J. Biochem. 131, 337–347 (2002)

Molecular Cloning and Genomic Analysis of Mouse Glucuronyltransferase Involved in Biosynthesis of the HNK-1 Epitope[1]

Shoji Yamamoto,* Shogo Oka,* Fumiko Saito-Ohara,† Johji Inazawa,† and Toshisuke Kawasaki*,[2]

**Department of Biological Chemistry and CREST (Core Research for Educational Science and Technology) Project, Japan Science and Technology Corporation, Graduate School of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8501; and †Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8591*

Received November 7, 2001; accepted December 4, 2001

cDNA and genomic clones encoding the mouse glucuronyltransferase (GlcAT-P) involved in biosynthesis of the HNK-1 carbohydrate epitope were isolated and the structural organization of the gene was determined. The predicted amino acid sequence of mouse GlcAT-P is 96.2 and 98.2% identical to those of the rat and human enzymes, respectively. Alternatively spliced isoforms of mouse GlcAT-P are present in the brain and encode two proteins that are identical throughout their length except for an additional 13 amino acids in the N-terminal cytoplasmic domain of the major form. The coding region of GlcAT-P is composed of 5 exons spanning approximately 6 kb, and the GlcAT-P gene was mapped to the A4 region of mouse chromosome 9. Upstream of the transcriptional start site, no typical TATA or CCAAT box was found, but binding sites for several known transcription factors including Sp1 and Krox-20 were identified. Transient transfection of luciferase reporter constructs demonstrated that a 207 bp fragment of the 5′-upstream region acts as a strong promoter in PC-12 cells, which express the HNK-1 epitope, but not in COS-1 cells. Thus, this minimal promoter region of GlcAT-P is suggested to be associated with the regulation of HNK-1 expression.

Key words: chromosomal mapping, genomic organization, glucuronyltransferase, HNK-1 carbohydrate epitope, molecular cloning.

The HNK-1 carbohydrate epitope, which is recognized by monoclonal antibody HNK-1, is characterized by its terminal 3′-sulfated glucuronic acid (*1*). The HNK-1 epitope is expressed on a series of cell adhesion molecules belonging to the immunoglobulin (Ig) superfamily, such as neural cell adhesion molecule (NCAM), myelin-associated glycoprotein (MAG), P0, L1, TAG-1, and telencephalin (*2–5*), as well as integrins, proteoglycans, and extracellular matrix glycoprotein tenascin-R (*6–9*). The HNK-1 epitope is also expressed on some glycolipids such as sulfoglucuronylglycolipids (SGGL)-1 and -2 (*10, 11*). The expression of the HNK-1 epitope is spatially and temporally regulated in the nervous system (*12, 13*). Strong expression of the HNK-1 epitope is observed in migrating neural crest cells, myogenic cells, rhombomeres, the cerebellum, and myelinating Schwann cells of peripheral nerves (*14–18*). This carbohydrate epitope is presumed to be associated with cell adhesion, migration, neurite extension and synaptic plasticity by modifying the functional properties of cell adhesion molecules (*19–22*). The structure of the HNK-1 epitope was demonstrated to be that of a sulfated trisaccharide, HSO_3-3GlcAβ1-3Galβ1-4GlcNAc, which is shared with the glycolipid and glycoprotein epitopes (*10, 11, 23*). Because the inner structure Galβ1-4GlcNAc is found commonly on various glycoproteins and glycolipids, glucuronyltransferase is considered to be a key enzyme in the biosynthesis of the HNK-1 epitope. However, the mechanisms underlying the regulation of HNK-1 epitope biosynthesis and the characteristic HNK-1 expression are poorly understood.

Recently, we purified a glucuronyltransferase (GlcAT-P) that is associated with biosynthesis of the HNK-1 epitope from rat brain (*24*), and cloned a full-length cDNA encoding GlcAT-P based on partial amino acid sequences (*25*). Using the GlcAT-P cDNA as a probe, we and others have cloned a second HNK-1 glucuronyltransferase (GlcAT-S) (*26, 27*). Furthermore, we also succeeded in cloning of the human GlcAT-P, which was mapped to human chromosome 11q25 (*28*).

Even though both GlcAT-P and GlcAT-S were specifically expressed in neural tissues, the distributions of these two enzymes were different in rat embryo and brain (*27*). The expression of GlcAT-P was detected more abundantly and more widely than that of GlcAT-S (*24, 26*), suggesting that

[1] This work was supported in part by a Grant-in-Aid for Scientific Research (B-12470497), and a Grant-in-Aid for Scientific Research (C-11680604) from the Japan Society for the Promotion of Sciences, and a Grant-in-Aid for Scientific Research on Priority Areas (A-12033204) from the Ministry of Education, Science, Sports and Culture of Japan. Shoji Yamamoto is the recipient of a Research Fellowship of the Japan Society for the Promotion of Science for Young Scientists. The nucleotide sequence data reported in this artide have been submitted to the DDBJ/EMBL/GenBank Data libraries under accession numbers AB055781 for the cDNA and AB055782 for the 5′-flanking region.

[2] To whom correspondence should be addressed. Phone: +81-75-753-4572, Fax: +81-75-753-4605, E-mail: kawasaki@pharm.kyoto-u.ac.jp

Abbreviation: GlcAT, glucuronyltransferase.

GlcAT-P is a major GlcAT involved in biosynthesis of the HNK-1 epitope. Furthermore, the acceptor specificities were significantly different in spite of the structural similarity of the two enzymes. These observations suggested that the HNK-1 epitopes synthesized by the two enzymes have distinct functions. To elucidate the biological function of the HNK-1 epitope and the molecular mechanism regulating its expression, we isolated cDNA and genomic clones of mouse GlcAT-P and characterized the mouse GlcAT-P gene, mapped it to mouse chromosome 9, and identified its promoter region. We also demonstrated the presence of two types of GlcAT-P mRNA produced through alternative splicing.

EXPERIMENTAL PROCEDURES

Isolation of Mouse GlcAT-P cDNA—A λgt10 adult BALB/c mouse brain cDNA library (CLONTECH, 5×10^5 phages) was screened with a ^{32}P-labeled rat GlcAT-P cDNA probe (a 410 bp PCR fragment corresponding to nucleotide numbers 465 to 874 of rat GlcAT-P cDNA) (*25*). Plaques were transferred to nitrocellulose filters and hybridized at 42°C in 5× SSC, 50 mM NaH_2PO_4, pH 7.0, 0.5% non-fat dried milk, 0.1% SDS, 100 μg/ml yeast RNA, and 50% formamide overnight, and then washed at 65°C in 0.2× SSC and 0.1% SDS. Positive plaques were purified, and phage DNA was isolated using standard procedures (*29*). DNA sequences were determined by the dideoxy chain termination method with a model 373 DNA sequencer (Applied Biosystems). The 5′-untranslated region of mouse GlcAT-P cDNA was directly amplified and determined by means of the cloning strategy of RACE-PCR using 10 μg of total RNA from 6-day-old BALB/c mouse brain as a template. The amplified products were subcloned into pCR2.1 (Invitrogen) and then sequenced.

Northern Blot Analysis—Total RNA was extracted from 6-day-old BALB/c mouse brain by the acidic guanidium thiocyanate/phenol/chloroform method. Ten micrograms of total RNA was separated in a 1.2% agarose-formaldehyde gel and then blotted onto a nylon membrane (Hibond N$^+$, Amersham Pharmacia Biotech) with RNA Markers (Promega). The blot was hybridized overnight with a ^{32}P-labeled full-length mouse GlcAT-P cDNA in 0.5 M NaH_2PO_4, pH 7.2, containing 7% SDS, 1 mM EDTA, and 1% BSA at 65°C (*30*). Then the blot was washed in 0.1× SSC and 0.1% SDS at 65°C. The radioactivity was visualized with an image analyzer (Fuji Photo Film, BAS2000).

Isolation of Mouse GlcAT-P Genomic Clones—A λfixII 129/Sv mouse genomic library (Stratagene) was screened with a ^{32}P-labeled full-length mouse GlcAT-P cDNA as a probe. About 2 million phages were plated and transferred to nitrocellulose filters. Hybridization and washes were performed by the same procedures as those for cDNA cloning. Then positive clones were purified, and DNA was prepared according to the standard procedure (*29*) and used for sequencing. None of the clones from the first library screening contained the first exon or the promoter region of the GlcAT-P gene. For the second screening, we used a 157 bp PCR product generated with a primer pair derived from the sequence of exon 1 (indicated by dotted arrows in Fig. 7) as a probe. About 2 million phages were plated, transferred to nitrocellulose filters, and screened. Four positive clones were isolated, which contained the 5′-flanking region of the GlcAT-P gene.

Genomic Southern Blot Analysis—DNA was prepared from mouse 129/Sv liver using standard procedures (*29*). Genomic DNA samples (5 μg each) were digested with restriction enzymes, electrophoresed through 0.6% agarose gels, and then transferred to a nylon membrane (GeneScreen Plus, DuPont NEN). The 444 bp *Sca*I–*Sac*I fragment of mouse GlcAT-P cDNA was labeled with [α^{32}P]dCTP using a random primer labeling kit (TaKaRa). The labeled fragment was hybridized with a membrane in 1 M NaCl containing 10% dextran sulfate, 1% SDS, and 100 μg/ml sheared salmon sperm DNA at 65°C. The membrane was washed in 0.2× SSC, and 0.1% SDS at 65°C, and then analyzed with a BAS 2000 image analyzer.

Chromosomal Mapping—Mouse chromosomes were prepared from skin fibroblast cultures that had been treated with thymidine (300 μg/ml) for 14–16 h, washed, and recultured for 4–5 h in the presence of BrdU (30 μg/ml), as described previously (*31*). For FISH analysis, the full-length mouse GlcAT-P cDNA was labeled with biotin-16-dUTP (Roche Diagnostics, Indianapolis, IN) by nick translation. The chromosomal DNA was hybridized with the probe in hybridization buffer (2× SSC, 50% formamide, 20% dextran sulfate, and 4 mg/ml of BSA). A signal amplification procedure (*31*) was applied for efficient detection of probe signals, and the chromosomes were counterstained with DAPI (4,6-diamino-2-phenylindole, Sigma, St. Louis, MO). The FITC (fluorescein isothiocyanate) signals of probes and the chromosome banding patterns were photographed with a CytoVision system (Applied Imaging, Santa Clara, CA).

Primer Extension Analysis—A fluorescently labeled oligonucleotide, 5′-FITC-TCTCTCTCGAGGCGTTCTCCGGTTCCG-3′, complementary to the sequence from position +331 to +355 of the first exon of the mouse GlcAT-P gene, was synthesized. The mouse brain poly(A)$^+$ RNA was prepared from 6-day-old mouse BALB/c brain total RNA using Oligo(dT)-Latex (Roche Diagnostics). A mixture of 300 ng of oligonucleotide with 5 μg of poly(A)$^+$ RNA was subjected to ethanol-precipitation. The DNA-RNA mixture was then redissolved in 30 μl of hybridization buffer (40 mM PIPES, pH 7.6, 1 mM EDTA, 0.4 M NaCl, and 80% formamide), denatured at 85°C for 10 min, and annealed at 30°C overnight. The annealed hybridization mixture was then ethanol-precipitated, washed, and redissolved in 20 μl of reverse transcriptase buffer (75 mM KCl, 50 mM Tris, pH 8.3, 20 mM dithiothreitol, 3 mM $MgCl_2$, 0.5 mM dNTPs, and 1 unit/μl RNase inhibitor). Subsequently, 200 units of Superscript II reverse transcriptase (Life Technologies) was added, and the reaction mixtures were incubated for 1 h at 42°C. The primer extension product was ethanol-precipitated, washed, and dissolved in 3 μl of loading buffer. The sample was heated at 95°C for 5 min and then separated on a 4.2% polyacrylamide sequencing gel along with the dideoxy chain termination sequencing reaction of the 5′-flanking region of the mouse GlcAT-P gene using the same fluorescently labeled oligonucleotide as a primer. Gel images were obtained and analyzed with an automated DNA sequencer, DSQ-1000L (SHIMADZU).

Reporter Constructs, Transfections, and Luciferase Assays—To create a –1598/+347 GlcAT-P promoter–luciferase construct, the λ DNA of a genomic clone (g72) containing the 1,598 bp 5′-flanking region and exon 1 was digested

with *Sac*I and *Xho*I. Then the restriction fragment was subcloned into pPicaGene-Basic II (pPGBII; Toyo Ink, Tokyo), which contains a luciferase gene in a promoterless background. To obtain various lengths of the 5′-flanking region and exon 1 (–905/+347, –207/+347, –21/+347, +30/+347, and +257/+347), PCR was performed with LA Taq (TaKaRa) using the λ DNA of g72 as a template. The amplification products were subcloned into pPGBII and then sequenced. As a control, plasmid pPGBII was transfected, pPGBII being inactive as to the expression of luciferase activity in the cell lines tested. The luciferase activity due to each luciferase reporter plasmid was normalized as to the *Renilla* luciferase activity of a cotransfected internal control plasmid, pRL-TK (Toyo Ink), containing the herpes simplex virus thymidine kinase promoter upstream of the cDNA encoding *Renilla* luciferase.

PC12 cells were cultured in RPMI 1640 containing 15% fetal bovine serum and 2 mM L-gluthamine, and COS-1 cells were cultured in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum and 2 mM L-gluthamine. Cells were seeded into 6-well plates at a density of approximately 5×10^4 cells per well 24 h prior to transfection.

Each luciferase reporter plasmid (1 μg) and pRL-TK (0.2 μg) were co-transfected into the cells with LipofectAMINE (Life Technologies) following the manufacturer's protocol. After 48 h of transfection, the cells were washed three times with phosphate-buffered saline and then lysed with cell lysing buffer (PGC-50; Toyo Ink). Luciferase activity was measured according to the manufacturer's instructions using a PicaGene Dual Assay System (Toyo Ink) with a luminometer, Lumat LB9501 (Berthold, Bad Weinsberg, Germany).

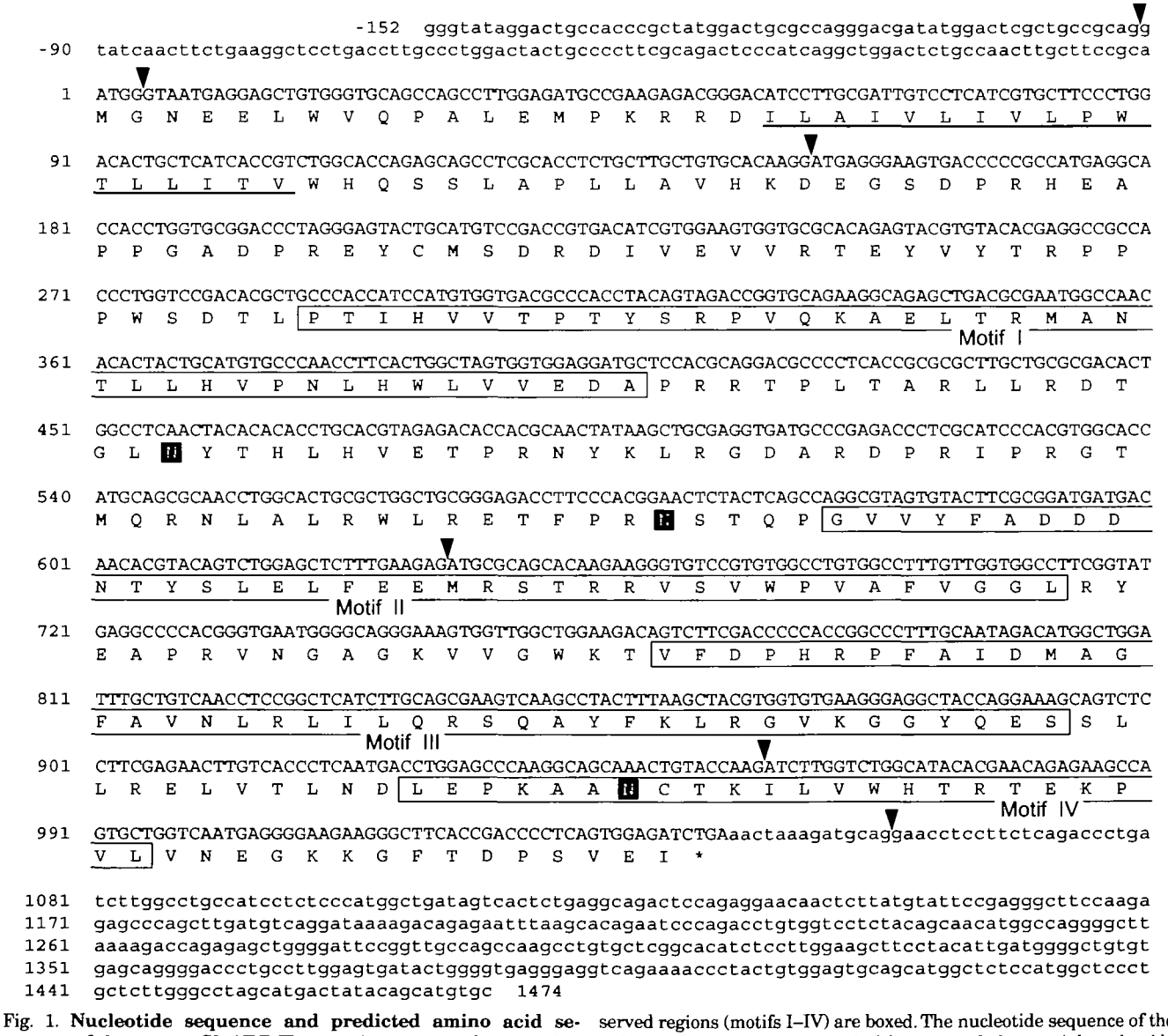

Fig. 1. **Nucleotide sequence and predicted amino acid sequence of the mouse GlcAT-P.** The putative transmembrane region is underlined. Black backgrounds indicate the potential *N*-glycosylation sites. The splice junctions are indicated by arrowheads. Four conserved regions (motifs I–IV) are boxed. The nucleotide sequence of the coding region is shown in capital letters, and the partial nucleotide sequences of the 5′- and 3′-untranslated regions in lowercase letters.

RESULTS

Cloning of the Mouse GlcAT-P cDNA—From 5×10^5 plaques of a mouse brain cDNA library screened using the rat GlcAT-P cDNA as a probe, two positive clones were obtained and sequenced (Fig. 1). The two sequences overlapped each other and contained the same open reading frame of 1,044 bp. These two clones contained 152 bp of a 5′-untranslated region and 466 bp of a 3′-untranslated region, respectively, although a polyadenylation signal and a poly(A)-like sequence were not found in these clones. The nucleotide sequence of the coding region of mouse GlcAT-P exhibits 96.2 and 88.6% identity with those of the rat and human GlcAT-P cDNAs, respectively. The translation initiation codon was assigned to the first in-frame ATG of the open reading frame, and the nucleotide sequence around it well matched the Kozak consensus sequence (*32*). The deduced amino acid sequence of mouse GlcAT-P encodes a protein of 347 amino acids including three potential *N*-glycosylation sites and a single potential transmembrane region consisting of 17 amino acids (Fig. 1). Alignment of the deduced amino acid sequence of the mouse GlcAT-P with those of the rat and human GlcAT-Ps revealed 99.7 and 98.2% sequence identity, respectively (Fig. 2). Only one and six amino acids of the mouse GlcAT-P are different from those of the rat and human GlcAT-Ps, respectively. Four conserved regions (motifs I–IV) previously identified as common features of GlcATs and three potential *N*-glycosylation sites are completely conserved in all species. Northern blot analysis of mouse brain revealed that mouse GlcAT-P transcripts exhibited in two isoforms as in the case of rat and human GlcAT-P transcripts: the predominant transcript was 4.2 kb and the minor one about 10 kb (data not shown).

Genomic Organization of the Mouse GlcAT-P Gene—

```
rat      MGNEELWAQPALEMPKRRDILAIVLIVLPWTLLITVWHQSSLAPLLAVHK
mouse    MGNEELWVQPALEMPKRRDILAIVLIVLPWTLLITVWHQSSLAPLLAVHK 50
human              MPKRRDILAIVLIVLPWTLLITVWHQSTLAPLLAVHK
                             Transmembrane

DEGSDPRHEAPPGADPREYCMSDRDIVEVVRTEYVYTRPPPWSDTLPTIHVVTPTYSRPV
DEGSDPRHEAPPGADPREYCMSDRDIVEVVRTEYVYTRPPPWSDTLPTIHVVTPTYSRPV 110
DEGSDPRRETPPGADPREYCTSDRDIVEVVRTEYVYTRPPPWSDTLPTIHVVTPTYSRPV

QKAELTRMANTLLHVPNLHWLVVEDAPRRTPLTARLLRDTGLNYTHLHVETPRNYKLRGD
QKAELTRMANTLLHVPNLHWLVVEDAPRRTPLTARLLRDTGLNYTHLHVETPRNYKLRGD 170
QKAELTRMANTLLHVPNLHWLVVEDAPRRTPLTARLLRDTGLNYTHLHVETPRNYKLRGD
      Motif I

ARDPRIPRGTMQRNLALRWLRETFPRNSTQPGVVYFADDDNTYSLELFEEMRSTRRVSVW
ARDPRIPRGTMQRNLALRWLRETFPRNSTQPGVVYFADDDNTYSLELFEEMRSTRRVSVW 230
ARDPRIPRGTMQRNLALRWLRETFPRNSSQPGVVYFADDDNTYSLELFEEMRSTRRVSVW
                                        Motif II

PVAFVGGLRYEAPRVNGAGKVVGWKTVFDPHRPFAIDMAGFAVNLRLILQRSQAYFKLRG
PVAFVGGLRYEAPRVNGAGKVVGWKTVFDPHRPFAIDMAGFAVNLRLILQRSQAYFKLRG 290
PVAFVGGLRYEAPRVNGAGKVVRWKTVFDPHRPFAIDMAGFAVNLRLILQRSQAYFKLRG
                                Motif III

VKGGYQESSLLRELVTLNDLEPKAANCTKILVWHTRTEKPVLVNEGKKGFTDPSVEI
VKGGYQESSLLRELVTLNDLEPKAANCTKILVWHTRTEKPVLVNEGKKGFTDPSVEI 347
VKGGYQESSLLRELVTLNDLEPKAANCTKILVWHTRTEKPVLVNEGKKGFTDPSVEI
                  Motif IV
```

Fig. 2. **Comparison of the deduced amino acid sequence of mouse GlcAT-P with the rat and human sequences.** The putative transmembrane and four conserved motifs are underlined with single and double lines, respectively. Black backgrounds indicate the different amino acid residues between the mouse GlcAT-P and the others.

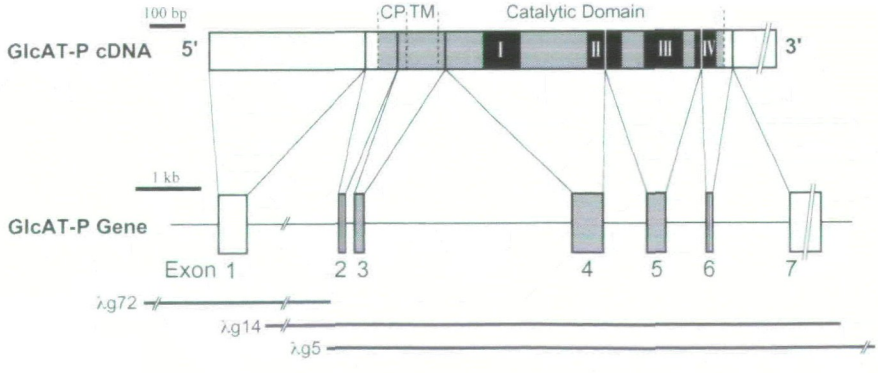

Fig. 3. **Structure of the mouse GlcAT-P gene.** Schematic representation of the mouse GlcAT-P cDNA and GlcAT-P gene. Open boxes indicate the 5′- and 3′-untranslated regions. Gray and black boxes indicate the coding region and conserved motifs I–IV, respectively. Exons are boxed and numbered 1–7. The three lambda clones analyzed are shown as bold lines. CP, cytoplasmic tail; TM, transmembrane region.

From 2×10^6 plaques of a 129/Sv mouse genomic library screened using the full-length mouse GlcAT-P cDNA as a probe, several overlapping clones encoding the GlcAT-P genomic sequence were isolated. However, none of the clones obtained on the first screening contained the first exon encoding the 5′-untranslated region. We performed second screening using a 158 bp PCR product corresponding to exon 1 as a probe, and isolated several clones containing the 5′-flanking region, exon 1, and intron 1. As shown in Fig. 3, the mouse GlcAT-P gene comprises at least seven exons and six introns, and spans over 23 kb of genomic DNA. All exon/intron boundaries were determined by DNA sequencing, and intron sizes were determined by PCR with primer pairs derived from flanking exons (Table I). All assigned exon/intron boundaries agree with consensus 5′-GT and 3′-AG splicing sequences (*33*). Exon 1 contains the 5′-untranslated region. Exon 2 contains the ATG translation initiation codon, and exon 3 contains the transmembrane region. Most of the catalytic region comprised exons 4, 5, and 6. Exon 7 contains the 3′-untranslated region. To confirm that the GlcAT-P gene is a single-copy gene in the mouse genome, we performed Southern blot analysis of mouse genomic DNA digested with *Eco*RI, *Bam*HI, *Spe*I, *Sph*I, *Xba*I, and *Hin*dIII. The *Sca*I–*Sac*I fragment (445 bp) of the GlcAT-P gene corresponding to a part of exon 4 was used as a probe. For each sample, a single band was detected, and the sizes of the bands showed good coincidence with those predicted on restriction enzyme site mapping of genomic DNA clones (data not shown). This indicates that GlcAT-P is encoded by a single-copy gene in the mouse genome.

Chromosomal Localization of the Mouse GlcAT-P Gene—FISH analysis was performed for chromosomal mapping of the mouse GlcAT-P gene (Fig. 4). The full-length mouse GlcAT-P cDNA was used as a probe. A total of 36 doublet fluorescent signals was observed, of which 33 (91.7%) were detected exclusively for chromosome 9. From comparison with the DAPI G-like banding pattern, the mouse GlcAT-P gene was assigned to band A4 on mouse chromosome 9 (9A4). This region well matched the locus (11q25) of the human GlcAT-P gene reported previously (*28*).

Alternative Splicing of the Mouse GlcAT-P Transcripts—Previously we demonstrated that two types of human mRNA were produced depending on the presence or absence of a 17 bp insertion immediately after the initiation codon and encoded two different lengths of the cytoplasmic region of GlcAT-P (*28*). The major isoform of human GlcAT-P mRNA had the 17 bp insertion sequence and was expected to produce a short form of GlcAT-P. To determine whether or not the insertion sequence is present in mouse GlcAT-P mRNA, RT-PCR analysis of mouse brain total RNA was performed with sense and antisense primers spe-

TABLE I. **Exon–intron organization of the mouse GlcAT-P gene.** Exon sequences are shown in uppercase letters and intron sequences in lowercase letters. Deduced amino acid residues and their residue numbers are displayed below the nucleotide sequence.

Exon	Exon size	Splice donor	Intron	Intron size	Splice acceptor
1	5′UTR*	GCTGCCGGAG gt cagtgtcc	1	~16 kb	tctcttac ag GTATCAACTT
2	5′UTR+4 bp	TCCGCA ATG G gt gggtgtga Met 1	2	0.14 kb	caccgctc ag GT AAT GAG CTG Asn Glu Leu 3 4 5
3	147 bp	GTG CAT AAG G gt gagcccac Val His Lys 48 49 50	3	3.2 kb	acctctgc ag AT GAG GGA AGT Glu Gly Ser 52 53 54
4	509 bp	C TTT GAA GAG gt gcgctagg Phe Glu Glu 218 219 220	4	0.5 kb	caccttgt ag ATG CGC AGC A Met Arg Ser 221 222 223
5	297 bp	C TGT ACC AAG gt aagaccct Cys Thr Lys 317 318 319	5	0.65 kb	ggatatgc ag ATC TTG GTC T Ile Leu Val 320 321 322
6	87 bp+3′UTR	AAAGATGCAG gt aggagaga	6	1.3 kb	ccccacac ag GAACCTCCTT
7	3′UTR				

*UTR, untranslated region.

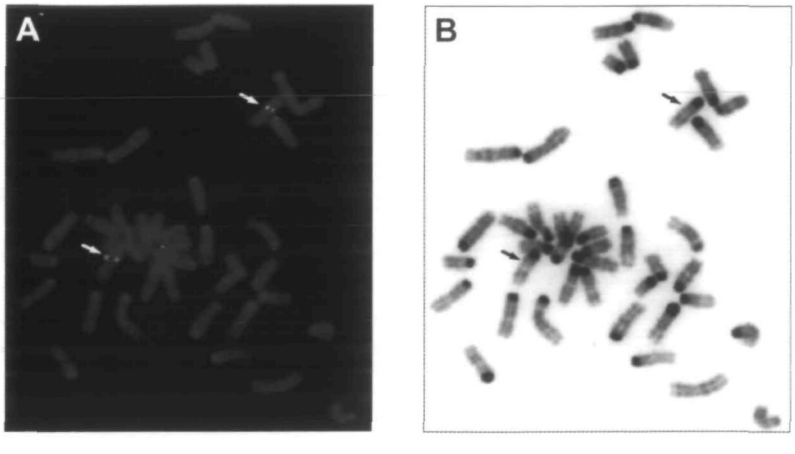

Fig. 4. **Chromosomal localization of the mouse GlcAT-P gene.** A, metaphase chromosomes hybridized with a biotin-labeled mouse GlcAT-P probe. Arrows indicate specific hybridization signals. B, chromosomes with specific signals were identified by post-hybridization DAPI-produced G-like banding as chromosome 9, and the signal was localized at band A4.

cific for the nucleotide sequences of exons 2 and 3, respectively (Fig. 5A). In addition to the major amplified product of 69 bp in length expected for the cloned GlcAT-P cDNA, a minor product of 85 bp in length was detected (Fig. 5B). Sequencing analysis of the two products revealed that the minor product contained a 16 bp insertion (Fig. 5C, sequence underlined) in the same position as the 17 bp insertion of human GlcAT-P. The nucleotide sequence of this insertion was found in intron 2 of GlcAT-P genomic DNA just after exon 2 (Fig. 5D). A consensus 5′-GT splicing donor sequence existed just after the insertion sequence in the genomic DNA. These results suggested that the two isoforms of mouse GlcAT-P mRNA were produced through alternative splicing of the primary gene transcript. This insertion induced a frameshift of the open reading frame from the first ATG codon, which resulted in translation from the second in-frame ATG and production of a truncated isoform of GlcAT-P protein lacking the N-terminal 13 amino acids (Fig. 5C).

Recently, the RIKEN Mouse Gene Encyclopaedia Project reported a mouse cDNA collection (*34*), which includes one clone of the mouse GlcAT-P (accession number AK003020). This cDNA clone lacks the 3′-terminus of the open reading frame, and includes the full sequence of intron 2 at the junction of exons 1 and 2. We also detected a weak band, of approximately 210 bp, corresponding to this unspliced isoform of intron 2 by the same RT-PCR analysis (data not shown). The translation initiation site for this type of GlcAT-P transcript is assigned as the second in-frame ATG.

These results suggested that one major and two minor transcripts are produced from the mouse GlcAT-P gene through alternative splicing, and long and short enzymes, different in the length of the cytoplasmic tail, are produced from these transcripts in mouse brain.

Determination of the Transcription Initiation Site of GlcAT-P—To define the transcription initiation site, 5′-RACE was performed with total RNA from mouse brain as a template. Based on the sequence of the 5′-untranslated region, we determined the precise transcription initiation site by primer extension analysis (Fig. 6). For the extension, a fluorescently labeled 27-mer oligonucleotide complementary to the exon 1 sequence was used. The extension reaction was carried out with poly(A)$^+$ RNA from 6-day-old mouse brain, and the end point of extension was determined by comparison with a sequence ladder derived from the genomic DNA template using the same labeled oligonucleotide as a primer. The extension analysis revealed a strong signal corresponding to the C (+1) residue 552 nucleotides upstream from the initiation codon, ATG. These results imply that there are no additional exons and no alternatively spliced 5′-untranslated exons.

Potential Regulatory Elements in the 5′-Flanking Sequence of the Mouse GlcAT-P Gene—To identify sequence elements that are involved in transcriptional regulation of the GlcAT-P gene, we determined the nucleotide sequence of approximately 1.6 kb of the 5′-flanking region (Fig. 7). The proximal 5′-flanking region of the GlcAT-P gene (bp −250 to +150) is characterized by a GC-rich sequence (GC content, 78%). The sequence around the transcription initiation site contained neither a canonical TATA box nor a CCAAT box. The AG-repeat sequence (purine-rich region) was found in the region between bp −850 and −725. A com-

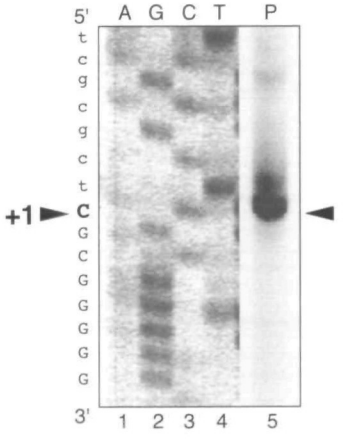

Fig. 6. **Determination of the 5′-end of the mouse GlcAT-P mRNA.** A primer extension analysis of mouse GlcAT-P. A fluorescently labeled oligonucleotide (indicated by an arrow in Fig. 7) was annealed to 5 μg of mouse brain poly(A)$^+$ RNA (lane 5), and then extended by reverse transcriptase. The extension product was separated on a 4.2% polyacrylamide sequencing gel in parallel with a DNA sequencing reaction (lanes 1–4).

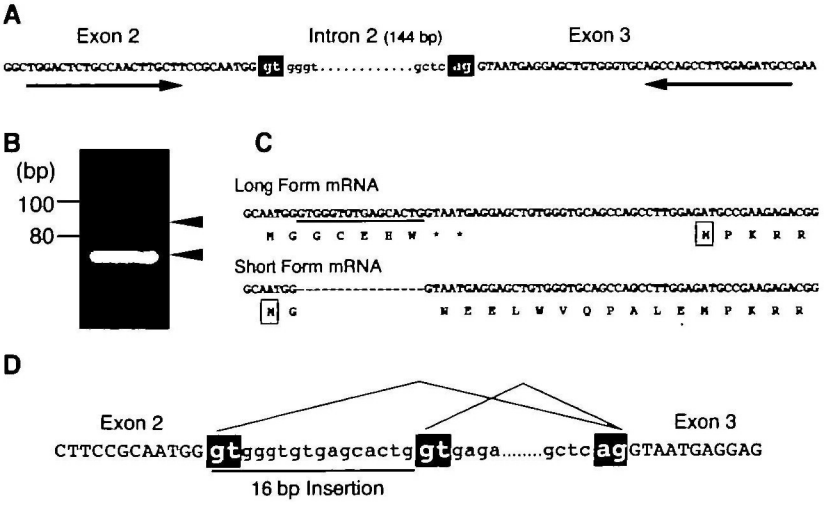

Fig. 5. **Splice variants of the GlcAT-P mRNA.** A, the partial nucleotide sequences of exons 2 and 3, and intron 2 are shown. Arrows indicate the primers used for RT-PCR analysis. Exon sequences are shown in capital letters, and intron sequences in lowercase letters. Black backgrounds indicate a splice donor and a splice acceptor. B, RT-PCR products with the primers shown in A. Arrowheads indicate two specific products. The positions and sizes of the DNA markers are shown at the left. C, the nucleotide sequences of the RT-PCR products. The insertion sequence (16 bp) is underlined. The methionines at the translation initiation site are boxed. D, alternative splicing of the mouse GlcAT-P transcript. The insertion sequence (16 bp) is underlined. Black backgrounds indicate two splice donors and a splice acceptor.

puter-assisted search revealed multiple sequence elements exhibiting homology to known binding sequences for transcription factors. Three Sp1 binding sequences (*35*) are distributed around the transcription initiation site at nucleotide positions +18, +5, and –67, and four additional Sp1 sites are also located at positions –182, –360, –582, and –696. A potential Krox-20 (Egr-2) site (*36*) is located at position +13, overlapping two Sp1 sites just after the transcription initiation site. The 5′-flanking region also included response elements for Ets (*37*), CRE-BP (CRE-binding protein) (*38*), C/EBPβ (CCAAT/enhancer binding protein beta) (*39*), MyoD (myoblast determination gene product) (*40*), AP-1 (activator protein 1) (*41*), p300 (E1A-associated 300 kDa protein) (*42*), POU factor Brn-2 (*43*), HSF2 (heat shock factor 2) (*44*), and CDP (cut-like homeodomain protein) (*45*).

Promoter Activity of the 5′-Flanking Region of the Mouse GlcAT-P Gene—To characterize the regions regulating the transcription of the GlcAT-P gene, we constructed a series of chimeric plasmids containing different lengths of the 5′-flanking region of the GlcAT-P gene fused to the promoterless luciferase gene in pPGBII (Fig. 8). The constructs were

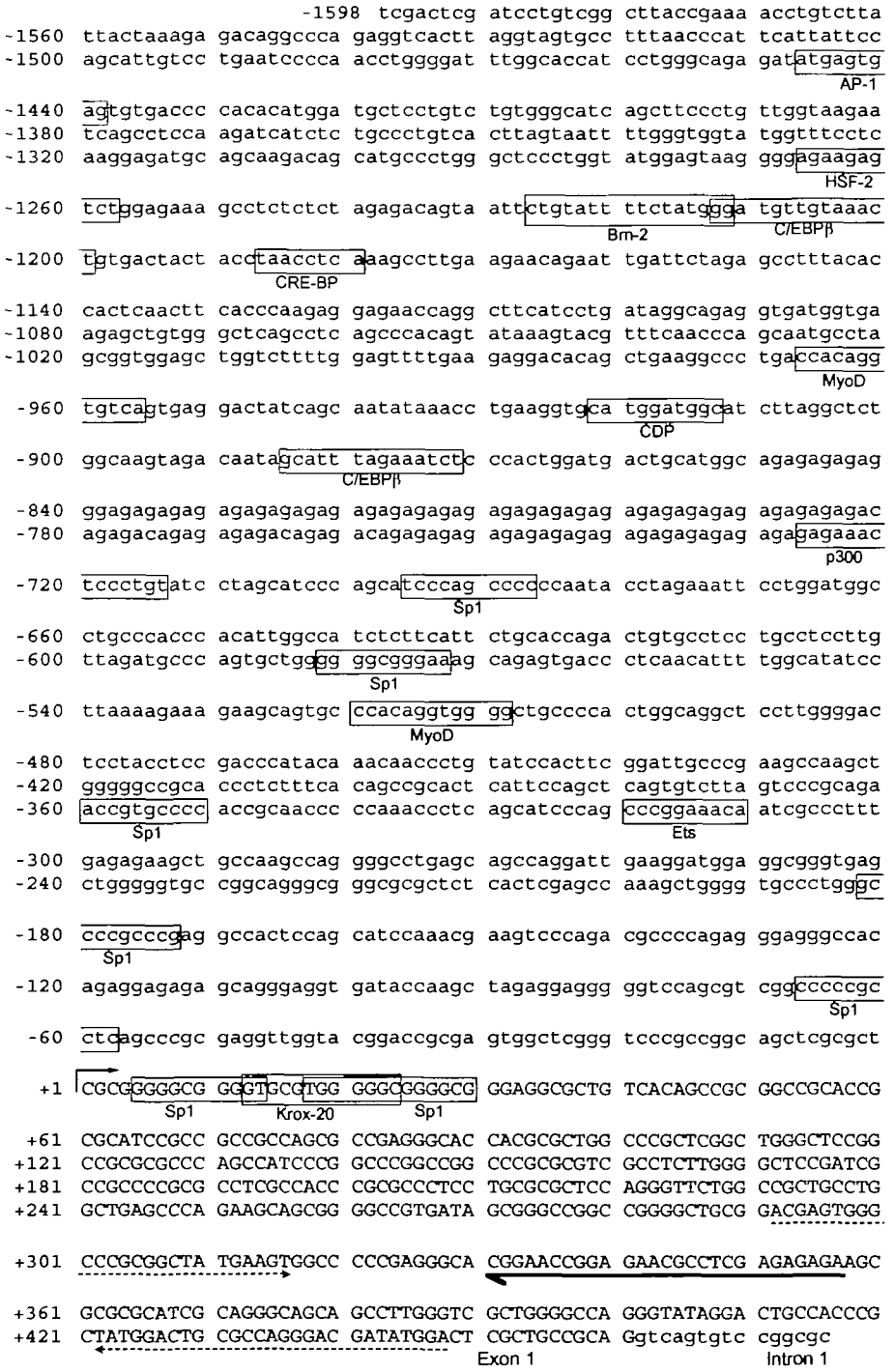

Fig. 7. **Nucleotide sequence of the 5′-flanking region of the mouse GlcAT-P gene.** The transcription initiation site (+1) is indicated by a bent arrow. The sequence of exon 1, shown in capital letters, is followed by a partial sequence of intron 1, in lowercase letters. Transcription factor binding sites are boxed. The oligonucleotide primer used in the primer extension analysis (Fig. 6) is indicated by an arrow. The primer pairs used in the screening of genomic DNA containing exon 1 are indicated by dotted arrows.

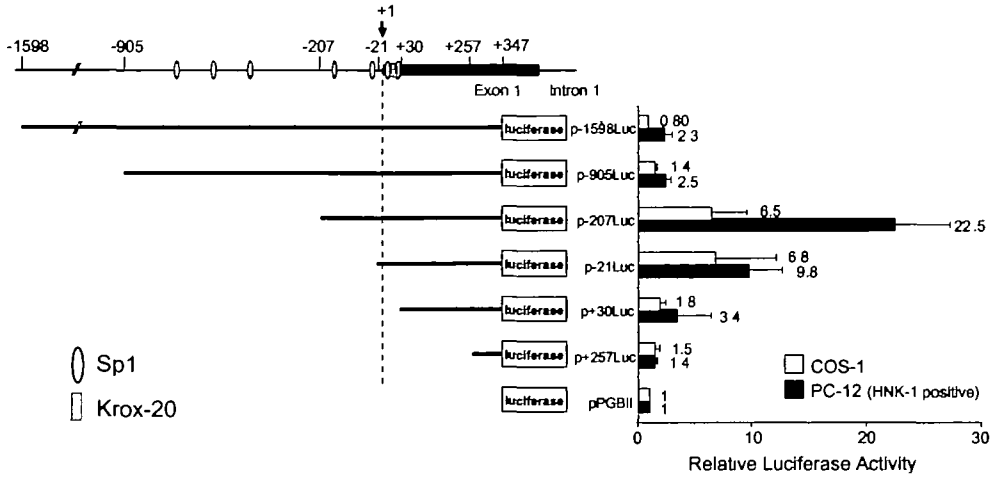

Fig. 8. **GlcAT-P gene promoter activity.** Schematic map of the Sp1 and Krox-20 binding sites, and schematic diagrams of various promoter-luciferase constructs are shown on the left. The binding sites of Sp1 and Krox-20 are shown in the diagrams. The vertical arrow indicates the transcriptional initiation site (+1) determined by primer extension analysis (Fig. 6). Exon 1 is shown as a gray box. Each construct was transfected into PC-12 cells (HNK-1 positive) or COS-1 cells (HNK-1 negative). Promoter activity was normalized as to the *Renilla* luciferase activity of a co-transfected internal control plasmid (pRL-TK). Luciferase activities are expressed relative to those of the same cells transfected with a promoterless luciferase plasmid, pPGBII. Data are the mean from at least three independent transfections performed in triplicate. Error bars indicate standard deviations.

transfected into pheochromocytoma PC-12 cells, which endogenously express the HNK-1 carbohydrate epitope (*46*), and COS-1 cells which do not express this epitope. The promoterless luciferase vector, pPGBII, was transfected into parallel cultures as a negative control. As shown in Fig. 8, plasmid p+257Luc containing the sequence from bp +257 to +347 generated little activity at the level of pPGBII. Elongation of the sequence to position –21 markedly increased the luciferase activity in both cell lines, p-21Luc generating approximately 10- and 7-fold higher luciferase activity than pPGBII in PC-12 and COS-1 cells, respectively. This increase suggests that this region, bp –21 to +347, functions as a promoter independent of the cell line, and that two Sp1 sites, located between bp –21 and +29, are likely to be associated with the basal transcriptional activity of the GlcAT-P gene promoter. Further elongation of the sequence to position –207 had different effects on the cell lines tested. In PC-12 cells, p-207Luc generated the highest activity of all constructs prepared (22.5-fold higher than that of pPGBII). The activity was 3.5-fold higher than that in COS-1 cells. In COS-1 cells, however, p-207Luc was unable to generate such a significant increase in activity compared with p-21Luc. The luciferase activities of p-207Luc and p-21Luc in COS-1 cells were 6.5- and 6.8-fold higher than that of pPGBII, respectively. These results implied that this region, bp –207 to –22, functions as a potent promoter specific for HNK-1 carbohydrate-expressing cells such as PC-12. A database search of this region revealed that, except for two Sp1 sites, no *cis*-elements so far identified are involved in the regulation of GlcAT-P expression in the nervous system or neuronal cells. Further elongation of the 5′-upstream sequence to position –905 greatly decreased the luciferase expression in both cell lines, and elongation to position –1598 maintained the activity at a low level. The marked decrease indicated the existence of repressive elements in this region, bp –905 to –208.

DISCUSSION

In this study, we isolated cDNA and genomic DNA of mouse GlcAT-P, determined the genomic structural organization and chromosomal localization of the mouse GlcAT-P gene, and identified its promoter. GlcAT-P is a glucuronyltransferase involved in HNK-1 carbohydrate epitope biosynthesis, and is thought to be a key enzyme in the HNK-1 biosynthesis pathway (*25*). Because expression of the HNK-1 carbohydrate epitope is strictly regulated during development, it is important to characterize the genomic structure of GlcAT-P and the mechanism involved in the regulation of GlcAT-P expression.

We succeeded in isolating two cDNA clones of mouse GlcAT-P, but no polyadenylation signal or polyadenylation-like sequence were found in the 3′-untranslated region. As reported previously, the human GlcAT-P cDNA has a 3′-untranslated region of 2.3 kb in length (*28*). Northern blot analysis of mouse brain revealed that the major transcript of GlcAT-P was 4.2 kb in length (data not shown). These results suggested the 3′-untranslated region of cloned mouse cDNA was incomplete and therefore contained no polyadenylation signal. The mouse GlcAT-P cDNA exhibits significant high homology with the rat and human cDNAs at the nucleotide and amino acid levels. Although two in-frame ATGs are present in the cytoplasmic region (nucleotide positions +1 and +40 in Fig. 1), the first ATG (+1) obeys Kozak's rule better (*32*) than the second one, and is more reasonable for the translation initiation site. The putative amino acid sequences of four conserved regions (motifs I–IV) of the catalytic domain were completely conserved in all three species. The *N*-glycosylation sites and the cysteine residue located in motif IV, which is essential for the enzymatic activity (*25*), are also conserved. These data indicate that GlcAT-P was highly conserved during phylogenesis and attest to its biological importance.

We also cloned the mouse GlcAT-P gene and defined its genomic organization. The mouse GlcAT-P gene consists of at least seven exons and six introns, and spans more than 23 kb. The coding region of GlcAT-P is composed of five exons spanning approximately 6 kb. Because length of intron 1 is over 15 kb, a second screening was necessary to isolate genomic clones containing the promoter region and exon 1 of the GlcAT-P gene. Most of the 5′-untranslated

region of cDNA was encoded in exon 1. The nucleotide sequence of exon 1 was found in the cDNA clones isolated by the library screening and all 5′-RACE cDNA fragments sequenced, indicating little possibility that the transcription starts between exons 1 and 2. RT-PCR analysis of the GlcAT-P transcript revealed that three forms of the GlcAT-P mRNA were generated through the alternative splicing of intron 2. The major splice isoform of GlcAT-P mRNA with no insertion between exons 1 and 2 is translated to the long-form enzyme carrying a cytoplasmic region composed of 19 amino acids. The other minor splice isoforms contained a 16 bp insertion sequence just after the first in-frame ATG or the full-length intron 2 without splicing. They are suggested to be translated from the second in-frame ATG codon and to generate the short-form enzyme lacking 13 amino acids at the N-terminus. The existence of alternative protein isoforms with or without a 13 amino acid extension in the N-terminus region is well known in the case of β1,4-galactosyltransferase (β4GalT-I) (*47*). However, the two forms of β4GalT-I mRNA are generated through an alternative mode of transcription involving either housekeeping or cell-specific promoters (*48, 49*). The short form of β4GalT-I is transcribed by the cell-specific promoter from a region of the genomic DNA between two in-frame ATG codons. The long and short β1,4-galactosyltransferases were shown to exhibit different intracellular distributions and thus were presumed to have different biological functions (*50*). The length of the extension of the cytoplasmic region was 13 amino acids in both cases (β4GalT-I and GlcAT-P), although the sequences of the extensions exhibited no homology with each other (data not shown). These results suggested the possibility that long and short GlcAT-Ps may be involved in different biological functions, like the two isoforms of β4GalT-I.

Southern blot analysis and chromosomal analysis by FISH revealed that GlcAT-P represents a single copy locus in the mouse genome. The GlcAT-P gene was localized to the A4 region on mouse chromosome 9. This region is syntenic with human chromosome 11q22-25. Therefore, this result is well coincident with our previous finding that the human GlcAT-P is localized to human chromosome 11q25 (*28*). Primer extension analysis identified a major transcriptional initiation site 552 nucleotides upstream of the translation initiation codon. Sequence analysis of the GlcAT-P 5′-flanking region revealed the presence of multiple potential *cis*-acting elements including Sp1, Krox-20, Ets, CREB, and C/EBPβ sites. The 5′ segment around the transcriptional initiation site is highly GC-rich and lacks a TATA or CCAAT box. A GC-rich element known as a CpG island often plays an important role in a promoter lacking a TATA box. These structural features are usually associated with housekeeping gene promoters, but also associated with many tissue-specific promoters including neural cell-specific TATA-less ones (*51*).

We performed the luciferase reporter assay with PC-12 and COS-1 cells. PC-12 cells express GlcAT-P (data not shown) and its product, the HNK-1 carbohydrate epitope (*46*), while COS-1 cells express neither. The proximal region, bp –21 to +347, generated promoter activity in PC-12 and COS-1 cells, suggesting that this is the basic promoter region, and two Sp1 sites located at +3 and +23 are possibly involved in its activity. The region between bp –207 and +347 markedly enhanced the promoter activity in PC-12 cells, the activity being 22.5-fold higher than that of the promoterless luciferase vector. On the other hand, no such enhancement occurred in COS-1 cells. Therefore this region was assumed to be the cell-specific promoter involved in the actual expression of GlcAT-P. Although the region bp –207 to –22 is essential for high promoter activity in HNK-1 expressing cells, no known putative binding sites of neural-specific transcription factors were found in this region. The novel transcription factor binding to this region may be involved in the promoter activity of the GlcAT-P gene. So far, the detailed mechanism of the cell-specific transcriptional activity of this promoter remains to be elucidated. However, it is noteworthy that this promoter region contains a Krox-20 (Egr-2) site. Krox-20 is a zinc finger transcription factor, which is thought to play important roles in vertebrate hindbrain segmentation or in the peripheral nerve myelination (*52*). Krox-20 is expressed in rhombomeres r3 and r5 of the hindbrain in the course of mouse and chick embryogenesis (*53, 54*). Alternate expression of the HNK-1 epitope in the same regions is observed in chick hindbrain (*16*). However, further experiments are needed to determine whether Krox-20 is involved in the strong promoter activity of the region bp –207 to +347. Further elon-

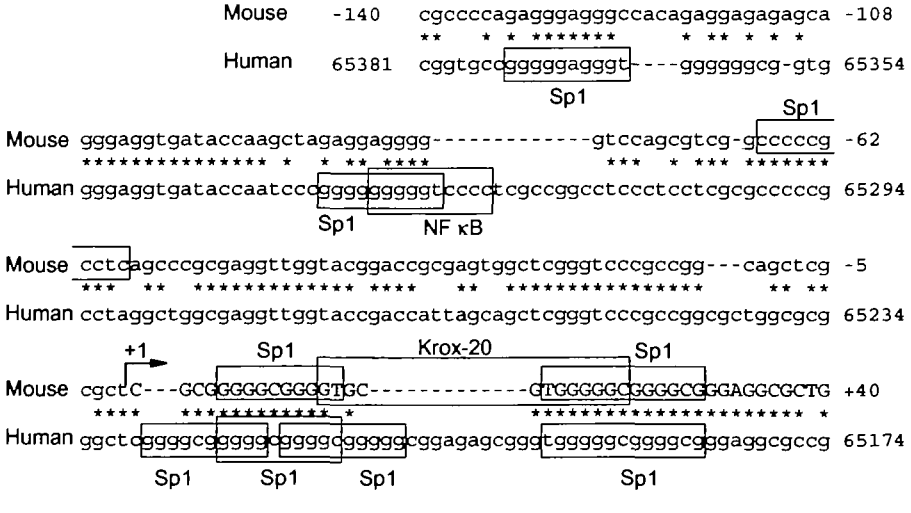

Fig. 9. **Sequence comparison of the 5′-flanking regions of the mouse and human GlcAT-P genes.** Human sequence flanking exon 1 (accession number AC013591) was obtained by searching the GenBank/EBI Data Bank. Numbering of the mouse sequence is as described in Fig. 7, and the human sequence is numbered according to the submitted sequence. Exon sequence is shown in capital letters. Transcription factor binding sites are boxed. Asterisks denote the identical nucleotides.

gation of the fragment of the 5′-flanking region in the reporter constructs resulted in a remarkable decrease in the promoter activity, and the reporter construct containing the sequence from bp −1598 to +347 generated low promoter activity even in PC-12 cells expressing the HNK-1 epitope endogenously. These results suggest the presence of regulatory regions associated with the transcription of the GlcAT-P gene in the further upstream region.

Our previous study showed that the human GlcAT-P mRNA was detected not only in neural tissues but also in liver on Northern blot analysis, although rodent GlcAT-P is expressed specifically in neural tissues (*28*). We obtained a partial sequence of the 5′ flanking region of human GlcAT-P by means of a database search (accession number AC-013591). Comparison of the 140 bp sequence upstream of exon 1 of the mouse GlcAT-P gene with the sequence of the human locus revealed 73% identity (Fig. 9). Compared with the mouse sequence, a 13 bp extra sequence was found in the human promoter region. Due to this sequence, a NF-κB binding motif is present in the human promoter. In the region just after the transcriptional initiation site defined in the mouse gene, another 13 bp extra sequence was found, and consequently the Krox-20 site was absent from the human GlcAT-P genomic DNA. The expression of human GlcAT-P in liver may be explained by these structural differences in promoter regions. Thus, comparison of the promoter structure of the GlcAT-P gene between man and rodents may enable us to elucidate the mechanism underlying the tissue specificity of GlcAT-P and HNK-1 expression.

Recently, a second glucuronyltransferase GlcAT-S involved in biosynthesis of the HNK-1 epitope was cloned (*26, 27*). GlcAT-P and GlcAT-S exhibited different expression patterns in rat brain (unpublished results) and different enzymatic properties (*24, 27*). Comparative analysis of the structures and promoter regions of GlcAT-P and GlcAT-S genes will provide us with important information for understanding the regulated expression and biological function of GlcATs.

We would like to thank H. Nishigaki and T. Honda for secretarial assistance.

REFERENCES

1. Abo, T. and Balch, C.M. (1981) A differentiation antigen of human NK and K cells identified by a monoclonal antibody (HNK-1). *J. Immunol.* **127**, 1024–1029
2. McGarry, R.C., Helfand, S.L., Quarles, R.H., and Roder, J.C. (1983) Recognition of myelin-associated glycoprotein by the monoclonal antibody HNK-1. *Nature* **306**, 376–378
3. Kruse, J., Mailhammer, R., Wernecke, H., Faissner, A., Sommer, I., Goridis, C., and Schachner, M. (1984) Neural cell adhesion molecules and myelin-associated glycoprotein share a common carbohydrate moiety recognized by monoclonal antibodies L2 and HNK-1. *Nature* **311**, 153–155
4. Schachner, M. and Martini, R. (1995) Glycans and the modulation of neural-recognition molecule function. *Trends Neurosci.* **18**, 183–191
5. Yoshihara, Y., Oka, S., Nemoto, Y., Watanabe, Y., Nagata, S., Kagamiyama, H., and, Mori, K. (1994) An ICAM-related neuronal glycoprotein, telencephalin, with brain segment-specific expression. *Neuron* **12**, 541–553
6. Pesheva, P., Horwitz, A.F., and Schachner, M. (1987) Integrin, the cell surface receptor for fibronectin and laminin, expresses the L2/HNK-1 and L3 carbohydrate structures shared by adhesion molecules. *Neurosci. Lett.* **83**, 303–306
7. Gowda, D.C., Margolis, R.U., and Margolis, R.K. (1989) Presence of the HNK-1 epitope on poly(N-acetyllactosaminyl) oligosaccharides and identification of multiple core proteins in the chondroitin sulfate proteoglycans of brain. *Biochemistry* **28**, 4468–4474
8. Xiao, Z.C., Bartsch, U., Margolis, R.K., Rougon, G., Montag, D., and Schachner, M. (1997) Isolation of a tenascin-R binding protein from mouse brain membranes. A phosphacan-related chondroitin sulfate proteoglycan. *J. Biol. Chem.* **272**, 32092–32101
9. Kruse, J., Keilhauer, G., Faissner, A., Timpl, R., and Schachner, M. (1985) The J1 glycoprotein—a novel nervous system cell adhesion molecule of the L2/HNK-1 family. *Nature* **316**, 146–148
10. Chou, D.K., Ilyas, A.A., Evans, J.E., Costello, C., Quarles, R.H., and Jungalwala, F.B. (1986) Structure of sulfated glucuronyl glycolipids in the nervous system reacting with HNK-1 antibody and some IgM paraproteins in neuropathy. *J. Biol. Chem.* **261**, 11717–11725
11. Ariga, T., Kohriyama, T., Freddo, L., Latov, N., Saito, M., Kon, K., Ando, S., Suzuki, M., Hemling, M.E., Rinehart, K.L., Kusunoki, S., and Yu, R.K. (1987) Characterization of sulfated glucuronic acid-containing glycolipids reacting with IgM M-proteins in patients with neuropathy. *J. Biol. Chem.* **262**, 848–853
12. Schwarting, G.A., Jungalwala, F.B., Chou, D.K., Boyer, A.M., and Yamamoto, M. (1987) Sulfated glucuronic acid-containing glycoconjugates are temporally and spatially regulated antigens in the developing mammalian nervous system. *Dev. Biol.* **120**, 65–76
13. Yoshihara, Y., Oka, S., Watanabe, Y., and Mori, K. (1991) Developmentally and spatially regulated expression of HNK-1 carbohydrate antigen on a novel phosphatidylinositol-anchored glycoprotein in rat brain. *J. Cell Biol.* **115**, 731–744
14. Bronner-Fraser, M. (1986) Analysis of the early stages of trunk neural crest migration in avian embryos using monoclonal antibody HNK-1. *Dev. Biol.* **115**, 44–55
15. Nagase, T., Shimoda, Y., Sanai, Y., Nakamura, S., Harii, K., and Osumi, N. (2000) Differential expression of two glucuronyltransferases synthesizing HNK-1 carbohydrate epitope in the sublineages of the rat myogenic progenitors. *Mech. Dev.* **98**, 145–149
16. Kuratani, S.C. (1991) Alternate expression of the HNK-1 epitope in rhombomeres of the chick embryo. *Dev. Biol.* **144**, 215–219
17. Eisenman, L.M. and Hawkes, R. (1993) Antigenic compartmentation in the mouse cerebellar cortex: zebrin and HNK-1 reveal a complex, overlapping molecular topography. *J. Comp. Neurol.* **335**, 586–605
18. Martini, R., Schachner, M., and Brushart, T.M. (1994) The L2/HNK-1 carbohydrate is preferentially expressed by previously motor axon-associated Schwann cells in reinnervated peripheral nerves. *J. Neurosci.* **14**, 7180–7191
19. Keilhauer, G., Faissner, A., and Schachner, M. (1985) Keilhauer, G., Faissner, A., and Schachner, M. (1985) Differential inhibition of neurone-neurone, neurone-astrocyte and astrocyte-astrocyte adhesion by L1, L2 and N-CAM antibodies. *Nature* **316**, 728–730
20. Künemund, V., Jungalwala, F.B., Fischer, G., Chou, D.K., Keilhauer, G., and Schachner, M. (1988) The L2/HNK-1 carbohydrate of neural cell adhesion molecules is involved in cell interactions. *J. Cell Biol.* **106**, 213–223
21. Martini, R., Xin, Y., Schmitz, B., and Schachner, M. (1992) The L2/HNK-1 carbohydrate epitope is involved in the preferential outgrowth of motor neuron on ventral roots and motor nerves. *Eur. J. Neurosci.* **4**, 628–639
22. Saghatelyan, A.K., Gorissen, S., Albert, M., Hertlein, B., Schachner, M., and Dityatev, A. (2000) The extracellular matrix molecule tenascin-R and its HNK-1 carbohydrate modulate perisomatic inhibition and long-term potentiation in the CA1 region of the hippocampus. *Eur. J. Neurosci.* **12**, 3331–3342
23. Voshol, H., van Zuylen, C.W.E.M., Orberger, G., Vliegenthart, J.F.G., and Schachner, M. (1996) Structure of the HNK-1 carbo-

hydrate epitope on bovine peripheral myelin glycoprotein P0. *J. Biol. Chem.* **271**, 22957–22960
24. Terayama, K., Seiki, T., Nakamura, A., Matsumori, K., Ohta, S., Oka, S., Sugita, M., and Kawasaki, T. (1998) Purification and characterization of a glucuronyltransferase involved in the biosynthesis of the HNK-1 epitope on glycoproteins from rat brain. *J. Biol. Chem.* **273**, 30295–30300
25. Terayama, K., Oka, S., Seiki, T., Miki, Y., Nakamura, A., Kozutsumi, Y., Takio, K., and Kawasaki, T. (1997) Cloning and functional expression of a novel glucuronyltransferase involved in the biosynthesis of the carbohydrate epitope HNK-1. *Proc. Natl. Acad. Sci. USA* **94**, 6093–6098
26. Seiki, T., Oka, S., Terayama, K., Imiya, K., and Kawasaki, T. (1999) Molecular cloning and expression of a second glucuronyltransferase involved in the biosynthesis of the HNK-1 carbohydrate epitope. *Biochem. Biophys. Res. Commun.* **255**, 182–187
27. Shimoda, Y., Tajima, Y., Nagase, T., Harii, K., Osumi, N., and Sanai, Y. (1999) Cloning and expression of a novel galactoside β1,3-glucuronyltransferase involved in the biosynthesis of HNK-1 epitope. *J. Biol. Chem.* **274**, 17115–17122
28. Mitsumoto, Y., Oka, S., Sakuma, H., Inazawa, J., and Kawasaki, T. (2000) Cloning and chromosomal mapping of human glucuronyltransferase involved in biosynthesis of the HNK-1 carbohydrate epitope. *Genomics* **65**, 166–173
29. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
30. Church, G.M. and Gilbert, W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995
31. Saito-Ohara, F., Ikeuchi, T., Matsumoto, M., and Kurata, S. (1997) Assignment of the mouse heme oxygenase genes: heme oxygenase-1 (Hmox1) to chromosome 10 band C1 and heme oxygenase-2 (Hmox2) to chromosome 16 band B1. *Cytogenet. Cell Genet.* **77**, 180–181
32. Kozak, M. (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125–8148
33. Shapiro, M.B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174
34. Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H.A., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S., and Hayashizaki, Y. (2001) Functional annotation of a full-length mouse cDNA collection. *Nature* **409**, 685–690
35. Thiesen, H.-J. and Bach, C. (1990) Target detection assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucleic Acids Res.* **18**, 3203–3209
36. Swirnoff, A.H. and Milbrandt, J. (1995) DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol. Cell. Biol.* **15**, 2275–2287
37. Nye, J.A., Petersen, J.M., Gunther, C.V., Jonsen, M.D., and Graves, B.J. (1992) Interaction of murine Ets-1 with GGA-binding sites establishes the Ets domain as a new DNA-binding motif. *Genes Dev.* **6**, 975–990
38. Benbrook, D.M. and Jones, N.C. (1994) Different binding specificities and transactivation of variant CRE's by CREB complexes. *Nucleic Acids Res.* **22**, 1463–1469
39. Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T., and Kishimoto, T. (1990) A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. *EMBO J.* **9**, 1897–1906
40. Davis, R.L., Cheng, P.-F., Lassar, A.B., and Weintraub, H. (1990) The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. *Cell* **60**, 733–746
41. Lee, W., Mitchell, P., and Tjian, R. (1987) Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* **49**, 741–752
42. Rikitake, Y. and Moran, E. (1992) DNA-binding properties of the E1A-associated 300-kilodalton protein. *Mol. Cell. Biol.* **12**, 2826–2836
43. Li, P., He, X., Gerrero, M.R., Mok, M., Aggarwal, A., and Rosenfeld, M.G. (1993) Spacing and orientation of bipartite DNA-binding motifs as potential functional determinants for POU domain factors. *Genes Dev.* **7**, 2483–2496
44. Kroeger, P.E. and Morimoto, R.I. (1994) Selection of new HSF1 and HSF2 DNA-binding sites reveals differences in trimer cooperativity. *Mol. Cell. Biol.* **14**, 7592–7603
45. Harada, R., Berube, G., Tamplin, O.J., Denis-Larose, C., and Nepveu, A. (1995) DNA-binding specificity of the cut repeats from the human cut-like protein. *Mol. Cell. Biol.* **15**, 129–140
46. Bock, E., Richter-Landsberg, C., Faissner, A., and Schachner, M. (1985) Demonstration of immunochemical identity between the nerve growth factor-inducible large external (NILE) glycoprotein and the cell adhesion molecule L1. *EMBO J.* **4**, 2765–2768
47. Shaper, N.L., Hollis, G.F., Douglas, J.G., Kirsch, I.R., and Shaper, J.H. (1988) Characterization of the full length cDNA for murine beta-1,4-galactosyltransferase. Novel features at the 5′-end predict two translational start sites at two in-frame AUGs. *J. Biol. Chem.* **263**, 10420–10428
48. Harduin-Lepers, A., Shaper, N.L., Mahoney, J.A., and Shaper, J.H. (1992) Murine beta 1,4-galactosyltransferase: round spermatid transcripts are characterized by an extended 5′-untranslated region. *Glycobiology* **2**, 361–368
49. Harduin-Lepers, A., Shaper, J.H., and Shaper, N.L. (1993) Characterization of two cis-regulatory regions in the murine beta 1,4-galactosyltransferase gene. Evidence for a negative regulatory element that controls initiation at the proximal site. *J. Biol. Chem.* **268**, 14348–14359
50. Lopez, L.C., Youakim, A., Evans, S.C., and Shur, B.D. (1991) Evidence for a molecular distinction between Golgi and cell surface forms of beta 1,4-galactosyltransferase. *J. Biol. Chem.* **266**, 15984–15991
51. Twyman, R.M and Jones, E.A. (1995) The regulation of neuron-specific gene expression in the mammalian nervous system. *J. Neurogenet.* **10**, 67–101
52. Topilko, P., Schneider-Maunoury, S., Levi, G., Baron-Van, Evercooren, A., Chennoufi, A.B., Seitanidou, T., Babinet, C., and Charnay, P. (1994) Krox-20 controls myelination in the peripheral nervous system. *Nature* **371**, 796–799
53. Wilkinson, D.G., Bhatt, S., Chavrier, P., Bravo, R., and Charnay, P. (1989) Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. *Nature* **337**, 461–464
54. Nieto, M.A., Bradley, L.C., and Wilkinson, D.G. (1991) Conserved segmental expression of Krox-20 in the vertebrate hindbrain and its relationship to lineage restriction. *Development* **2 (Suppl.)**, 59–62