [PubMed: 9192838]
4. Dhoot GK, Gustafsson MK, Ai X, Sun W, Standiford DM, Emerson CP Jr. *Science* 2001;293:1663–1666. [PubMed: 11533491]
5. Robertson DA, Freeman C, Morris CP, Hopwood JJ. *Biochem J* 1992;288:539–544. [PubMed: 1463457]
6. Ohto T, Uchida H, Yamazaki H, Keino-Masu K, Matsui A, Masu M. *Genes Cells* 2002;7:173–185. [PubMed: 11895481]
7. Kyte J, Doolittle RF. *J Mol Biol* 1982;157:105–132. [PubMed: 7108955]
8. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P. *Nucleic Acids Res* 2002;30:242–244. [PubMed: 11752305]
9. Lupas A. *Trends Biochem Sci* 1996;21:375–382. [PubMed: 8918191]
10. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. *Protein Eng* 1997;10:1–6. [PubMed: 9051728]
11. Thompson JD, Higgins DG, Gibson TJ. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]
12. Nakayama K. *Biochem J* 1997;327:625–635. [PubMed: 9599222]
13. Yoshida K, Miyauchi S, Kikuchi H, Tawada A, Tokuyasu K. *Anal Biochem* 1989;177:327–332. [PubMed: 2499215]
14. Blake JA, Eppig JT, Richardson JE, Davisson MT. *Nucleic Acids Res* 2000;28:108–111. [PubMed: 10592195]
15. Burset M, Seledtsov IA, Solovyev VV. *Nucleic Acids Res* 2001;29:255–259. [PubMed: 11125105]
16. Waldow A, Schmidt B, Dierks T, von Bulow R, von Figura K. *J Biol Chem* 1999;274:12284–12288. [PubMed: 10212197]
17. Steiner DF. *Curr Opin Chem Biol* 1998;2:31–39. [PubMed: 9667917]
18. Litjens T, Bielicki J, Anson DS, Friderici K, Jones MZ, Hopwood JJ. *Biochem J* 1997;327:89–94. [PubMed: 9355739]
19. Wasley LC, Rehemtulla A, Bristol JA, Kaufman RJ. *J Biol Chem* 1993;268:8458–8465. [PubMed: 8473289]
20. Wang X, Pei D. *J Biol Chem* 2001;276:35953–35960. [PubMed: 11470782]
21. Reichsman F, Smith L, Cumberledge S. *J Cell Biol* 1996;135:819–827. [PubMed: 8909553]
22. Lin X, Perrimon N. *Nature* 1999;400:281–284. [PubMed: 10421372]
23. Gallagher JT. *J Clin Invest* 2001;108:357–361. [PubMed: 11489926]
24. Bernfield M, Gotte M, Park PW, Reizes O, Fitzgerald ML, Lincecum J, Zako M. *Annu Rev Biochem* 1999;68:729–777. [PubMed: 10872465]
25. Lander AD, Selleck SB. *J Cell Biol* 2000;148:227–232. [PubMed: 10648554]
26. Esko JD, Lindahl U. *J Clin Invest* 2001;108:169–173. [PubMed: 11457867]

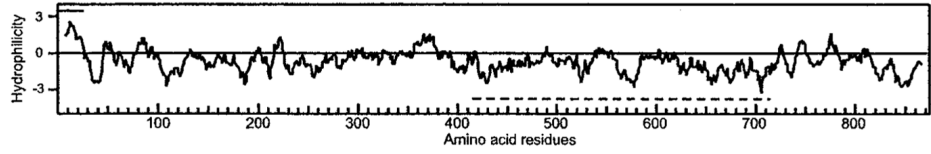




```

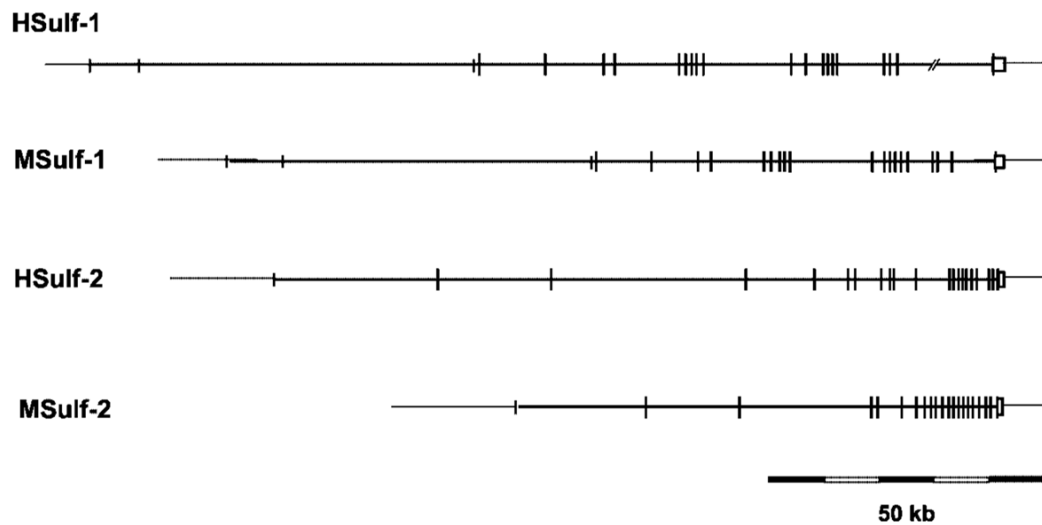
1  gggccatttctggacaacagctgctattttcacttgagcccaagtaatttctcggggagttctcggggcgcacaggcagctcggtttgcctcggattgagctcggggtcgcgccgg 120
121  cgcggcctctcaatggcaaatgtgtgctgggagggcagcgcgaggtcttcggcaaaagcagctcagattgttcagaccggggcgagctcctgtaaaagcagataaaagaaacattt 240
241  ataaagtgatcatagagggagcgcggcggggctgtcgactcccgaggaaatctgggtcctccagctccagagagagaaagagggaaagaggcagattcactt 360
361  cgtttccagccaagtgactgacgatggccctcctgaaattatcagcatattgatttattagcagatcccctggtttgtgtgttacgcaacaacacagctgcacaaggtcctgct 480
481  cgcttccctccctcgtttccagctcctgggccaatccacatctgtttcaactcccgccgagggcagcagagcagagtgctcgaatctcagatcgcagatgaagaggagcaggaaaga 600
601  aaacaagccacagagcaacttgagactcccgatcccaaaaagaagcaccagatcagcaaaaaaagagTGGGCCCGAGCTCGTGTCGTCGCGCAACTGTGTTCTCC 720
1  M G P P S L V L C L L S A T V F S 17
721  CTGTGGGTGGAAGCTCGGCCCTTCCTGTGCGCACCCGCTCAAAGGCGAGTTCAGAGGGCCCGCAGGAACATCCGCCCAACATCATCTGVTGTGACGGACGACAGGATGTGGAG 840
18  L L G G S S A C F L S H H R L K R L G R D R R I R P N I I L V L T D D Q D V E 57
841  CTGGTTCATGTCAGGTGATGAACAAGACCCGGCGCATCATGGAGCAGGGCGGGCGCACTTCATCAACGCCCTTCGTGACACACCATGTGCTGCCCTCAGGCTCCTCCATCTCACC 960
58  L G S M Q V M K T R R I M E Q G G A H F I N A F V T T P M C C P S R S S I L T 97
961  GGCAAGTACGTCACAACCACAACCTACACCAACAATGAGAATCTCGCCCTCTGGCAGGCACAGCAGAGGCCACCTTTCGCGTGTACCTCAATAGCACTGGCTACCCG 1080
98  G K Y V H N H N T Y T N N E C S S P S W Q A Q H E S R T F A V Y L S T G Y R 137
1081  ACAGCTTTCCTCGGGAAGTATCTTAATGAATACAACCGGCTCTACGTGCCACCCGGCTGGAAGGAGTGGGTGCGACTCCTTAAAACTCCCGCTTTTATAACTACACGCTGTCCGGAAC 1200
138  T A F F G K Y L N E Y G S Y V P P G W K E W V G L L K N S R F Y Y T L C R N 177
1201  GGGTGAAGAGCAGCGCTCCGACTACCAAGGATTAACCTACAGCACTCAACCAATGACAGCTGAGCTTCTCCGACGTCGAAGAGATGTATCCGACAGGCGACCTCCTC 1320
178  G V K E K H G S K D Y L T D L I T D S V S P F R T S K K M Y P H R F V L 217
1321  ATGGTCATCAGCCTGACGCCCCCAGCGCTGAGGATTCAGGCCCAACAATATTCAGCCCTTCCCAACGCATCTCAGCAGCATCAGCGCGACTACAACCGCCCAACCGGAC 1440
218  M V I S H A A P H G P E D S A P Q Y S R L F P A S Q H I T P S Y N Y A P N P D 257
1441  AAACACTGGATCATCGCTACACGGGGCCATGAAGCCCATCCACATGAATTCACCAACATGCTCCAGCGGAAGCGTTCAGACCCCTCATGTCCGTGGACGACTCCATGGAGCAGATT 1560
258  K H W I M R Y T G P M K P I H M E F T N M L Q R K R L Q T L M S V D D S M E T I 297
1561  TACAACATGCTGGTGTAGACGGGCGAGCTGGACAACAGTACATCGTATACCCGGCAGCCGGTACACATCGCCGAGTTTGGCTGTGAAGGAAATCCATCCGATGATGAGTTT 1680
298  Y N M L V E T G E L D N T Y I V T Y T A D H G Y H I G Q F G L V K G K S M P Y E F 337
1681  GACATCAGGTCGCTTCTACGTGAGGGCCCCAACGTGGAAGCGGCTGTGAATCCCCACATCGCTCCTCAACATTGACCTGGCCCCACCCTCTGGACATTCAGGCTGGACATA 1800
338  D I R V P F Y V R G P N V E A G C L N P H I V L N I D L A P T I L D I A G L D I 377
1801  CCTCGGATATGGACGGGAATCCATCTCAAGCTGTGGACAGCGCGCGGTAATCGGTTTCACTTGAAGAAAGATGAGGVTCTGGCGGACTCCTTCTGGTGGAGAGGCG 1920
378  P A D M T G K S I L K L L D T E R P F H L K R L Q R K R L Q T L M S V D D S M E T I 417
1921  AAGCTGTACACAAGAGACAATGACAAGTGGACGCCAGGAGGAGAACTTTCGCCAAGTACAGCGTGTGAAGGACTGTGTGACGCTGTGAGTACAGACGGCCTGAGGAG 2040
418  K L L H K R D A Q E E N F L P K Y Q R V K D L C Q R A E Y Q T A C E G 457
2041  CTGGGACAGAAGTGGCAGTGTGTGGAGGACGCCAGGGGAAGTGAAGCTCATAAGTCAAGGGCCCATGCGGCTGGGGCGGACGAGCCCTCTCCAACTCGTGCCCAAGTACTAC 2160
458  L G Q K W Q C V E D A T G K L K L H K C K G P M R L G G S R A L S N L V P K Y Y 497
2161  GGGCAGGCGAGCCCTGCACCTGTGACAGCGGGACTACAAGCTCAGCCTGCGGACGCCGAAAAAATCTTCAAGAAGAAGTACAAGCCAGCTATGTCCGCGACTCGCTCCATC 2280
498  G Q G S C D S G D Y K L S L A G R R K K L F K K K Y K A S Y V R S R S I 537
2281  CGCTCAGTGGCCTCGAGGTCGACGGGCGCAGGTTGACACAGTACGCTGGTGTGCGCCAGCCGAAACCTCAACAGCGGCATGGCCAGGGCCCTGAGGACCAAGATGACAAG 2400
538  R S V A I E V D G R V Y H V G L G D A Q P R N L T K R H W P G C P E D Q D D K 577
2401  GATGGTGGGACTTCAGTGGCAGTGGAGCCCTCCCGACTACTACGCCCAACCCCATTAAGATGACACATCGTGTGTACATCTTAGAGAACGACACAGTCCAGTGTGACCTGGACCTG 2520
578  D G G D F S G T G G L P D Y S A A N P I K V T H R C Y I L E D T V Q C D L D L 617
2521  TACAAGTCCCTGAGGCTGGAAAGACCAAGCTGACATCGACAGCAGATGAAACCCCTGCAGAACAAAATTAAGAACCTGAGGGAAGTCCGAGGTCACCTGAGGAAAAAGCGCCA 2640
618  Y K S L Q A W K D H K L H I D H E I E T L Q N K I K N L R E V R G H L K K R P 657
2641  GAAGAATGTGACTGTACAAAATCAGTACCACACCCAGCACAAGGCGCCCTCAAGCAGGAGGCTGACCTGCTCCTTTCAGGAAGGGCTGCAAGGAAGGACAAGGTGTGCTG 2760
658  E C D C H K I S Y H T Q H K G R L K H R G S S L H P F R K G L Q E A G K D K V W L 697
2761  TTGGGGAGCAGAAGCGCAGAAGAACTCCGCAAGCTGCTCAAGCGCTCGAGAACACAGCAGCTGCAGCATGCCAGGCTCAGTGTCTCACCCACGACAACCCAGCAGTGGCAGG 2880
698  L R E Q K R K K K L R K L L K R L Q N D T C S M P G L T C F T H D N Q H W Q T 737
2881  GCGCCTTCTGACACTGGGCTTCTGTGCTGCACCCAGCCAAACATAACACGTACTGGTGCATGAGGACCATCAATGAGACTCAAAATTCCTCTCTGTGAATTTGCAACTGGC 3000
738  A P F W T L G P F A C T S A N T Y N M C R T I E T H N F L F C E F A T G 777
3001  TTCCTAGACTCTTGTATCTCAACACAGACCCCTACCAGCTGATGAATGAGTGAACACACTGACAGGGATGCTCCTCAACGACTACAGCTACAGCTGAGGACTGAGGAGCTCAG 3120
778  F L E Y F L N T D P Y Q L M A V N T L D R D V L N Q L H V C L M E A G R S C K 817
3121  GGTTACAAGCAGTGAATCCCGGACTCGAAACATGACCTGGGACTTAAAGATGGAGGAAGCTATGAGCAATACAGGAGTTCAGCGCTCGAAAGTGGCCAGAATGAAGAGACCTTCT 3240
818  G Y K Q C N P R T R N M D L G L K D G G S Y E Q Y R Q F Q R K K W P E M K R P S 857
3241  TCCAAATCACTGGGCAACTGGAAGGCTGGAAGGTTAgaacaacagaggtggacctccaaaacatagaggcatcacctgactgcacaggaatgaaaaacctgtgggtgatt 3360
858  S K S L G Q L W E G W E G * 870
3361  tccagcagacctgtgctattggccaggagcctgagaaagcaagcagcactctcagtaacaatgaagattctggaggataaacagcaggagcagagataaactcaggaagtcatttt 3480
3481  tgcacctgtttgtttggattatacctcaccagctgcacaaaatgcatttttctgatatcaaaaagtcaccactaacctccccagaagctcacaaggaagaaacggagagcagagc 3600
3601  agagagatttctctggaaatttctcccaagggcgaaagtcattggaatttttaaatcataggggaaagcagctcctgttcaaatcctcttattctttgtgttgcacaagaaaggaac 3720
3721  taagaagcaggacagggcaacgtggagaggtgaaaacagtgacagagcgtttgacaatgagtcagtagcacaagagagatgacattacatagcatalaaccctgggtgcctctgaa 3840
3841  gaaactgccttcatgtatagtgactattacatgtaatacaatcagggaaactttaggggaaactaataagaaatcccaatttccaggagtggtggtgcaataaacctgtctgga 3960
3961  cagtgtaaaaagaaaaaaatctggacatttctgttccctgcagataccattctccctagatattctctgtatgtgctccagaactgagtttttttttaaggtactgaaaag 4080
4081  aatgaagtgtatgtgtcccaagtttgatgaaactgtatttgtaaaaaaatttgytagtttaagtattctctgatacagagtggtcaaaacccagcgaatgaccagcagctggtatgaa 4200
4201  gaaccttgacattttgaaaagggcatttcttggggaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaaaaataaaaaaa 4279

```



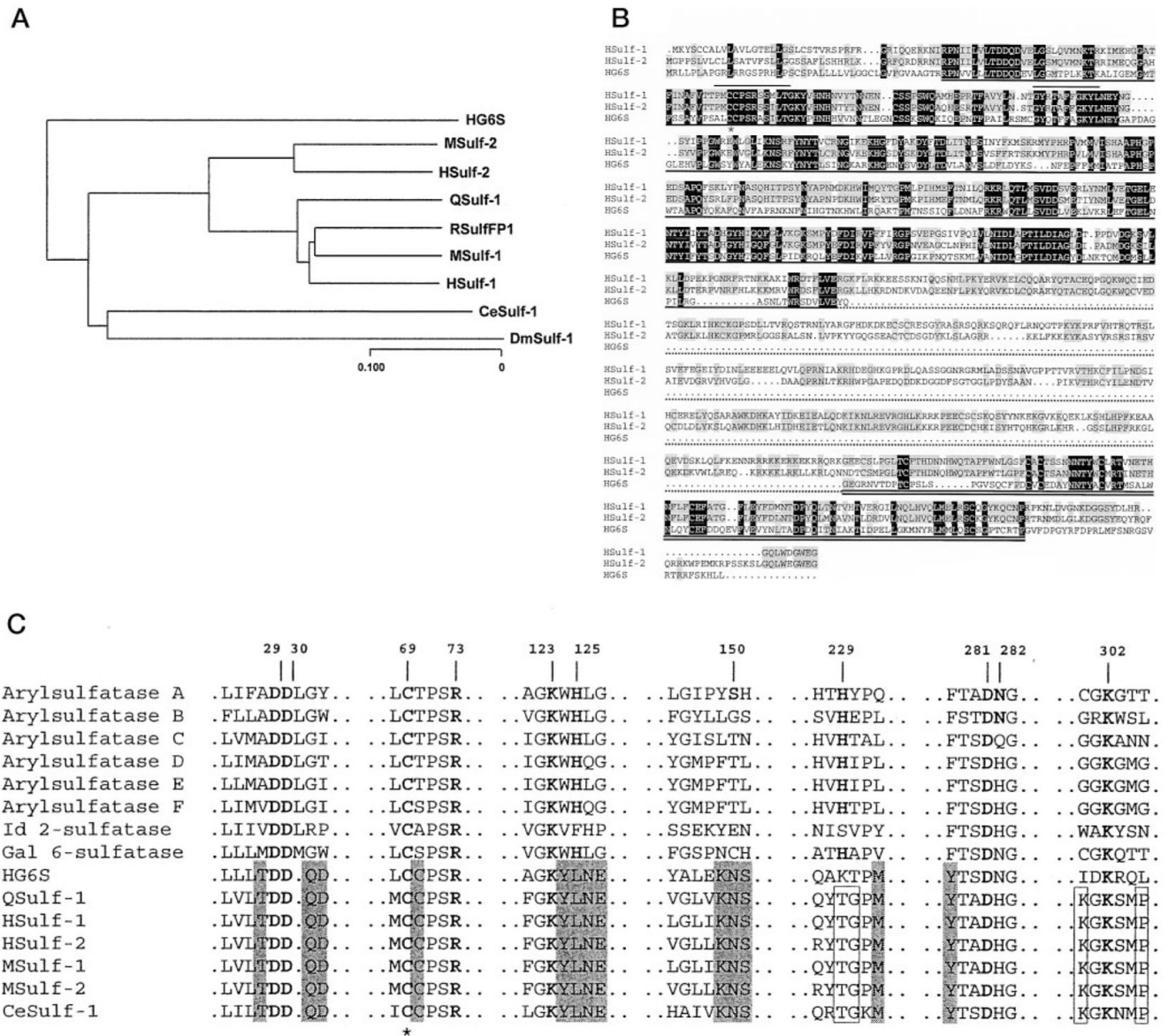
**Fig. 2. HSulf-2 cDNA sequence and deduced amino acid sequence**  
 Potential sites for N-linked glycosylation are indicated by the circles. Potential furin cleavages sites are indicated by the arrows. The coiled-coil boundaries are underlined. Bottom panel, hydropathy analysis of the predicted protein. The line above the plot indicates the predicted signal sequence, and the line below is the hydrophilic region. The GenBank™ accession number is AY101176.





**Fig. 3. Genomic organization of Sulf genes**

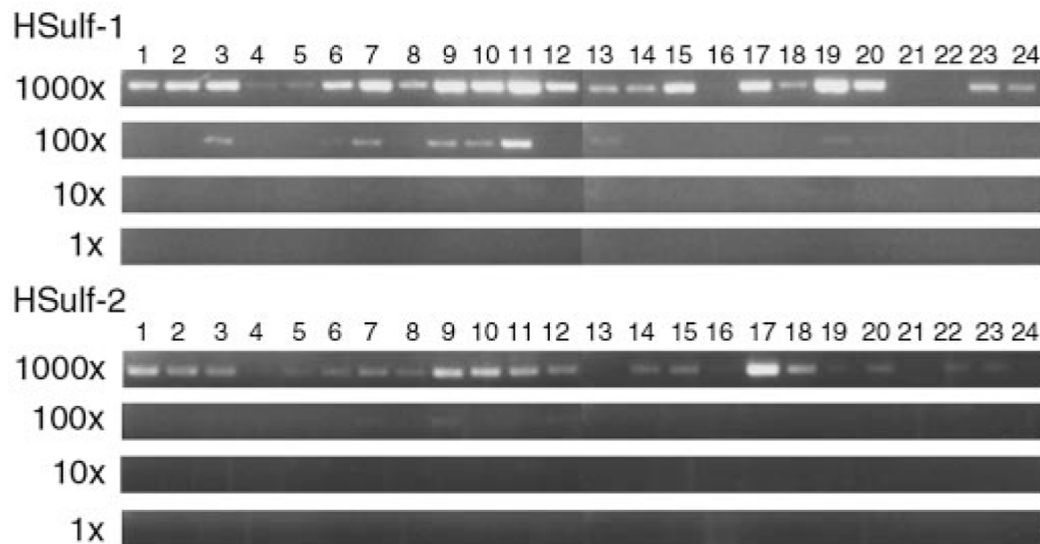
*Long vertical bars* indicate coding exons, whereas *short vertical bars* indicate noncoding exons. The box at the 3'-end denotes the last exon, which in all cases encodes the C-terminal 9 amino acids and also contains the sequence that is transcribed into the entire 3'-UTR of the mature cDNA. The intron-exon structures of the genes are described in greater detail in Table I. The sizes of the Sulf genes are  $\geq 167$  kb for HSulf-1 (gap within last intron after nucleotide 155,101 of the human Sulf-1 gene), 140.6 kb for MSulf-1, 129.2 kb for HSulf-2, and 81.7 kb for MSulf-2.



**Fig. 4. Sequence relationship among members of the Sulf family**

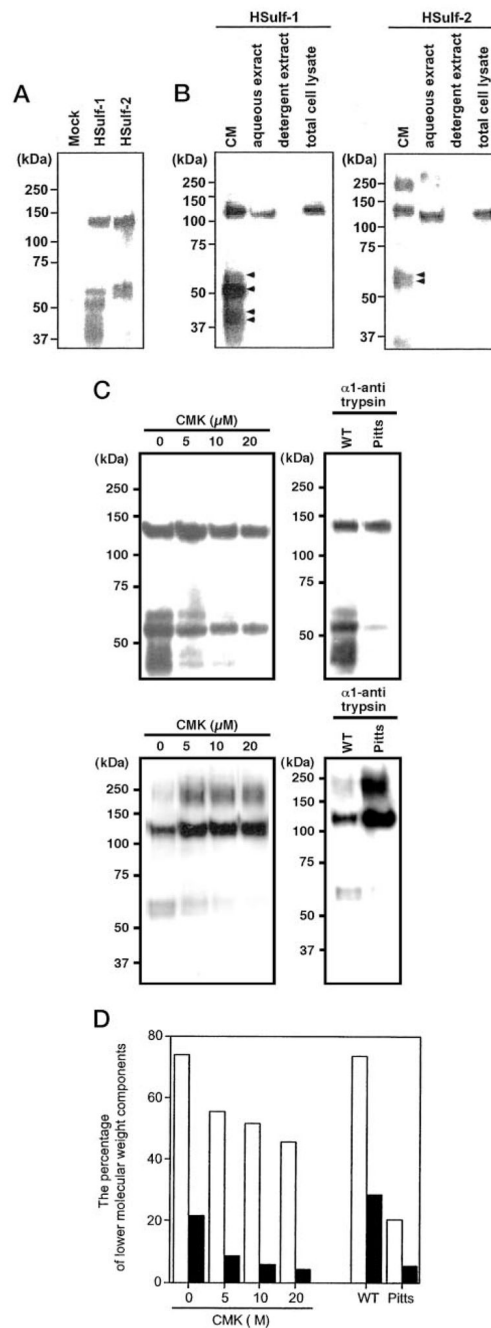
**A**, dendrogram of the members of the Sulf family based on protein sequences. **B**, ClustalW alignment of HSulf-1, HSulf-2, and HG6S. Dark shading indicates identity for all three proteins, and gray shading signifies identity in two of the three proteins. The dark underline denotes the sulfatase domain. The short lines above the sequence indicate the two signature regions for sulfatases (PS00523 and PS00149) containing consensus sequences that are highly conserved among the entire sulfatase family. The star indicates the cysteine residue that is predicted to undergo an *N*-formylglycine modification. The dotted underline designates the hydrophilic region of the Sulfs. The double underline indicates the C-terminal region with homology among the three proteins. **C**, conservation of sulfatase domain active site residues within the Sulfs. Shown are Sulf sequences containing residues identical to (boldface letters) or homologous to the putative active sites of arylsulfatase A, numbered from its N terminus, based on the crystal structure (16). Gray boxes denote active site amino acids that are shared with HG6S, QSulf-1, HSulfs, MSulfs, and CeSulf-1, and open boxes denote residues

that are shared within the Sulf family. The *star* indicates the cysteine residue that is predicted to undergo an *N*-formylglycine modification.



**Fig. 5. Detection of Sulf mRNAs in human tissues**

cDNAs derived from various tissues were subjected to PCR to amplify products for HSulf-1 (*top*) or HSulf-2 (*bottom*). The cDNAs were diluted in 10-fold steps over a 1000-fold range. The tissues were as follows: 1, brain; 2, heart; 3, kidney; 4, spleen; 5, liver; 6, colon; 7, lung; 8, small intestine; 9, muscle; 10, stomach; 11, testis; 12, placenta; 13, salivary; 14, thyroid; 15, adrenal; 16, pancreas; 17, ovary; 18, uterus; 19, prostate; 20, skin; 21, plasma blood leukocytes; 22, bone marrow; 23, fetal brain; 24, fetal liver.

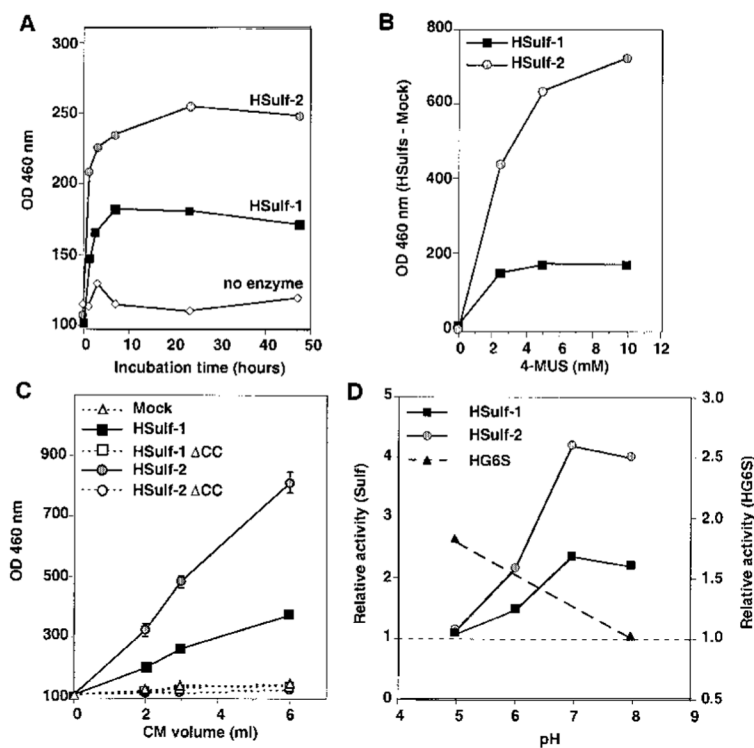


### Fig. 6. Expression of Sulf proteins in transfected CHO cells

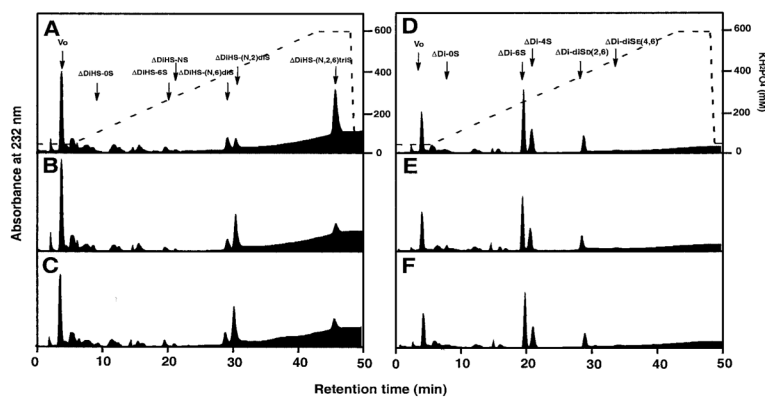
**A**, conditioned medium of CHO cells, transfected with a cDNA for HSulf-1, HSulf-2, or the empty vector was collected, concentrated, and subjected to Western blotting for the Myc tag. **B**, the amount of Sulfs was determined by Western blotting in conditioned medium of cultured cells (*CM*), a high salt aqueous extract of crude membranes (*aqueous extract*), a detergent extract of the membrane residue remaining after the aqueous extraction (*detergent extract*), and a detergent extract of whole cells (*total cell lysate*). Equal numbers were used for each fraction. The *arrows* indicate the lower molecular weight species present in the *CM*. **C**, CHO cells were transfected with a cDNA for HSulf-1 (*upper panel*) or HSulf-2 in the presence of varying concentrations of CMK or were co-transfected with a cDNA encoding either wild-type



$\alpha$ 1-antitrypsin (WT) or the Pittsburgh mutant of  $\alpha$ 1-antitrypsin (active form of inhibitor, Pitts). The conditioned medium from each transfection was collected, concentrated, and subjected to Western blotting for detection of the Myc tag. The high molecular mass form of Sulf-2 (~250 kDa) that is observed in the presence of CMK or the active form of the antitrypsin inhibitor is presumed to be a SDS-resistant dimer. *D*, the bars indicate the percentage of lower molecular weight components (63, 58, 46, and 44 kDa for Hsulf-1 and 64 and 60 kDa for HSulf-2) relative to the total quantities of tagged proteins. The data were obtained by scanning the blots in *C*. *Open bars* show the results for HSulf-1, and *solid bars* show those for HSulf-2.



**Fig. 7. Arylsulfatase activity of expressed Sulfs and lack of the activity in HSulf mutants**  
 HSulf-1, HSulf-2, or their mutated forms (HSulf-1  $\Delta$ CC and HSulf-2  $\Delta$ CC) were purified from the conditioned medium of transfected CHO cells by binding to Ni-NTA beads. *A*, the bead-bound material was tested for arylsulfatase activity as a function of time against 10 mM 4-MUS substrate at pH 8. The “no enzyme” control was based on testing conditioned medium from vector control-transfected CHO cells. No activity was detected in the absence of added substrate (not shown). The same results were obtained in three different experiments. *B*, the concentrated conditioned medium was tested for arylsulfatase activity at pH 8 for 2 h at different concentrations (1–10 mM) of 4-MUS. To eliminate background effects, the activity in vector control material was subtracted from that of Sulf-transfected material. *C*, the eluted material from Ni-NTA was tested for arylsulfatase activity at pH 8 for 2 h as a function of input volume of conditioned medium. The same results were obtained in three different experiments. *D*, bead-bound Sulfs were tested for arylsulfatase activity (1 h) at the indicated pH values. The activity of each Sulf was determined relative to that of beads exposed to an equivalent volume of vector-control conditioned medium. The activity of HG6S was determined (24-h incubation) relative to that of the buffer. The same results were obtained in three different experiments.



**Fig. 8. Endoglucosamine-6-sulfatase activity of expressed Sulf6**

The conditioned medium of CHO cells transfected with the empty vector alone (*A* and *D*), HSulf-1 (*B* and *E*), or HSulf-2 (*C* and *F*) was prepared as described under “Experimental Procedures.” Porcine intestinal heparin (*A–C*) or shark cartilage chondroitin 6-sulfate (*D–F*) were incubated with the conditioned medium and then subsequently digested with either a mixture of bacterial heparinases or chondroitinase ABC. The resulting disaccharide fractions were analyzed by as described. The *arrows* correspond to the elution positions of authentic unsaturated disaccharide markers. The *dotted lines* indicate the concentrations of  $\text{KH}_2\text{PO}_4$  used for elution.

Table 1

## Genomic organization of Sulf genes

Exon	HSulf-1		MSulf-1		HSulf-2		MSulf-2	
	Size	Amino acids	Size	Amino acids	Size	Amino acids	Size	Amino acids
	<i>bp</i>							
1	280	5'-UTR	255	5'-UTR	569	5'-UTR	141 <sup>a</sup>	5'-UTR
2	95	5'-UTR	95	5'-UTR	275	5'-UTR, 1-58	267	5'-UTR, 1-58
3	73	5'-UTR	73	5'-UTR	240	59-138	240	59-138
4	232	5'-UTR, 1-57	231	5'-UTR, 1-57	152	139-189	152	139-189
5	240	58-137	240	58-137	170	190-246	170	190-246
6	152	138-188	152	138-188	151	247-296	151	247-296
7	170	189-257	170	189-257	176	297-355	176	297-355
8	151	246-295	151	246-295	129	356-398	129	356-398
9	176	296-354	176	296-354	57	399-417	57	399-417
10	129	355-397	129	355-397	130	418-460	130	418-460
11	57	398-416	57	398-416	196	461-525	211	461-530
12	130	417-459	130	417-459	229	526-602	229	532-607
13	217	460-531	217	460-531	97	603-634	97	608-639
14	256	532-617	253	531-616	95	635-666	95	640-671
15	97	618-649	97	617-648	60	667-686	60	672-691
16	95	650-681	95	649-680	170	687-742	170	692-747
17	66	682-703	66	681-702	143	743-790	143	748-795
18	176	704-761	176	703-760	124	791-831	124	796-836
19	143	762-809	143	761-808	34	832-843	34	837-848
20	124	810-850	124	809-849	54	844-861	54	849-866
21	34	851-862	34	850-861	986	862-870, 3'-UTR	902	867-875, 3'-UTR
22	1546	863-871, 3'-UTR	1566	862-870, 3'-UTR				

<sup>a</sup>Exon 1 derived from the 5'-end of EST AW210030 (nr 24-164).

**Table II**

## Features of the Sulf proteins

Feature	HSulf-1	MSulf-1	HSulf-2	MSulf-2
Length (amino acids)	871	870	870	875
No. of <i>N</i> -linked sites	10	10	11	11
Signal sequence <sup>a</sup>	1–22	1–27	1–24	1–24
Sulfatase domain <sup>b</sup>	42–414	42–414	43–415	43–415
Hydrophilic region	415–735	415–734	416–715	416–721
G6S-related region	736–843	735–842	717–824	722–829
Coiled-coil region <sup>c</sup>	639–673	638–672	623–658	629–663
Furin cleavage sites	408–409	408–409	409–410	409–410
	576–577	575–576	423–424	423–424
	661–662	660–661	538–539	543–544
	669–670		565–566	
	732–733	731–732	646–647	651–652
		656–657	661–662	
		848–849	853–854	

<sup>a</sup>Signal sequences were predicted with the Signal P program.

<sup>b</sup>Sulfatase domains were identified by analysis with SMART.

<sup>c</sup>Coiled-coil regions were identified by analysis with SMART.



**Table III**  
**Endoglucosamine-6-sulfatase activity of the Sulfs**

Endoglucosamine-6-sulfatase activity was measured against intact heparin as described under “Experimental Procedures” using CM derived from CHO cells transfected with the indicated vectors. HSulf-1  $\Delta$ CC and HSulf-2  $\Delta$ CC denote the C87A/C88A and C88A/C89A mutants, respectively. Endosulfatase activity was defined by calculating the moles of unsaturated trisulfated disaccharides from which the sulfate group on the C-6 position was liberated in the standard assay (Fig. 8) by the Sulfs as compared with levels in untreated control samples. Values shown are means  $\pm$  S.D. based on three independent reactions.

Plasmid	Endoglucosamine-6-sulfatase activity
	<i>pmol/h/ml medium</i>
Mock	9.4 $\pm$ 0.2
HSulf-1	87.8 $\pm$ 0.1
HSulf-1 $\Delta$ CC	7.7 $\pm$ 1.2
HSulf-2	96.6 $\pm$ 0.5
HSulf-2 $\Delta$ CC	5.6 $\pm$ 0.6