Molecular cloning and characterization of chondroitin-4-O-sulfotransferase-3: A novel member of the HNK-1 family of sulfotransferases

DOI:
10.1074/jbc.M204907200

Document Version
Final published version

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 38, Issue of September 20, pp. 34766–34772, 2002
Printed in U.S.A.

Molecular Cloning and Characterization of Chondroitin-4-*O*-sulfotransferase-3

A NOVEL MEMBER OF THE HNK-1 FAMILY OF SULFOTRANSFERASES*

Received for publication, May 17, 2002, and in revised form, June 19, 2002
Published, JBC Papers in Press, June 21, 2002, DOI 10.1074/jbc.M204907200

Hyung-Gyoo Kang‡§, Matthias R. Evers§¶, Guoqing Xia§¶, Jacques U. Baenziger‡, and Melitta Schachner¶‖

From the ‡Department of Pathology, Washington University School of Medicine, St. Louis, Missouri 63110 and ¶Zentrum fuer Molekulare Neurobiologie, Universitaet Hamburg, Martinistrasse 52, D-20246 Hamburg, Germany

We have identified and characterized an *N*-acetylgalactosamine-4-*O*-sulfotransferase designated chondroitin-4-sulfotransferase-3 (C4ST-3) (GenBank™ accession number AY120869) based on its homology to HNK-1 sulfotransferase (HNK-1 ST). The cDNA predicts an open reading frame encoding a type II membrane protein of 341 amino acids with a 12-amino acid cytoplasmic domain and a 311-amino acid luminal domain containing a single potential *N*-linked glycosylation site. C4ST-3 has the greatest amino acid sequence identity when aligned with chondroitin-4-*O*-sulfotransferase 1 (C4ST-1) (45%) but also shows significant amino acid identity with chondroitin-4-*O*-sulfotransferase 2 (C4ST-2) (27%), dermatan-4-*O*-sulfotransferase 1 (29%), HNK-1 ST (26%), *N*-acetylgalactosamine-4-*O*-sulfotransferase 1 (26%), and *N*-acetylgalactosamine-4-*O*-sulfotransferase 2 (23%). C4ST-3 transfers sulfate to the C-4 hydroxyl of β1,4-linked GalNAc that is substituted with a β-linked glucuronic acid at the C-3 hydroxyl. The open reading frame of C4ST-3 is encoded by three exons located on human chromosome 3q21.3. Northern blot analysis reveals a single 2.1-kilobase transcript. C4ST-3 message is expressed in adult liver and at lower levels in adult kidney, lymph nodes, and fetal liver. Although C4ST-3 and C4ST-1 have similar specificities, the highly restricted pattern of expression seen for C4ST-3 suggests that it has a different role than C4ST-1.

We and others have recently reported the cloning and functional characterization of members of the HNK-1 family of sulfotransferases. These include human HNK-1 sulfotransferase (HNK-1 ST)[1] itself (1, 2), GalNAc-4-ST1 (3–5), GalNAc-4-ST2 (5, 6), dermatan-4-sulfotransferase-1 (D4ST-1) (7), chondroitin-4-sulfotransferase-1 (C4ST-1) (8, 9), and chondroitin-4-sulfotransferase-2 (C4ST-2) (8). With the exception of HNK-1 ST itself, which transfers sulfate to the C-3 hydroxyl of terminal glucuronic acid in the sequence GlcUAβ1,3Galβ1,4GlcNAc-R to produce the HNK-1 epitope SO_4-3-GlcUAβ1,3Galβ1,4GlcNAc-R, the members of this family of sulfotransferases are all GalNAc-4-sulfotransferases. GalNAc-4-ST1 and GalNAc-4-ST2 both transfer sulfate to the C-4 hydroxyl of terminal β1,4-linked GalNAc on *N*-linked oligosaccharides such as those found on the glycoprotein hormones lutropin and thyrotropin (10, 11). Whereas C4ST-1, C4ST-2, and D4ST-1 are also GalNAc-4-*O*-sulfotransferases, they only transfer sulfate to the C-4 hydroxyl of internal β1,4-linked GalNAc moieties within the repeating disaccharide sequences found in chondroitin and dermatan (7–9). We now report the cloning of another member of this family of sulfotransferases. Like C4ST-1, this new sulfotransferase, chondroitin-4-sulfotransferase-3 (C4ST-3) transfers sulfate to the C-4 hydroxyl of β1,4-linked GalNAc that is flanked by GlcUA residues in chondroitin. In contrast to C4ST-1, C4ST-3 is labile at 37 °C and has a restricted distribution, suggesting that it may have a unique biological role *in vivo*.

EXPERIMENTAL PROCEDURES

Molecular Cloning of a cDNA Encoding Human Chondroitin-4-O-sulfotransferase-3—A human genomic BAC clone, RP11-390G14, derived from human chromosome 3 (GenBank™ accession number AC024558) was identified in TBLASTN searches (12) against the nonredundant data base at the NCBI using deduced protein sequences of human and rat HNK-1 sulfotransferases (1, 2) as query sequences. The putative open reading frame (ORF) encodes a protein that shows homology to the luminal domain of C4ST-1, a member of the HNK-1 family of sulfotransferases (8, 9). Subsequent BLASTN queries of the dbEST data set using this region of homology identified two matching EST sequences (GenBank™ accession numbers BF448098 and AI074149, respectively). The corresponding cDNA clones (IMAGp998H017603 and IMAGp998M214153, respectively) were obtained from the RZPD (Berlin, Germany) (13) and sequenced on both strands. The partial ORF of C4ST-3, encoding most of its luminal domain, was amplified by PCR. The 5′-specific primer, 5′-cca agc ttg cca cca tgt ttg gaa aca gag ccc t-3′, contains a *Hin*dIII restriction enzyme site, the Kozak consensus sequence GCCACC, and an artificial start codon. The 3′-specific primer, 5′-gct cta gac tag agc agc cgc agg tag ga-3′, contains an *Xba*I site and a stop codon. The product was directionally subcloned into pcDNA3.1 (Invitrogen, Karlsruhe, Germany) and designated pcDNA3.1-C4ST-3-ORF309.

Subsequent BLASTN queries of the dbEST using the partial ORF of C4ST-3 identified an EST (GenBank™ accession number BI908522) that overlaps the NH_2 terminus of the sequence used to construct

* This work was supported by National Institutes of Health Grant R01-DK41738 (to J. U. B.), Deutsche Forschungsgemeinschaft Grant SCHA185/15-1 (to M. S.), and a German Academic Exchange Service postdoctoral fellowship (to G. X.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY120869.

§ These three authors contributed equally to this work.

‖ To whom correspondence should be addressed: Zentrum fuer Molekulare Neurobiologie, Universitaet Hamburg, Martinistr. 52, 20246 Hamburg, Germany. Tel.: 49-40-42803-6246; Fax: 49-40-42803-6248; E-mail: melitta.schachner@zmnh.uni-hamburg.de.

[1] The abbreviations used are: HNK-1 ST, HNK-1 sulfotransferase; GlcUA, glucuronic acid; C4ST, chondroitin-4-*O*-sulfotransferase; GalNAc-4-ST, *N*-acetylgalactosamine-4-*O*-sulfotransferase; D4ST, dermatan-4-*O*-sulfotransferase; MCO, $(CH_2)_8$-$COOCH_3$; ORF, open reading frame; PAPS, 3′-phosphoadenosine-5′-phosphosulfonate; UTR, untranslated region; EST, expressed sequence tag; NTA, nitrilotriacetic acid; CHO, Chinese hamster ovary.

A

```
1    ATGGGGAGGCGCTGCTGCCGGCGGCGCGTGCTGGCGGCCGCCTGTCTGGGCGCCGCGCTCCTGCTCCTATGCGCCGCGCCCCGCTCCCTGCGCCCGGCATTTGGAAACAGAGCCCTGGGC  120
1    M  G  R  R  C  C  R  R  R  V  L  A  A  A  C  L  G  A  A  L  L  L  L  C  A  A  P  R  S  L  R  P  A  F  G  N  R  A  L  G   40

121  TCCAGCTGGCTTGGTGGGGAGAAGAGAAGCCCCCTGCAGAAGCTCTATGACCTGGATCAGGACCCGCGCTCGACCCTGGCGAAGGTGCACCGTCAGCGGCGCGACCTGCTGAACAGCGCC  240
41   S  S  W  L  G  G  E  K  R  S  P  L  Q  K  L  Y  D  L  D  Q  D  P  R  S  T  L  A  K  V  H  R  Q  R  R  D  L  L  N  S  A   80

241  TGTAGCCGCCACTCACGCCGGCAGCGCCTGCTACAGCCGGAGGACCTGCGGCACGTGCTGGTGGACGACGCGCATGGCCTGCTCTACTGCTACGTGCCCAAGGTGGCCTGCACCAACTGG  360
81   C  S  R  H  S  R  R  Q  R  L  L  Q  P  E  D  L  R  H  V  L  V  D  D  A  H  G  L  L  Y  C  Y  V  P  K  V  A  C  T  N  W  120

361  AAGCGCGTGCTGCTGGCGCTGAGCGGCCAAGCCCGCGGCGACCCGCGCGCCATCTCCGCGCAAGAGGCGCACGCGCCTGGCCGCCTGCCCTCACTGGCCGACTTCAGCCCCGCCGAGATC  480
121  K  R  V  L  L  A  L  S  G  Q  A  R  G  D  P  R  A  I  S  A  Q  E  A  H  A  P  G  R  L  P  S  L  A  D  F  S  P  A  E  I  160

481  AACCGGCGCCTGCGCGCCTACTTGGCCTTCCTGTTCGTGCGGGAGCCCTTCGAGCGCCTGGCATCGGCTTACCGCAACAAGCTCGCGCGCCCCTACAGCGCCGCCTTCCAGAGGCGCTAC  600
161  N  R  R  L  R  A  Y  L  A  F  L  F  V  R  E  P  F  E  R  L  A  S  A  Y  R  N  K  L  A  R  P  Y  S  A  A  F  Q  R  R  Y  200

601  GGTGCACGCATCGTTCAGCGCCTGCGGCCGCGCGCGCTCCCCGACGCCCGGGCCCGCGGCCACGACGTGCGCTTCGCGGAGTTCCTGGCCTACCTGCTGGACCCGCGCACGCGGCGTGAG  720
201  G  A  R  I  V  Q  R  L  R  P  R  A  L  P  D  A  R  A  R  G  H  D  V  R  F  A  E  F  L  A  Y  L  L  D  P  R  T  R  R  E  240

721  GAGCCCTTCAACGAGCACTGGGAGCGCGCGCACGCGCTCTGCCACCCGTGTCGCCTCCGCTACGACGTCGTGGGCAAGTTCGAGACGCTGGCGGAGGACGCGGCCTTCGTGCTGGGCCTG  840
241  E  P  F  N  E  H  W  E  R  A  H  A  L  C  H  P  C  R  L  R  Y  D  V  V  G  K  F  E  T  L  A  E  D  A  A  F  V  L  G  L  280

841  GCGGGCGCATCCGACCTGAGCTTCCCTGGGCCGCCGCGGCCCCGGGGAGCCGCCGCCTCCCGCGACCTGGCAGCGCGCCTCTTCCGGGACATCAGCCCCTTCTACCAGCGGCGCCTCTTC  960
281  A  G  A  S  D  L  S  F  P  G  P  P  R  P  R  G  A  A  A  S  R  D  L  A  A  R  L  F  R  D  I  S  P  F  Y  Q  R  R  L  F  320

961  GACCTCTACAAGATGGACTTCCTGCTTTTCAACTACTCCGCCCCCTCCTACCTGCGGCTGCTCTAGCGGTCCTGGAGGTCCTGTGGCCACGCGGGGCAAGTGCCTTTCCGACAAGACCCC 1080
321  D  L  Y  K  M  D  F  L  L  F  N  Y  S  A  P  S  Y  L  R  L  L  *                                                          341

1081 CGGGGAATGCAGGTGCTGCCGGCCCCAGGACCCCTCTTCAAGAGCCACTGCGTGCACTCACCTGGCCGCCGGGCCAGCGGGCGCAGGGCACACCTGGCCAGGCTTGGGGGCAGCCCATCT 1200
1201 CAGGTGGCCCTGCACGCGTGTGCCTGCCTCGGCCTGTCGCCTGAGGCCTGCTTCCTCCACTTGCTCCAGCTGACAGGCACCTCTCCAGGCCCCGTAGATGGGCAAGGACTTGATAACCAG 1320
1321 GGTTTTAGGCTTTTAAAGGCCATTTTGGGGGTCAGCCCTGCCCCTGAACCTGTTCATGGTGCATCAGAACATAATGCTGACACCGGTGTCAGTGTGGCCCGAGCCTGTGTCCTCCCCACC 1440
1441 TCGCCCACCCTGGCAAGGACAGCTGCGGCCAAGGACGAAAGCCCTCCCTTGGCTGGCCTCACGATGGGGCCGTCCCGGGAGCCAGGTGGGAGCTGCCTTCCACTGCCATCGGGTCTCCTC 1560
1561 TCCTCTCCCACGCGGCTGGCCCTACCCAGGCGCCACCTTCGGTCTCAGTCTGGCAAGACGCTGGGTCTTCAGGCTCCATGCCAACAGAGCCCCTGGTGCAATGCGGTCACAGGTTTTATG 1680
1681 GGACTTTGGTGAGCTGGGCGGTCATGGTTTTGAAATAAATGTATTTTGTTACTTTCTGAAAAAAAAA 1747
```

B

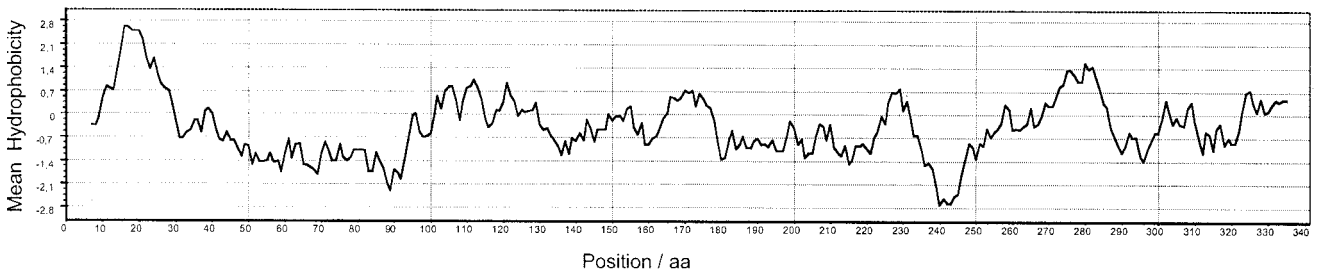


FIG. 1. **Nucleotide and deduced amino acid sequence of human C4ST-3 cDNA (GenBank™ accession number AY120869).** *A*, the predicted amino acid sequence of C4ST-3 is denoted by *capital letters* below the nucleotide sequence. The single predicted membrane-spanning domain and a single potential *N*-linked glycosylation site are indicated by the *thick underlines* and by the *underlines* with a *black dot* below the glycosylated Asn, respectively. *B*, Kyte-Doolittle mean hydrophobicity plot for C4ST-3 (scan window size was 13 amino acids).

pcDNA3.1-C4ST-2-ORF309. BI908522 was used to evaluate a segment of the working draft of chromosome 3 (GenBank™ accession number NT_005588; also see below) for the presence of exons predicted by the program FGENES (version 1.6). Based on genomic sequence information, the start codon of C4ST-3 was predicted to be ~40 bp to the 5′-end of the genomic sequence represented by BI908522. The 5′-terminal part of the C4ST-3 ORF was amplified by PCR from human placenta first-strand cDNA (generated with Omniscript Reverse Transcriptase; Qiagen, Hilden, Germany) using the 5′-specific primer 5′-atg ggg agg cgc tgc tgc cgg cgg cgc g-3′ and the 3′-specific primer 5′-ttg atc tcg gcg ggg ctg aag t-3′. The 482-bp product was subcloned into the pGEM®-T Easy vector (Promega) and sequenced on both strands. Finally, the full-length ORF was assembled in the pBluescript® KS vector (Stratagene, Amsterdam, The Netherlands) by sequentially subcloning an 850-bp *Pst*I-*Xba*I fragment derived from pcDNA3.1-C4ST-3-ORF310 and an additional 170-bp *Hin*dIII-*Pst*I fragment from the pGEM®-T construct above to produce pBKS-C4ST-3-ORF341.

5′-Rapid amplification of cDNA ends using Marathon Ready cDNA (adult kidney; BD Biosciences-CLONTECH, Heidelberg, Germany) as a template was performed using a number of different primer combinations; however, the 5′-UTR could not be amplified. The 3′-UTR was obtained by amplifying the cDNA using two different primers predicted on the basis of the genomic sequence. The 5′-specific primer 5′-tca act act ccg ccc cct cct acc-3′ and the 3′-specific primer 5′-acc gcc cag ctc acc aaa gtc c-3′ were used to amplify the predicted 3′-end of the C4ST-3 cDNA from a SMART™ cDNA library (BD Biosciences-CLONTECH) derived from a mixture of several human tissues/cell lines (*e.g.* generated from mRNA representing the human kidney cell line XYZ). The 714-bp DNA product was gel-purified (Invitrogen) and directly sequenced on both strands. The amplified sequence was assembled in conjunction with corresponding EST sequences (GenBank™ accession numbers AA677272, AI041547, AA887547, and BI464064, respectively) using the Lasergene (DNASTAR Inc., Madison, WI) software suite.

The deduced protein sequences of all members of the HNK-1 family of sulfotransferase were analyzed by multiple alignments using the ClustalW (version 1.4) algorithm (14) implemented in the Bioedit suite.

Construction of pcDNA3.1-C4ST-3—The ORF of C4ST-3 was amplified from pBKS-C4ST-3-ORF341 (see above) by PCR using 1) the 5′-specific primer 5′-tta agc ttg cca cca tgg gga ggc gct gct gcc ggc ggc gcg-3′ containing a *Hin*dIII site, the consensus Kozak sequence GCCACC (15), and a start codon and 2) the 3′-specific primer 5′-gct cta gac tag agc agc cgc agg tag ga-3′ containing a stop codon and an *Xba*I site. The 1026-bp PCR fragment was directionally subcloned into the eukaryotic expression vector pcDNA3.1 (Invitrogen), completely sequenced on both strands, and designated pcDNA3.1-C4ST-3.

Genomic Organization and Chromosomal Localization of C4ST-3 and C4ST-1—The cDNA sequences of C4ST-3 and C4ST-1 (GenBank™ accession number AJ289131; see also RefSeq entry NM_018413) were used to query human genome (Genomic BLAST accessible at the NCBI Web site) sequences using BLASTN in order to elucidate the genomic structure of these human genes. Filtering of repetitive elements (Repeat Masker, version 07/16/00, University of Washington Genome Center) and BLASTN (version 2.1.3) searches against the databases of GenBank™, EMBL, and DDBJ EST divisions (dbEST) at the NCBI were carried out on the identified working draft sequence segments to verify exon sequences and to locate neighboring genes. CpG islands as defined by Gardiner-Garden and Frommer (16) were analyzed by using the WWWCPG program (17).

Northern Blot and Expression Array Analysis—Human Multiple Tissue Northern (MTN®) blots and human multiple tissue expression

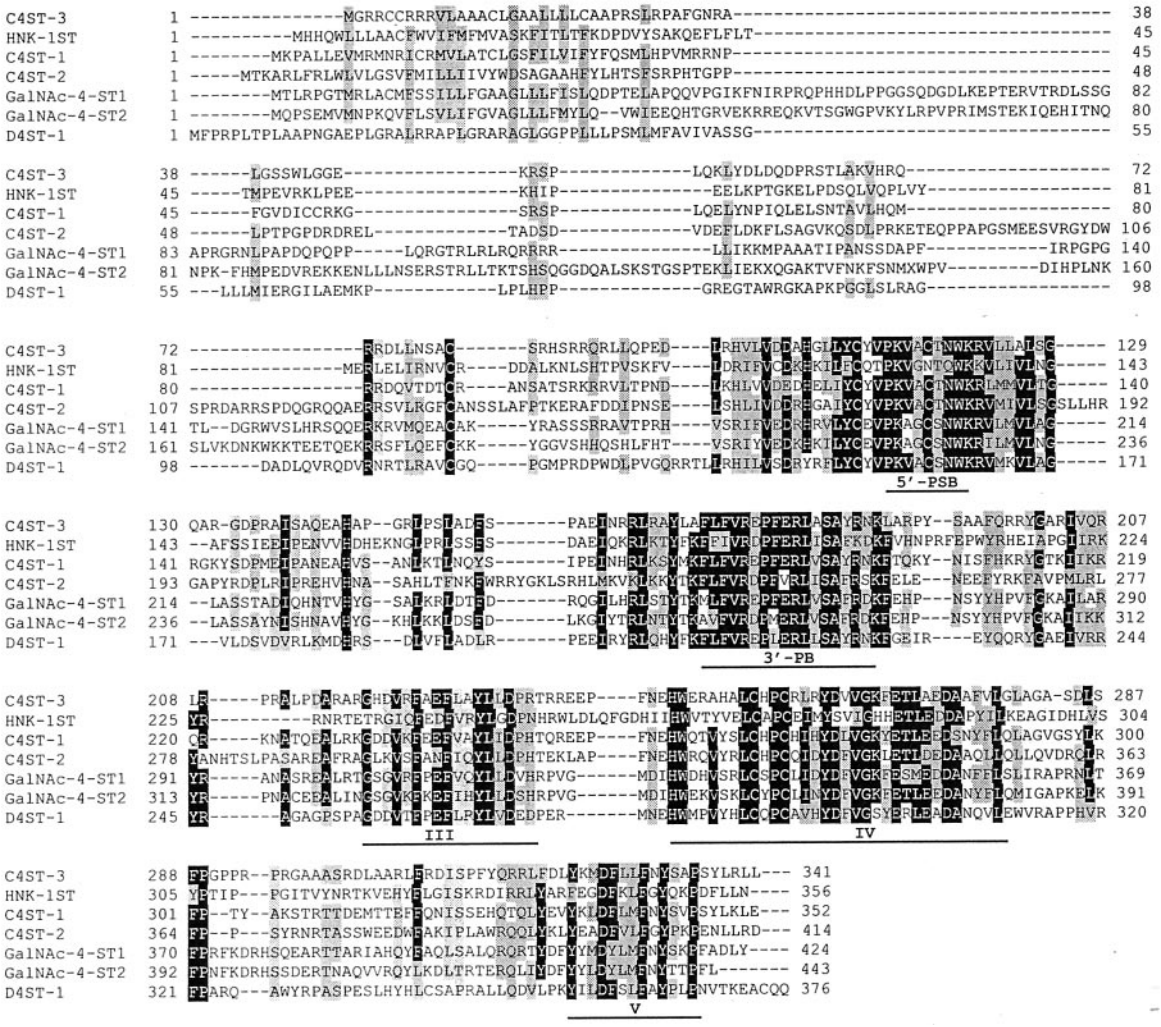


FIG. 2. **Multiple amino acid sequence alignment of members of the human HNK-1 family of sulfotransferases.** Alignment was performed using the ClustalW algorithm. Introduced gaps are shown as *hyphens*, and aligned amino acids are *boxed* (*black* for identical residues and *dark gray* for similar residues). Putative binding sites for the 5′-phosphosulfonate group (*5′-PSB*) and 3′-phosphate group (*3′-PB*) of PAPS and three additional highly conserved domains (III, IV, and V) are indicated.

(MTE™) arrays were purchased from CLONTECH (BD Biosciences-CLONTECH, Heidelberg, Germany). They were hybridized with 5–15 × 10^6 cpm of a specific α-^{32}P-labeled cDNA probe ([α-^{32}P]dCTP and Megaprime™ labeling kit purchased from Amersham Biosciences) and washed according to the manufacturer's specifications. The membranes were exposed to Biomax MS films (Eastman Kodak Co.) for 2–5 days at −80 °C with intensifying screens. The 264-bp probe corresponds to nucleotides 217–480 of the C4ST-3 cDNA (GenBank™ accession number AY120869).

Transient Expression of Human D4ST-1, C4ST-1, and C4ST-3—CHO/Tag cells were transfected with 13 μg of pcDNA3.1-D4ST-1, pcDNA3.1-C4ST-1, pcDNA3.1-C4ST-3, or pcDNA3.1 using 35 μg of LipofectAMINE (Invitrogen) in serum-free medium for 6 h according to the manufacturer's protocol. Sixty hours after transfection with native forms of the sulfotransferases, the cells and medium were collected separately for analysis. Cells were lysed with 200 μl of 20 mM HEPES buffer, pH 7.4, 5 mM $MgCl_2$, 175 mM KCl, 2% Triton X-100, protease inhibitors (23 millitrypsin inhibitor units of aprotinin and 4 μg each of leupeptin, antipain, pepstatin, and chymostatin) per 100-mm diameter culture plate. The homogenate was mixed by rotation for 1 h and sedimented at 12,000 × *g* for 20 min. The supernatant was designated as the cell extract. The culture medium was pooled and sedimented at 12,000 × *g* for 20 min. The culture supernatant was adjusted to a final concentration of 20 mM HEPES, pH 7.4, and protease inhibitors were added as noted above.

The cytosolic and transmembrane domains of D4ST-1, C4ST-1, and C4ST-3 were substituted with the signal sequence of Igκ in the pSec constructs. Furthermore, the *myc* epitope followed by six histidine residues was added at the carboxyl terminus of each of these sulfotransferases in the pSec constructs. Following transfection with pSec-D4ST-1, pSec-C4ST-1, pSec-C4ST-3, or pSec, the culture medium was collected as described above. The culture medium (20 ml) was incubated with 200 μl of Ni^{2+}-NTA-agarose (Qiagen) overnight at 4 °C. The Ni^{2+}-NTA-agarose was washed with 20 mM HEPES, pH 7.4, 200 mM NaCl, 10 mM imidazole. Ni^{2+}-NTA-agarose with bound sulfotransferase was suspended in 800 μl of 10 mM imidazole, pH 6.8, and aliquots were assayed directly for sulfotransferase activity.

Sulfotransferase Activity Analysis—Chondroitin- and dermatan-4-sulfotransferase activities were assayed as described (3, 8). Each reaction (50 μl) contained 50 mM imidazole-HCl, pH 6.8, 2 mM dithiothreitol, 5 × 10^5 cpm [^{35}S]PAPS, 2 μM PAPS, and protamine (0.005% (w/v) for chondroitin or 0.05% (w/v) for dermatan), and protease inhibitors (see solubilization buffer above). Desulfated chondroitin (Seikagaku America, Inc.) and desulfated dermatan from porcine intestine (Sigma) (18) were used as acceptors. Incorporation was determined to be linear for 2–3 h under these conditions. Assays were carried out for 2 h at 28 °C and terminated by boiling for 3 min. The [^{35}S]SO_4-labeled chondroitin and dermatan products were precipitated by the addition of 0.1 volume of 4 M potassium acetate and 3 volumes of ethanol. After sedimentation, the precipitates were dissolved in 140 μl of H_2O and separated from any remaining [^{35}S]PAPS and degradation products by passage over Bio-Spin 6 columns (Bio-Rad). Transfer to GalNAcβ1,4GlcNAcβ1,2Manα-MCO (19) and GlcUAβ1,3Galβ1,4GlcNAcβ1,3Galβ1,4Glcβ1,C_2H_4NHC-OCF_3 was assayed as described (2, 3). When sulfotransferase bound to Ni^{2+}-NTA-agarose was analyzed, the reaction was gently mixed every 30 min.

Product Characterization—[^{35}S]SO_4-chondroitin and [^{35}S]SO_4-dermatan products were digested with 50 milliunits of chondroitinase AC

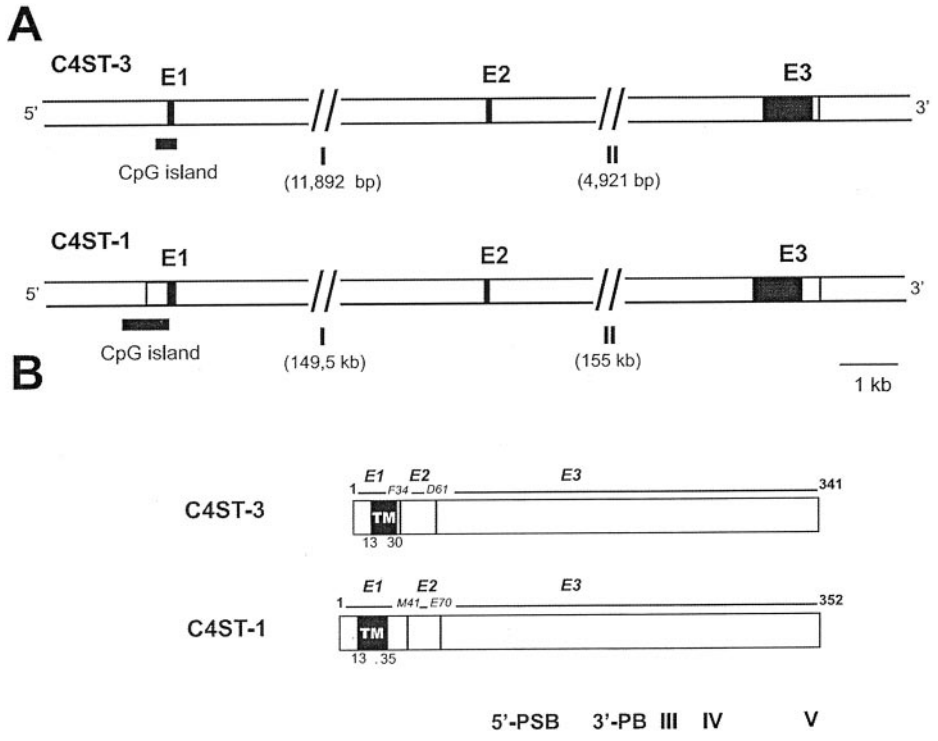


FIG. 3. **Structure of the human C4ST-3 and C4ST-1 genes.** *A*, exonic sequences that contribute to coding regions of C4ST-3 and C4ST-1 are shaded in *black*, and untranslated regions are shaded in *gray*. The size of intronic regions not shown are indicated in parentheses. The locations of CpG islands are indicated by the *solid bars*. *B*, schematic of the aligned protein sequences of C4ST-3 and C4ST-1. The *numbers* and *letters above* each sequence indicate the first amino acid encoded by the particular exons (*E1–E3*). The location of the transmembrane domain (*TM*) and the conserved motifs/regions, 5′-phosphosulfonate binding site (5′-*PSB*), 3′-phosphate binding site (3′-*PB*), III, IV, and V, are also indicated.

TABLE I
Substrate specificities of C4ST-3, C4ST-1, C4ST-2, D4ST-1, GalNAc-4-ST-1, and GalNAc-4-ST2

CHO/Tag cells were transfected with pSec-C4ST-3, pSec-C4ST-1, pSec-C4ST-2, pSec-D4ST-1, pSec-GalNAc-4-ST1, pSec-GalNAc-4-ST2, or the pSec vector. After 60 h in culture, the medium was collected, and secreted sulfotransferases were incubated with Ni^{2+}-NTA-agarose (Qiagen). The beads were washed, and sulfotransferase activity was determined as described under "Experimental Procedures" using fixed aliquots of beads. Results are expressed as pmol incorporated/h/100-mm plate.

Transferase	Chondroitin	Dermatan	GGnMMCO
		pmol/h/plate	
C4ST-3	195.4	46.9	3.5
C4ST-1	1053.9	143.4	3.2
C4ST-2	86.6	126.6	3.8
D4ST-1	89.1	1156.0	3.7
GalNAc-4-ST1	73.3	51.8	1678.5
GalNAc-4-ST2	64.3	51.7	1249.7
Vector	19.8	21.6	7.7

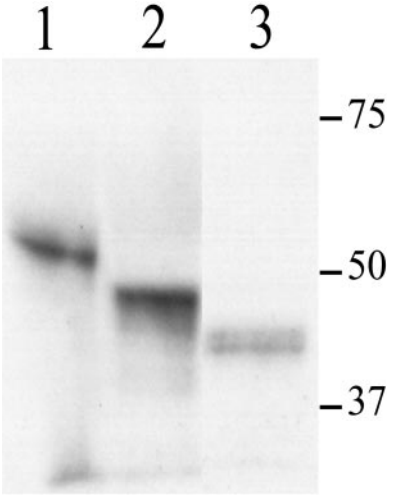


FIG. 4. **Expression of pSec-C4ST-3, pSec-C4ST-1, and pSec-D4ST-1 in CHO cells.** CHO/Tag cells were transfected with pSec-C4ST-3, pSec-C4ST-1, pSec-D4ST-1, or the pSec vector alone. Medium was collected at 60 h following transfection, and the proteins were precipitated with chloroform/methanol (38). The precipitated proteins were subjected to SDS-PAGE in 7.5% acrylamide gels and electrophoretically transferred to polyvinylidene difluoride membranes. Western blot analysis with anti-*myc* antibody was used to estimate the amount of sulfotransferase secreted into the medium. *Lane 1*, pSec-C4ST-1 from 400 μl of medium; *lane 2*, pSec-D4ST-1 from 400 μl of medium; *lane 3*, pSec-C4ST-3 from 400 μl of medium. The location and molecular masses in kDa for standards are indicated to the *right* of the gel.

I (Calbiochem) in 100 mM Tris acetate buffer, pH 7.3, for 16 h at 37 °C or with 25 milliunits of chondroitinase B (Calbiochem) in 100 mM Tris acetate buffer, pH 8.0, for 16 h at 37 °C. The digestion products were analyzed by high pressure liquid chromatography on a 4.6 × 250-mm MicroPak AX-5 column (Varian) using a linear gradient of 10–450 mM KH_2PO_4 over 40 min with a flow rate of 1.0 ml/min (20). Standards were detected by absorbance at 210 or 234 nm, and fractions were collected at 0.5-min intervals for determination of radioactivity. [^{35}S]SO_4-labeled products separated on a Micro Pak AX-5 column were pooled separately and further characterized by gel filtration on a 16 × 500-mm Superdex™ 30 preparative grade gel filtration column (Amersham Biosciences) eluted at 1.0 ml/min in 100 mM NH_4HCO_3. The location of the sulfate on the disaccharide digestion products was confirmed by digestion with chondroitin-4-sulfatase (50 milliunits) and analysis on Micro Pak AX-5 columns as above.

RESULTS

Identification of a Human cDNA Related to HNK-1 ST—The nonredundant data base at the NCBI was probed with the deduced amino acid sequences of human and rat HNK-1 STs (GenBank™ accession number AF033827). A BAC clone RP11–390G14 derived from human chromosome 3 (GenBank™ accession number AC024558) that contained an ORF with a length of 843 bp in the region displaying homology was identified. This sequence was used for further BLASTN searches against dbEST. Retrieval of EST BI908522 that overlapped the 5′-region of the ORF in the BAC clone allowed us to evaluate FGENES-predicted exons on a corresponding working draft sequence segment (GenBank™ accession number NT_005588) of chromosome 3. An ORF with a length of 1026 bp that is encoded by three exons was identified and cloned (Fig. 1*A*). 5′-Rapid amplification

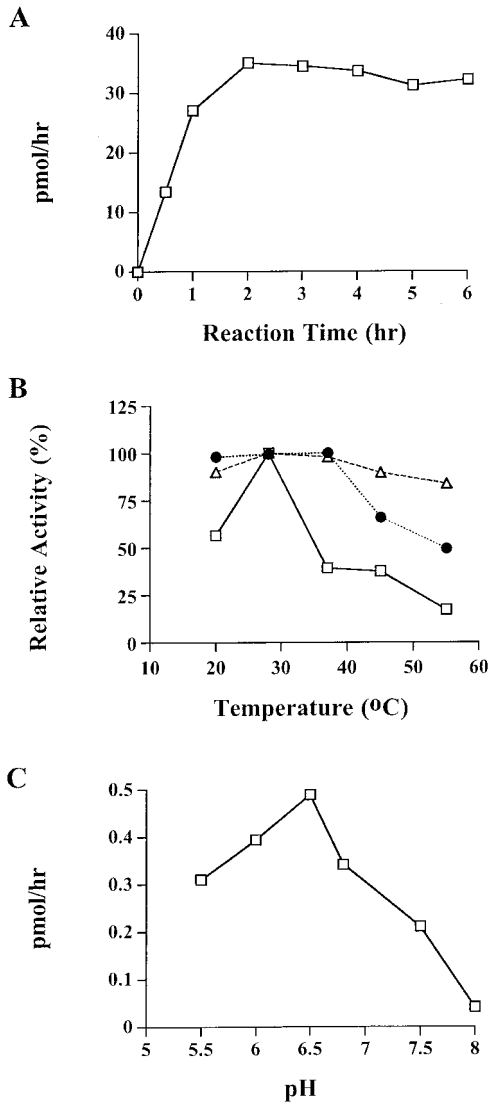


FIG. 5. **Properties of pSec-C4ST-3.** pSec-C4ST-3, pSec-C4ST-1, and pSec-D4ST-1 were immobilized on Ni^{2+}-NTA-agarose. *A*, pSec-C4ST-3 (25-μl suspension containing 5 μl of agarose) was incubated with 50 μg of desulfated chondroitin and 2 μM [^{35}S]PAP in a 50-μl reaction at 28 °C for the times indicated. The amount of [^{35}S]SO_4 incorporated was determined as described under "Experimental Procedures." *B*, the amount of [^{35}S]SO_4 transferred to chondroitin by immobilized pSec-C4ST-1 (■) and pSec-C4ST-3 (□) or dermatan by pSec-D4ST-1 (△) was determined following incubation for 2 h at the temperatures indicated with 50 μg of desulfated chondroitin or desulfated dermatan and 2 μM PAP[^{35}S] in a 50-μl reaction. *C*, the transfer of [^{35}S]SO_4 to chondroitin by pSec-C4ST-3 was monitored using imidazole-HCl at the pH values indicated following incubation for 2 h at 28 °C.

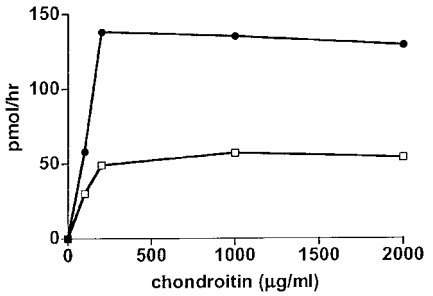


FIG. 6. **Concentration dependence for transfer of sulfate to chondroitin by pSec-C4ST-1 and pSec-C4ST-3.** pSec-C4ST-1 (●) and pSec-C4ST-3 (□) were immobilized on Ni^{2+}-NTA-agarose and incubated with increasing concentrations (0–2000 μg/ml) of desulfated chondroitin for 2 h at 37 °C, and the amount of [^{35}S]SO_4 incorporated was determined. The amount of [^{35}S]SO_4 incorporated by an equal amount of medium from cells transfected with the pSec vector has been subtracted.

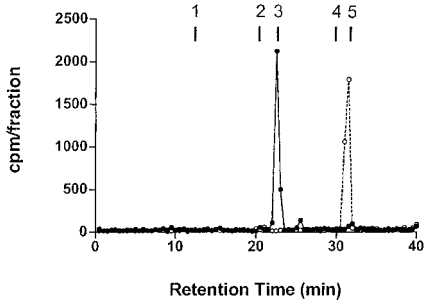
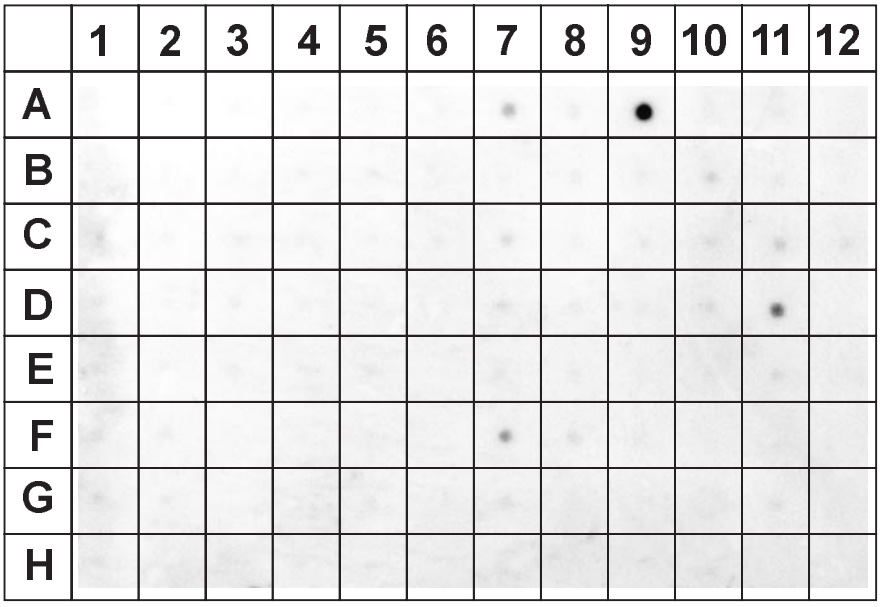


FIG. 7. **Characterization of the [^{35}S]SO_4-chondroitin and [^{35}S]SO_4 product produced by pSec-C4ST-3.** Desulfated chondroitin (50 μg) was incubated with immobilized pSec-C4ST-3 (25-μl suspension containing 5 μl of agarose) in a 50-μl reaction containing 2 μM [^{35}S]PAPS for 2 h at 28 °C. The [^{35}S]SO_4-labeled products were isolated by gel filtration and digested with 50 milliunits of chondroitinase AC I in 100 mM Tris acetate buffer, pH 7.3, for 16 h at 37 °C. The digestion product was analyzed on a MicroPak AX-5 column (Varian) developed with a linear gradient of 10–450 mM KH_2PO_4 over 40 min at a flow rate of 1.0 ml/min. (●). The product of the chondroitinase AC I digestion was further digested with chondroitin-4-sulfatase and analyzed on the MicroPak AX-5 column under identical conditions (○). The elution positions of standards are as follows: GalNAc-4-SO_4 (*1*), D-glucono-6-enepyranoside β-1,3-GalNAc-4-SO_4 (*2*), D-glucono-4-enepyranoside β-1,3-GalNAc-4-SO_4 (*3*), D-gluco-4-enepyranoside β1,3 GalNAc-4,6-diSO_4 (*4*), and SO_4, free sulfate (*5*).

of cDNA ends and amplification of the 3′-UTR were carried out to obtain the full-length sequence of the cDNA but only succeeded in elongation of the 3′-end (see "Experimental Procedures"). Fig. 1*A* shows the 1747-bp C4ST-3 cDNA including the 1026-bp ORF that encodes a protein of 341 amino acid residues with a single potential *N*-glycosylation site and a calculated molecular mass of 39 kDa. The deduced protein, designated C4ST-3 (GenBank™ accession number AY120869) is a type II transmembrane protein with a 12-amino acid cytosolic domain at the amino terminus (see Kyte-Doolittle hydrophobicity profile in Fig. 1*B*).

Multiple alignment of the protein sequence of C4ST-3 with other members of the HNK-1 sulfotransferase family was performed using the ClustalW algorithm as implemented in the BioEdit software suite (Fig. 2). The alignment indicates that C4ST-3 is 45% identical to C4ST-1, 27% identical to C4ST-2, 26% identical to GalNAc-4-ST1, 23% identical to GalNAc-4-ST2, and 26% identical to HNK-1 ST (all of the protein sequences shown in Fig. 2 are of human origin). The regions with the highest degree of identity are the putative 5′-phosphosulfate binding site (*5′-PSB*), the putative 3′-phosphate binding site (*3′-PB*), and three regions of unknown function designated III, IV, and V that are carboxyl-terminal to the 3′-phosphate binding site (Fig. 2). Identical and similar amino acids are *shaded* if they occur at a specific position in at least five of the seven sequences shown in the multiple alignment in Fig. 2.

Genomic Organization and Chromosome Localization of C4ST-1—A BLAST analysis of the genomic sequence available through the NCBI Web site using C4ST-3 cDNA as a query sequence mapped the C4ST-3 gene to human chromosome 3q21.3. The ORF and the 3′-UTR of C4ST-3 are encoded by three exons (Fig. 3). The genomic sequence was examined for the presence of CpG islands as defined by Gardiner-Garden and Frommer (16). A CpG island was identified extending from 620 bp upstream to 880 bp downstream of the C4ST-3 start codon. Such CpG islands have been detected in 82% of analyzed genes that show widespread expression and are indicative of the presence of a promoter region.



	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G												
H												

A1 brain, whole	B1 cerebral cortex	C1 frontal lobe	D1 parietal lobe
A2 cerebellum, left	B2 cerebellum, right	C2 corpus callosum	D2 amygdala
A3 substantia nigra	B3 accumbens nucleus	C3 thalamus	D3 pituitary gland
A4 heart	B4 aorta	C4 atrium, left	D4 atrium, right
A5 esophagus	B5 stomach	C5 duodenum	D5 jejunum
A6 colon, transverse	B6 colon, descending	C6 rectum	D6 empty
A7 kidney	B7 skeletal muscle	C7 spleen	D7 thymus
A8 lung	B8 placenta	C8 bladder	D8 uterus
A9 liver	B9 pancreas	C9 adrenal gland	D9 thyroid gland
A10 leukemia, HL-60	B10 HELA S3	C10 leukemia, K-562	D10 leukemia, MOLT-4
A11 brain, fetal	B11 heart, fetal	C11 kidney, fetal	D11 liver, fetal
A12 yeast total RNA	B12 yeast tRNA	C12 E.coli rRNA	D12 E. coli DNA
E1 occipital lobe	F1 temporal lobe	G1 p.g. of cerebral cortex*	H1 pons
E2 caudate nucleus	F2 hippocampus	G2 medulla oblongata	H2 putamen
E3 spinal cord	F3 empty	G3 empty	H3 empty
E4 ventricle, left	F4 ventricle, right	G4 interventricular space	H4 apex of the heart
E5 ileum	F5 ilocecum	G5 appendix	H5 colon, ascending
E6 empty	F6 empty	G6 empty	H6 empty
E7 periph. blood leuk.**	F7 lymph node	G7 bone marrow	H7 trachea
E8 prostate	F8 testis	G8 ovary	H8 empty
E9 salivary gland	F9 mammary gland	G9 empty	H9 empty
E10 Burkitt`s ***	F10 Burkitt`s ****	G10 SW480 *****	H10 lung carcinoma, A549
E11 spleen, fetal	F11 thymus, fetal	G11 lung, fetal	H11 empty
E12 Poly r(A)	F12 human C_0t-1 DNA	G12 human DNA, 100 ng	H12 human DNA, 500 ng

FIG. 8. **RNA dot blot analysis of C4ST-3 transcripts.** The human multiple tissue expression (MTE™) array shown was hybridized with a ^{32}P-labeled human C4ST-3-specific cDNA probe. Tissue sources for the RNA are indicated *below* the blot. C4ST-3 expression is detected in adult liver (A9) and at significantly lower levels in adult kidney (A7), lymph nodes (F7), and fetal liver (D11). *, paracentral gyrus of cerebral cortex; **, peripheral blood leukocytes; ***, Burkitt's lymphoma Raji; ****, Burkitt's lymphoma Daudi; *****, colorectal adenocarcinoma, SW280.

C4ST-3 Is a Chondroitin-specific GalNAc-4-O-sulfotransferase—Whereas C4ST-3 displayed the highest percentage of identical amino acids, 45.1%, when compared with C4ST-1, it also displayed significant homology with other members of the HNK-1 family of sulfotransferases. Initial experiments using C4ST-3 expressed by CHO/Tag cells revealed detectable levels of an activity in both cell extracts and medium able to transfer sulfate from [^{35}S]PAPS to chondroitin (not shown). Due to the low levels of activity, a secreted form of C4ST-3 was prepared by substituting the cytosolic and transmembrane domains of C4ST-3 with the signal sequence of human IgGκ (Invitrogen) to produce pSec-C4ST-3. The identical constructs were prepared for C4ST-1, C4ST-2, D4ST-1, HNK-1 ST, GalNAc-4-ST1, and GalNAc-4-ST2 and designated pSec-C4ST-1, pSec-C4ST-2, pSec-D4ST-1, pSec-GalNAc-4-ST1, and pSec-GalNAc-4-ST2. The *myc* epitope followed by six histidines was located at the carboxyl terminus of each of these constructs. Following transfection into CHO/Tag cells, the secreted transferases were allowed to bind to Ni^{2+}-NTA-agarose and assayed for transfer of sulfate to chondroitin, dermatan, and GGnM-MCO while bound to the agarose beads as summarized in Table I.

pSec-C4ST-3 transferred sulfate to chondroitin (195 pmol/h/plate) but not to either dermatan or GGnM-MCO (Table I). Likewise, pSec-C4ST-1 was highly active with chondroitin but not dermatan or GGnM-MCO. pSec-D4ST-1 and pSec-GalNAc-4-ST1 and -ST2 were active with dermatan and GGnM-MCO, respectively. Whereas the transfer of sulfate to chondroitin was 10-fold higher than seen with mock-transfected cells for pSec-C4ST-3 (195 *versus* 20 pmol/h/plate), the rate of transfer was one-tenth of that seen for pSec-C4ST-1. Since this may reflect differences in the level of expression, the relative amounts of pSec-C4ST-1 and pSec-C4ST3 were estimated by Western blot analysis using an anti-*myc* antibody following SDS-PAGE and electrophoretic transfer to polyvinylidene difluoride (Fig. 4). The level of expression was significantly lower for pSec-C4ST-3 than for either pSec-C4ST-1 or pSec-D4ST-1, accounting for much of the apparent lower level of activity.

In addition to lower levels of expression, pSec-C4ST-3 was found to be labile at temperatures above 28 °C (Fig. 5*B*). Whereas pSec-C4ST-1 and pSec-D4ST-1 displayed similar levels of activity at 37 and 28 °C, the transfer of sulfate by pSec-C4ST-3 at 37 °C was 35% of that seen at 28 °C (Fig. 5*B*).

As with pSec-C4ST-3, the transfer of sulfate by pSec-C4ST-1 was reduced at 45 and 55 °C, whereas the transfer of sulfate to dermatan by D4ST-1 was not markedly reduced at these temperatures. When transferase reactions were carried out at 28 °C, incorporation of sulfate by C4ST-3 into chondroitin remained linear for up to 2 h under the assay conditions utilized (Fig. 5*A*). As with C4ST-1 and pSec-C4ST-1, the incorporation of [^{35}S]SO_4 into chondroitin by pSec-C4ST-3 is saturated at an acceptor concentration of 200 μg/ml (Fig. 6). pSec-C4ST-3 has a pH optimum of 6.5 for transfer of sulfate to chondroitin (Fig. 5*C*).

Digestion of the ^{35}S-sulfated chondroitin product with chondroitinase AC I yielded a single peak that comigrated with D-glucono-4-enepyranoside β-1,3-GalNAc-4-SO_4 (Fig. 7). No product was obtained upon digestion with chondroitinase B (not shown). The location of the sulfate on the C-4 hydroxyl of the GalNAc was confirmed by the release of free sulfate upon digestion of this product with chondroitin-4-sulfatase (Fig. 7). Thus, C4ST-3 is a GalNAc-4-sulfotransferase that is specific for chondroitin sequences (*i.e.* GlcUAβ1,4GalNAcβ1,3).

Tissue Expression Pattern of C4ST-3—Macroarray and Northern blot analyses were used to determine the expression pattern for C4ST-3 in human tissues. As evident by MTE™ macroarray analysis (Fig. 8), C4ST-3 transcripts are most highly represented in adult liver (*A9*). Significantly lower expression levels were detected in adult kidney (*A7*), lymph nodes (*F7*), and fetal liver (*D11*). By Northern blot analysis (not shown), a single 2.1-kb transcript was detected in adult liver (Fig. 8). The level of expression was not sufficient to detect a transcript in the kidney upon Northern blot analysis.

DISCUSSION

C4ST-3 represents the seventh member of the HNK-1 family of sulfotransferases. C4ST-3 has the highest percentage of identical amino acids when aligned with C4ST-1, 45.1% identical. Like C4ST-1, C4ST-3 transfers sulfate to the C-4 hydroxyl of GalNAc substituted at C-3 with β-linked GlcUA (*i.e.* GlcUAβ,3-GalNAc). Neither sulfotransferase transfers sulfate to the C-4 hydroxyl of GalNAc substituted at C-3 with α-linked IdoUA (*i.e.* dermatan) or to terminal GalNAc in the sequence GalNAcβ1,4GlcNAcβ1,2Man found at the terminus of certain *N*-linked oligosaccharides. Based on the levels of mRNA expression, C4ST-1 is widely expressed in tissues (8, 9), whereas C4ST-3 has a restricted pattern of expression. Since C4ST-3 and C4ST-1 have similar, if not identical, specificities and are both expressed in liver, kidney, and lymph nodes, the biological role played by C4ST-3 is not clear. The lability of C4ST-3 at 37 °C as compared with C4ST-1 and the lower levels of expression seen following transfection as compared with C4ST-1 suggest that C4ST-3 may have a different and perhaps highly specialized role *in vivo*.

With the exception of HNK-1 ST itself, the seven members of the HNK-1 family of sulfotransferases have all proved to be GalNAc-4-*O*-sulfotransferases. The GalNAc-sulfotransferases add sulfate to the C4 hydroxyl of either terminal β1,4-linked GalNAc (GalNAc-4-ST1 (3, 5) and GalNAc-4-ST2 (6)) or to β1,4-linked GalNAc that is substituted at its C-3 hydroxyl with either GlcUA (C4ST-1 (8, 9) and C4ST-3) or IdoUA (D4ST-1 (7)). In contrast, HNK-1 ST adds sulfate to the C-3 hydroxyl of terminal β1,3-linked GlcUA (1, 2). Each member of this family of sulfotransferases is thus highly specific, and the unique sulfated structures that are produced have distinct biological roles. The HNK-1 structure is involved in neural recognition and synaptogenesis in the central and peripheral nervous system (21–23). Oligosaccharides terminating with SO_4-4-GalNAcβ1,4GlcNAcβ1,2Man are recognized by a receptor, the Man/GalNAc-4-SO_4 receptor (24–26), that regulates the half-life of glycoproteins such as the glycoprotein hormone lutropin (27, 28). This recognition is essential for attaining maximal biologic activity *in vivo* (29). The highly regulated expression of GalNAc-4-ST1 (3, 30) in other tissues such as the brain suggests that it will have additional roles that remain to be defined. In addition to their importance for the formation and maintenance of cartilage (31, 32), there is also evidence that chondroitin sulfate proteoglycans are important for neural cell adhesion, neurite outgrowth, synaptic plasticity, and regeneration (33–37). A number of chondroitin sulfate-bearing proteoglycans are produced by tissues (31, 32). Since the functions of these proteoglycans are modulated and/or require the addition of sulfate, the existence of multiple chondroitin-4-GalNAc sulfotransferases provides the potential for additional regulation and specificity. The availability of the cloned sulfotransferases provides tools to investigate the biological roles of these complex sulfated structures.

REFERENCES

1. Ong, E., Yeh, J. C., Ding, Y., Hindsgaul, O., and Fukuda, M. (1998) *J. Biol. Chem.* **273,** 5190–5195
2. Bakker, H., Friedmann, I., Oka, S., Kawasaki, T., Nifant'ev, N., Schachner, M., and Mantei, N. (1997) *J. Biol. Chem.* **272,** 29942–29946
3. Xia, G., Evers, M. R., Kang, H.-G., Schachner, M., and Baenziger, J. U. (2000) *J. Biol. Chem.* **275,** 38402–38409
4. Okuda, T., Mita, S., Yamauchi, S., Fukuta, M., Nakano, H., Sawada, T., and Habuchi, O. (2000) *J. Biol. Chem.* **275,** 40605–40613
5. Hiraoka, N., Misra, A., Belot, F., Hindsgaul, O., and Fukuda, M. (2001) *Glycobiology* **11,** 495–504
6. Kang, H. G., Evers, M. R., Xia, G., Baenziger, J. U., and Schachner, M. (2001) *J. Biol. Chem.* **276,** 10861–10869
7. Evers, M. R., Xia, G., Kang, H. G., Schachner, M., and Baenziger, J. U. (2001) *J. Biol. Chem.* **276,** 36344–36353
8. Hiraoka, N., Nakagawa, H., Ong, E., Akama, T. O., Fukuda, M. N., and Fukuda, M. (2000) *J. Biol. Chem.* **275,** 20188–20196
9. Yamauchi, S., Mita, S., Matsubara, T., Fukuta, M., Habuchi, H., Kimata, K., and Habuchi, O. (2000) *J. Biol. Chem.* **275,** 8975–8981
10. Manzella, S. M., Hooper, L. V., and Baenziger, J. U. (1996) *J. Biol. Chem.* **271,** 12117–12120
11. Baenziger, J. U., and Green, E. D. (1991) in *Biology of Carbohydrates*, Vol. 3 (Ginsberg, V., and Robbins, P. W., eds) pp. 1–46, JAI Press Ltd., London
12. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
13. Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996) *Genomics* **33,** 151–152
14. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
15. Kozak, M. (1978) *Cell* **15,** 1109–1123
16. Gardiner-Garden, M., and Frommer, M. (1987) *J. Mol. Biol.* **196,** 261–282
17. Milanesi, L., D'Angelo, D., and Rogozin, I. B. (1999) *Bioinformatics* **15,** 612–621
18. Cole, L. A., Perini, F., Birken, S., and Ruddon, R. W. (1984) *Biochem. Biophys. Res. Commun.* **122,** 1260–1267
19. Skelton, T. P., Hooper, L. V., Srivastava, V., Hindsgaul, O., and Baenziger, J. U. (1991) *J. Biol. Chem.* **266,** 17142–17150
20. Baenziger, J. U. (1994) *Methods Enzymol.* **230,** 237–249
21. Kruse, J., Mailhammer, R., Wernecke, H., Faissner, A., Sommer, I., Goridis, C., and Schachner, M. (1984) *Nature* **311,** 153–155
22. Schachner, M., and Martini, R. (1995) *Trends Neurosci.* **18,** 183–191
23. Jungalwala, F. B. (1994) *Neurochem. Res.* **19,** 945–957
24. Fiete, D., and Baenziger, J. U. (1997) *J. Biol. Chem.* **272,** 14629–14637
25. Fiete, D., Beranek, M. C., and Baenziger, J. U. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 11256–11261
26. Fiete, D. J., Beranek, M. C., and Baenziger, J. U. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 2089–2093
27. Baenziger, J. U., Kumar, S., Brodbeck, R. M., Smith, P. L., and Beranek, M. C. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 334–338
28. Fiete, D., Srivastava, V., Hindsgaul, O., and Baenziger, J. U. (1991) *Cell* **67,** 1103–1110
29. Mi, Y.-L., Shapiro, S. D., and Baenziger, J. U. (2002) *J. Clin. Invest.* **109,** 269–276
30. Dharmesh, S. M., Skelton, T. P., and Baenziger, J. U. (1993) *J. Biol. Chem.* **268,** 17096–17102
31. Kjellén, L., and Lindahl, U. (1991) *Annu. Rev. Biochem.* **60,** 443–475
32. Iozzo, R. V. (1998) *Annu. Rev. Biochem.* **67,** 609–652
33. Faissner, A., Clement, A., Lochter, A., Streit, A., Mandl, C., and Schachner, M. (1994) *J. Cell Biol.* **126,** 783–799
34. Bukalo, O., Schachner, M., and Dityatev, A. (2001) *Neuroscience* **104,** 359–369
35. Lemons, M. L., Howland, D. R., and Anderson, D. K. (1999) *Exp. Neurol.* **160,** 51–65
36. Bradbury, E. J., Moon, L. D., Popat, R. J., King, V. R., Bennett, G. S., Patel, P. N., Fawcett, J. W., and McMahon, S. B. (2002) *Nature* **416,** 636–640
37. Ruoslahti, E. (1996) *Glycobiology* **6,** 489–492
38. Wessel, D., and Flugge, U. I. (1984) *Anal. Biochem.* **138,** 141–143