Human glycosaminoglycan glucuronyltransferase I gene and a related processed pseudogene: genomic structure, chromosomal mapping and characterization

Hiroshi KITAGAWA¹, Masateru TAOA, Yuko TONE and Kazuyuki SUGAHARA

Department of Biochemistry, Kobe Pharmaceutical University, Higashinada-ku, Kobe 658-8558, Japan

Here we describe the characterization of the human glycosaminoglycan glucuronyltransferase I gene (*GlcAT-I*) and a related pseudogene. The *GlcAT-I* gene was localized to human chromosome 11q12–q13 by *in situ* hybridization of metaphase chromosomes. *GlcAT-I* spanned 7 kb of human genomic DNA and was divided into five exons. Northern blot analysis showed that *GlcAT-I* exhibited ubiquitous but markedly different expressions in the human tissues examined. The *GlcAT-I* promoter was approx. 3-fold more active in a melanoma cell line than in a hepatoma cell line, providing evidence for the differential regulation of the gene's expression. Stepwise 5' deletions of the promoter identified a strong enhancer element between –303 and –153 bp that included binding motifs for Ets, CREB (cAMP-response-element-binding protein) and STAT (signal

transducers and activators of transcription). Screening of a human genomic library identified one additional distinct genomic clone containing an approx. 1.4 kb sequence region that shared an overall 95.3% nucleotide identity with exons 1–5 of *GlcAT-I*. However, a lack of intron sequences, as well as the presence of several nucleotide mutations, insertions and deletions that disrupted the potential *GlcAT-I* reading frame, suggested that the clone contained a processed pseudogene. The pseudogene was localized to chromosome 3. The human genome therefore contains two related *GlcAT-I* genes that are located on separate chromosomes.

Key words: gene expression, gene structure, glycosaminoglycans, proteoglycan.

INTRODUCTION

Glycosaminoglycans (GAGs) have been implicated in the regulation and maintenance of cell adhesion, cell proliferation, cytodifferentiation and tissue morphogenesis (reviewed in [1–3]). In particular, crucial roles of GAGs, especially heparan sulphate, in developmental processes and signalling pathways of Wingless, Hedgehog and fibroblast growth factor in *Drosophila* have recently become evident [2]. The biological activities of GAGs depend on their ability to interact with various proteins including morphogens, growth factors, cytokines and extracellular matrix proteins through specific saccharide sequences embedded in GAG chains (reviewed in [4]). Therefore the biosynthesis of GAGs is believed to be strictly regulated.

Sulphated GAGs, including heparin/heparan sulphate, chondroitin sulphate and dermatan sulphate, are covalently bound to serine residues in the core proteins through the common carbohydrate–protein linkage region structure, GlcA β 1-3Gal β 1-3Gal β 1-4Xyl β 1-O-Ser [5]. Heparin/heparan sulphate is synthesized once GlcNAc has been transferred to the common linkage region, whereas chondroitin sulphate is formed if GalNAc is added first. The biosynthesis of these GAGs is initiated by the addition of a xylose residue to specific serine residues in the core protein and is followed by the addition of two galactose residues, a GlcA residue and then the alternate addition of *N*-acetylhexosamine and GlcA residues. The addition of each sugar is

thought to be catalysed by a specific glycosyltransferase, which uses the corresponding uridine diphosphate sugar as a donor substrate [5].

The glucuronyl transfer to the galactose residue is catalysed by a β 1,3-glucuronyltransferase termed glucuronyltransferase I (*GlcAT-I*), which transfers GlcA from UDP-GlcA to the trisaccharide-serine, Gal β 1-3Gal β 1-4Xyl β 1-O-Ser, finalizing the formation of the common linkage region, GlcA β 1-3Gal β 1-3Gal β 1-4Xyl β 1-O-Ser. We recently cloned the cDNA encoding the *GlcAT-I* from human placenta [6] on the basis of the information of the amino acid sequence alignment of rat glycoprotein-specific glucuronyltransferase (*GlcAT-P*) involved in the biosynthesis of the carbohydrate epitope HNK-1 with putative proteins in *Caenorhabditis elegans* and *Schistosoma mansoni* [7]. Interestingly, Bai et al. [8] showed that the transfection of Chinese hamster ovary cell mutants defective in *GlcAT-I* with *GlcAT-I* cDNA augmented GAG synthesis to levels approximately double that in wild-type cells, implying that *GlcAT-I* might be rate-limiting. In addition, the transfection of the *GlcAT-I* cDNA into COS-1 cells induced a significant expression of the HNK-1 epitope, suggesting that the high expression of *GlcAT-I* rendered the cells capable of synthesizing the HNK-1 epitope [9,10]. It is therefore important to clarify the regulation of *GlcAT-I* expression.

In the present study we examined the structural organization and functional expression of human *GlcAT-I*. In addition, we

Abbreviations used: DAPI, 4,6-diamidino-2-phenylindole; FISH, fluorescence *in situ* hybridization; GAG, glycosaminoglycan; *GlcAT-I*, glucuronyltransferase I (EC 2.4.1.135); *GlcAT-P*, glycoprotein-specific glucuronyltransferase [7]; SV40, simian virus 40.

¹ To whom correspondence should be addressed (e-mail kitagawa@kobepharm-u.ac.jp).

The nucleotide sequence data reported will appear in DDBJ, EMBL and GenBank® Nucleotide Sequence Databases under the accession numbers AB047194 and AB049116.

identified a related processed pseudogene and localized *GlcAT-I* and the pseudogene to two distinct human chromosomes, 11q12–q13 and 3 respectively.

EXPERIMENTAL

Cells

Cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum. The cell lines used were G361 human melanoma cells (A.T.C.C. CRL-1424), C32 human amelanotic melanoma cells (A.T.C.C. CRL-1585), HepG2 human hepatocellular carcinoma cells (A.T.C.C. HB-8065) and MFH-7 human malignant fibrous histiocytoma cells [11].

Isolation of human *GlcAT-I* genomic clones

A human placenta genomic library in a λ FIX II phage vector (Stratagene, La Jolla, CA, U.S.A.) was screened with the cloned human *GlcAT-I* cDNA fragment of 1.4 kb [6]. Three independent clones (C1–C3) were isolated and characterized. Insert DNA fragments were initially characterized by restriction digestion and Southern blot analysis. Human genomic DNA fragments that hybridized to the *GlcAT-I* cDNA probe were subcloned into Bluescript plasmid vectors and sequenced.

Identification of the 5' end of *GlcAT-I* mRNA

The site of transcription initiation in human placenta was determined by using a Cap site cDNA™ kit (Wako, Osaka, Japan) in accordance with the manufacturer's instructions. The kit is based on a method that consists of removing the cap with tobacco acid pyrophosphatase and ligating an oligoribonucleotide to decapped mRNA species with T4 RNA ligase [12]. The first PCR with an anti-sense primer, 5'-ATCCTTCTGCCGT-AGCTGCTC-3' (the nucleotide sequence corresponding to 124–144; see Figure 1 in [6]), and an anchor primer (1RC) provided by the manufacturer was followed by nested PCR with an anti-sense primer, 5'-TTGCAGCTGGGAAATCCTCAG-3' (the nucleotide sequence corresponding to 145–165; see Figure 1 in [6]), and an anchor primer (2RC) also provided by the manufacturer. For each reaction, 35 cycles (94 °C for 30 s, 55 °C for 30 s, 72 °C for 45 s) were run with *EX Taq* polymerase (Takara Shuzo Co., Kyoto, Japan) in 5% (v/v) DMSO. The final PCR products were subcloned into pGEM®-T Easy vector (Promega, Tokyo, Japan) and sequenced in a 377 DNA sequencer (PE Applied Biosystems, Foster City, CA, U.S.A.). Several clones were sequenced to compensate for misreading by *EX Taq* polymerase.

Northern blot analysis

A commercial human Multiple Tissue Northern Blot I membrane (ClonTech, Tokyo, Japan) was used for the analysis; 2 μ g of polyadenylated RNA was loaded in each lane. The membrane was probed with a gel-purified, radiolabelled (more than 10⁹ c.p.m./ μ g), human *GlcAT-I* cDNA fragment of 1.4 kb [6].

Functional analysis of the promoter

A 1230 bp *GlcAT-I* promoter region, which contained the 1198 bp sequence upstream of the transcription initiation site and the 32 bp exon I 5' sequence, was amplified by PCR with the sense primer 5'-CTAGCTAGCGCGCCAGCTGCTAACA-

GCA-3' (the nucleotide sequence corresponding to –1198 to –1179; see Figure 5), and the anti-sense primer 5'-CCCAA-GCTTGCAGAGAAACACGTTCTTAGC-3' (the nucleotide sequence corresponding to 12–32; see Figure 5), with the *GlcAT-I* genomic clone (C1) as a template. In addition, various lengths of the *GlcAT-I* promoter region were amplified by PCR with one of the sense primers 5'-CTAGCTAGCGGACACAAAAGTGCCTGTGGC-3' (the nucleotide sequence corresponding to –604 to –583; see Figure 5), 5'-CTAGCTAGCTTTTGACTGCCGGCCACGC-3' (the nucleotide sequence corresponding to –303 to –285; see Figure 5) or 5'-CTAGCTAGCCAGCTGGGAGCTGGTAGGGGA-3' (the nucleotide sequence corresponding to –153 to –131; see Figure 5) and the anti-sense primer shown above, with the *GlcAT-I* genomic clone (C1) as a template. *NheI* and *HindIII* restriction sites were introduced at the 5' ends of the sense and anti-sense primers respectively. Each PCR product was digested with *NheI* and *HindIII* and then subcloned into a pGL3-Basic luciferase reporter vector (Promega). The presence of the correct insert was confirmed by direct sequencing. Each construct was mixed with the pRL-TK vector (Promega) and these were transfected into cells on 60 mm plates with FuGENE™ 6 (Roche Molecular Biochemicals, Tokyo, Japan) in accordance with the manufacturer's instructions. The pRL-TK vector, which contains the thymidine kinase promoter from herpes simplex virus and the *Renilla* luciferase reporter gene, was used as an internal control to normalize for the efficiency of the transient transfection. After 48 h the luciferase activity was assayed with the dual luciferase reporter assay system (Promega). As a control we used an empty firefly luciferase vector (pGL3-Basic) and a pGL3-Control vector (Promega) in which the firefly luciferase expression is driven by the simian virus 40 (SV40) promoter/enhancer.

In situ chromosome hybridization

In situ hybridization was performed on chromosome preparations obtained from phytohaemagglutinin-stimulated human lymphocytes cultured for 68–72 h. 5-Bromodeoxyuridine was added for the final 7 h of culture (0.18 mg/ml of medium) to synchronize the cell population. A 3.1 kb *BglII* genomic fragment containing exons I and II (see Figure 1) was used as the probe, which was biotinylated with dATP by using the BioNick labelling kit (Life Technologies, Gaithersburg, MD, U.S.A.). The procedure for fluorescence *in situ* hybridization (FISH) detection was performed as described previously [13,14]. In brief, the chromosome slides were baked at 55 °C for 1 h. After RNase treatment, the slides were treated in 70% formamide in 2 × SSC for 2 min at 70 °C to denature the chromosomes followed by dehydration with ethanol. The probes were denatured in a hybridization mixture consisting of 50% (v/v) formamide and 10% (w/v) dextran sulphate at 75 °C for 5 min. The probes were loaded on the denatured chromosomal slides. After hybridization overnight, slides were washed and signals were detected and amplified with the published method [13]. FISH signals and the 4,6-diamidino-2-phenylindole (DAPI) banding pattern were recorded separately by taking photographs; the assignment of the FISH mapping data with chromosomal bands was achieved by superimposing FISH signals on DAPI-banded chromosomes [14].

RESULTS

Isolation of *GlcAT-I* and determination of the exon/intron structure

Previously we identified a 1005 bp open reading frame in a 1.47 kb human *GlcAT-I* cDNA [6]. To determine the genomic

Figure 1 Genomic organization of human *GlcAT-I*

EcoRI (E), *BglII* (Bg) and *BamHI* (Ba) restriction sites are indicated. Exon regions are denoted by boxes: filled boxes represent the coding sequence and open boxes denote the 5' and 3' untranslated sequences. The translation initiation codon (ATG) and the termination codon (TGA) are also shown. A black horizontal bar denotes the intron.

Table 1 Exons of human *GlcAT-I*

For each exon the cDNA residues and positions of the amino acid sequence predicted by the open reading frame are given as numbered in Figure 1 in [6].

Exon number	Exon length (bp)	Corresponding cDNA residues	Amino acid residues encoded
I	119	— 37 to 82	1–39
II	175	83–257	39–98
III	361	258–618	99–218
IV	291	619–909	218–315
V	504	910–1413	315–335

Table 2 Intron junctions in human *GlcAT-I*

Introns are named after the two exons that they separate. For example, the intron between exon I and exon II is represented as I/II. Capital letters represent exon sequences; small letters indicate intron sequences. Bold letters represent agreement with the consensus sequence reported in [15].

Intron	Intron length (bp)	Splice junction sequence	
		Donor	Acceptor
I/II	1194	CUCG:gugagc	cugggaccag:GCCA
II/III	3155	CCAG:guaggg	ucccugcag:GCUG
III/IV	190	GGAG:gugagc	ccucgguag:AUGC
IV/V	705	UCG:guuagg	ccuggcacag:GUAC

structure of the human *GlcAT-I* gene, we screened a human placenta genomic library with the cloned *GlcAT-I* cDNA fragment and found three independent clones, C1–C3. Insert DNA fragments were characterized by restriction digestion and Southern blot analysis. Sequencing revealed that the two clones (C1 and C2) overlapped each other and contained all exons corresponding to human *GlcAT-I* cDNA. As summarized in Figure 1, the human *GlcAT-I* gene spanned approx. 7 kb and the coding sequence of *GlcAT-I* was divided into five exons. All exons contained the coding sequences indicated in Figure 1. In addition, exon V contained the 3' untranslated region, which included the polyadenylation signal AATAAA. The locations of the *GlcAT-I* exons were determined by sequencing the genomic clones. Sequencing was also done to determine the exact size of the exons and the introns as shown in Table 1 and to determine

Ψ	-60	ggagaaaaataatatttgagagaccccgctgctggggcagggcagcggtggcc
cDNA	-29	*****
Ψ	1	ATGAAGCTGAAGTGAAGACGTGTTCTCGCCTACTTCTGGTGCATCGCGGCCTC
cDNA	1	ATGAAGCTGAAGTGAAGACGTGTTCTCGCCTACTTCTGGTGCATCGCGGCCTC
Ψ	61	CTCTACGCGCTGG-----CCAGCCATGTGACTGCCCTCCCTCCCTGCGGCGAGCA
cDNA	71	CTCTACGCGCTGGTACAGCTCGGCGAGCCTGTGACTGCCCTCCCTGCGGCGAGCA
Ψ	121	GCTGAGCAGCTACGGCAGAGGATCTGAGGATTCCAGCTGCAAGCGGAATCCGACGG
cDNA	131	GCGGAGCAGCTACGGCAGAGGATCTGAGGATTCCAGCTGCAAGCGGAATCCGACGG
Ψ	171	CCACCCCTGCGCCTGCCAGCCCTGACCCGAGGCGCTGCCTACTATCTATGTTGTT
cDNA	181	CCACCCCTGCGCCTGCCAGCCCTGACCCGAGGCGCTGCCTACTATCTATGTTGTT
Ψ	231	ACCCCACTATGCCAGGCTGTACAGAGGAGATTTGGTACGGCTGTCCAGACACTG
cDNA	241	ACCCCACTATGCCAGGCTGTACAGAGGAGATTTGGTACGGCTGTCCAGACACTG
Ψ	291	AGCTGTGCTCCCGGCTGCATTGAGTGTGGTGAAGGATGCTGAGGGTCCACCCCGCTG
cDNA	301	AGCTGTGCTCCCGGCTGCATTGAGTGTGGTGAAGGATGCTGAGGGTCCACCCCGCTG
Ψ	351	GTCTCAGGGCTGTGGTGTGCTCTGGCTCCTCTTACACACCTGGTGGTCTTACGCC
cDNA	361	GTCTCAGGGCTGTGGTGTGCTCTGGCTCCTCTTACACACCTGGTGGTCTTACGCC
Ψ	411	AAAGCCAGCGGCTTCGGGAGGCGCAAGCTGGCTGGTTCATCCCATGGTGTGAGCAG
cDNA	421	AAAGCCAGCGGCTTCGGGAGGCGCAAGCTGGCTGGTTCATCCCATGGTGTGAGCAG
Ψ	471	CGAACAAGGCCCTGGACTGGCTCCGGGCGAGGAGTGTCTGGGTGGGAAGAGGAC
cDNA	481	CGAACAAGGCCCTGGACTGGCTCCGGGCGAGGAGTGTCTGGGTGGGAAGAGGAC
Ψ	531	CCATCACCACCGGACCAAGGAGTGTGTACTTTGCTGACAGTGGCAACACCTACAGC
cDNA	541	CCATCACCACCGGACCAAGGAGTGTGTACTTTGCTGACAGTGGCAACACCTACAGC
Ψ	591	TGGGAGCTGTTTGAGGAGTGGCTGGACCCGCTGGTGTCTCAGTGTGGCTGTGGGGCTG
cDNA	601	TGGGAGCTGTTTGAGGAGTGGCTGGACCCGCTGGTGTCTCAGTGTGGCTGTGGGGCTG
Ψ	651	GTGGGTGGCTGCGAGTCAAGGCGCCCTCAGGTACAGGACAGCGGGTGTGGGTCTCCAC
cDNA	661	GTGGGTGGCTGCGAGTCAAGGCGCCCTCAGGTACAGGACAGCGGGTGTGGGTCTCCAC
Ψ	711	ACAGCATGGGAGCCCAACGGGCTTCCCTGTGGATATGGCCGATTTGCTGTGGCCCTG
cDNA	721	ACAGCATGGGAGCCCAACGGGCTTCCCTGTGGATATGGCCGATTTGCTGTGGCCCTG
Ψ	771	CCCTTGTGTAGCTAAGCCCAATGCCAGTTTGTATTCACCTGCTCCCGGAGCCACCTG
cDNA	781	CCCTTGTGTAGCTAAGCCCAATGCCAGTTTGTATTCACCTGCTCCCGGAGCCACCTG
Ψ	831	GAGAGCAGTCTTCAAGCCACCTTGTGGATCCCAAGGACCTGGAGCCACGGGCTGCCAAC
cDNA	841	GAGAGCAGTCTTCAAGCCACCTTGTGGATCCCAAGGACCTGGAGCCACGGGCTGCCAAC
Ψ	891	TGCATCGGGTACTGGTGTGACATACAGGACAGCGAAGCCCAAGATGAAGCAGAAGGAG
cDNA	901	TGCATCGGGTACTGGTGTGACATACAGGACAGCGAAGCCCAAGATGAAGCAGAAGGAG
Ψ	951	CAGCTGCAGCGGCGAGGCGCGGCTCAGACCCAGCAATTGAGGTGTGATggcgccccac
cDNA	961	CAGCTGCAGCGGCGAGGCGCGGCTCAGACCCAGCAATTGAGGTGTGATggcgccccac
Ψ	1011	cccaactaccacctcttttcaggcacagaccttgggactgggcccagggcctgcccag
cDNA	1021	cccaactaccacctcttttcaggcacagaccttgggactgggcccagggcctgcccag
Ψ	1071	catgtggtttttcaggctgacctggaagcagaagtgcccctctgcccctccaga
cDNA	1081	gatgtggtttttcaggctgacctggaagcagaagtgcccctctgcccctccagg
Ψ	1131	ccagggtcatgtctctgctgcttcaatcttccccagcctgcatgtggcactgcccaca
cDNA	1141	ccagggtcatgtctctgctgcttcaatcttccccagcctgcatgtggcactgcccaca
Ψ	1191	ggctggggacagcagccctgtattgagcaggttggccctgtctagggtggaacagaa
cDNA	1201	ggctggggacagcagccctgtattgagcaggttggccctgtctagggtggaacagaa
Ψ	1251	ggacagacgggacccaggaggaggagcagtgatttaactgggtaacttattggggctgggc
cDNA	1261	ggacagatggggccaggaggaggagcagtgatttaactgggtaacttattggggctgggc
Ψ	1311	aagcactggggggctggaggagctgggctggacccctccctacatgagcatgtgaccccc
cDNA	1321	atgactggggggctggaggagctgggctggacccctccctacatgagcatgtgaccccc
Ψ	1371	ttctacctccagaataaagaatctcaacctgaaaaaaagaaaaaatatttgaaggag
cDNA	1381	ttctacctccagaataaagaatctcaacctgaaaaaaagaaaaaatatttgaaggag

Figure 2 Comparison of human *GlcAT-I* cDNA and the processed pseudogene sequences

The 1.4 kb sequence region of the *GlcAT-I* pseudogene (*ψ*) homologous with the human *GlcAT-I* cDNA sequence is shown above the cDNA sequence. The initiation and termination codons of the cDNA are in bold. The sequences are numbered relative to the translation initiation site of the cDNA. Gaps in the nucleotide sequence alignments are indicated by dashes and are not taken into account in nucleotide numbering. The nucleotide sequences corresponding to the open reading frame are shown in capitals. Asterisks indicate nucleotides that are identical in the two sequences.

the sequences of the intron/exon junctions as shown in Table 2. All the intron/exon junctions followed the GT/AG rule and were flanked by conserved sequences [15]. The sequence encoded

Figure 3 Mapping of *GlcAT-I* to human chromosome 11

Upper panels: an example of FISH mapping of the *GlcAT-I* probe. The left panel shows the FISH signals on chromosome 11; the right panel shows the same mitotic figure stained with DAPI to identify chromosome 11. Lower panel: diagram of FISH mapping results of the *GlcAT-I* probe. Each dot represents the double FISH signals detected on human chromosome 11.

by these exons was identical with that of the human *GlcAT-I* cDNA [6].

Identification of a pseudogene related to human *GlcAT-I*

The mapping data of the remaining clone, C3, was not consistent with that obtained for clones C1 and C2. A sequencing analysis of clone C3 revealed that it contained a related but distinct human *GlcAT-I* sequence. A sequence comparison between the *GlcAT-I* cDNA and the related clone, shown in Figure 2, showed that they shared 95.3% identity and the related sequence contained multiple nucleotide changes, insertions and deletions. These sequence alterations would cause a frameshift and non-sense mutations, which would prematurely terminate the translation of the potential *GlcAT-I* open reading frame that follows the translation initiation start codon located in the same relative position. These findings suggest that clone C3 contained a processed *GlcAT-I* pseudogene.

Figure 4 Northern blot analysis of *GlcAT-I* in human adult tissues

Northern blots with RNA from various human adult tissues were hybridized with a probe for *GlcAT-I*, as described in the Experimental section. The positions of size markers are indicated at the left.

Chromosomal mapping of *GlcAT-I* and the related pseudogene

By using a genomic fragment of *GlcAT-I*, *in situ* hybridization to normal metaphase chromosomes was performed to determine the chromosomal localization of the human gene. Of 100 metaphase cells examined by this hybridization, 61 showed signals on one pair of the long arms of chromosome 11. The detailed position was further determined on the basis of the summary of 10 photos (Figure 3, upper panels). No additional locus was identified by FISH detection under the conditions used (see the Experimental section). These findings clearly indicate that *GlcAT-I* is located at human chromosome 11q12–q13 (Figure 3, lower panel). While this manuscript was in preparation, the database search of the Human Genome Project showed a human chromosome 11q12.2 sequence (accession number AC004230) that encompassed the *GlcAT-I* sequence and was consistent with the present findings. Notably, the database search also revealed that *GlcAT-I* was located downstream of the *ROM1* gene encoding the retinal outer segment membrane protein 1 [16] and separated by an intergenic region of only 194 bp. The two genes are transcribed in opposite directions.

In addition, a search of the unfinished high-throughput genomic sequence database identified a human chromosome 3 'working draft' sequence (accession number AC012057) identical with a *GlcAT-I* processed pseudogene sequence. Thus the *GlcAT-I* processed pseudogene has been mapped to chromosome 3.

Differential expression of *GlcAT-I*

To determine the size and expression pattern of the *GlcAT-I* mRNA in human tissues, Northern blot analysis of mRNA was performed; the results demonstrated a single band of 1.5 kb for all human tissues examined (Figure 4). However, expression of the mRNA varied considerably, exhibiting a unique tissue-specific pattern of expression. These differential expression patterns have also been described in mouse tissues [10].

Promoter region of *GlcAT-I*

The 5'-flanking region of exon I of the *GlcAT-I* gene was sequenced to study the putative promoter region (Figure 5). The

-1200 CCGCGCCAGCTGCTAACAGCACATTTTAAAGTTTCCCATTTGTTTGTAGTACAGGAAGTA
TATA SRY Ets

-1140 TACTGTTCCCTTCCCTCATTTTGTAGACTGACACAGTCAAGATATTTTACCAG
MZF-1 MZF-1 AML-1a

-1080 AAACTTTGTGGGGTATATTTTGTGAGACGGTCTCACCCAGTACCCAGCTGG
Lyf-1 AP-1

-1020 AGTGCAGTGGCATGATCAGGCTCACACAGCCTCAGCCTCCCATGTGAGCTCCCAAGT
AML-1a Lyf-1

-960 AGCTGGGACTGTAGCCCGCGGCATCACACCCAGCTAATTTTCTCTTTTGTAGAG
GATA-1 SRY

-900 ACGGGTTTGTGCTTGTGCCAGGCTGATCTGAAACTCTGGGCTTCTAAAGTCCAGG
Nkx-2.5

-840 ATTACAGCATGAGCCACTGGGCGGGCCCTTTTCTTTTAAAGGAAAAATGAGAAG
TATA

-780 GATGCTTGTGCTTGTGAGGAGGAAGGGATGCATTCATGGAATCCCGAGCTCAGGAAC
c-Rel MZF-1

-720 TATTTCCAATCAATTCTCTGGTAGTTAAACACTTCCTTGGCACCCCTGTTGTCATCTGT
CDP CR3+HD Ets

-660 TGAAGAAATGTGGGCCAGAGGACACACATAGAGATGCTGGAATCGGGTGTCCAGAAAGGAC

-600 ACAAACCTGCTGTGGCAATGACATAATTTTGTCAACTTTGGTTTCTTATCTATCTCA
Nkx-2.5 GATA-1

-540 AAGTATTATAGTGAGGATTCAGGAGAGACACCCAGTGGATCATGTTGGTGGGCTTAGAT
AML-1a

-480 TTCATTCGTGTAATAAGTATTGAGCAACAGCTATTGGGTGGCTGACTGACTGGATGGAT
Oct-1 C/EBP

-420 GAATAAGCTCAGATCTGGGCGCCAGGACTGGCTTCTCAGGCAATTCCTGATGAGGGG
deltaEF1 GATA-1

-360 GCGGGTCTCGAGGTAGGTACACAGTGCACCTACTCGAGATTCTGCACCTACCCGGGTTT
Sp1 USF GATA-1

-300 TGACTGCCGCCACCGAGTTTGGCACGCGGATGTGACGTACACCCCGGTGGTTCCCG
Ets CREB STAT

-240 GAACCTGGCGGTGGCGCGGAAATCAAGTCTCAAGCGGGAGGGTGCACAGATGACCGG
Ets Ets

-180 AAGCTGCGCTGGCGCGCTGCCAGCCCCAGCTGGGAGCTGTAGGGGATGCGAGGAAG
p300

-120 GCGGGACGACGGGCGGAGCCGCTGCAGGGCGGGCTGCGGACAGGAGCCGGGTTTGG
Sp1

-60 GCGGGGAAACCCCTCGTCCCTGACAGCCCGCTGCTCGGGCGGGCGGGCGGCGCGGCG
Sp1 Sp1

+1 ATGAGCTGAAGCTGAAGAACGTTTCTCGCTACTTCTCGTGTGATGCGCGGCTC

Figure 5 Nucleotide sequences of the human *GlcAT-I* promoter

The two putative start sites of transcription are shown by bent arrows. The consensus binding sites for the transcription factors are underlined. The sequences are numbered relative to the translation initiation site ATG, which is boxed.

transcription initiation site in RNA from human placenta was determined by using the Cap site cDNA™ kit as described in the Experimental section. Two putative start sites were found and located at positions -37 and -29 (relative to A⁺TG) in *GlcAT-I* (Figure 5). The sum total (1.47 kb) of the revealed exon sequences was consistent with the size of the *GlcAT-I* mRNA, which was estimated to be 1.5 kb by Northern analysis (see Figure 4). Analysis of 1.2 kb of the 5'-flanking region revealed that the region has a subregion with high G+C content (70% in the 400 bp proximal region; see Figure 5) and four potential Sp1-binding sites, which are present in many constitutive gene promoters. Although the region lacks canonical TATA or CCAAT boxes, there are two potential TATA boxes (at -801 and -1177 bp in Figure 5) and one potential CCAAT box (at -464 bp in Figure 5) upstream of the GC-rich region. In addition, the 5'-flanking region contains many other well characterized promoter elements, as shown in Figure 5. These include response elements for SRY (sex-determining region Y gene product) [17], Ets [18], MZF-1 (myeloid zinc-finger protein) [19], AML-1a (acute myeloid leukaemia gene product) [20], Lyf-1

Figure 6 Functional analysis of the human *GlcAT-I* promoter

Luciferase expression assays with the *GlcAT-I* promoter construct containing a 1230 bp *GlcAT-I* promoter region (see Figure 5) were performed in four different cultured human cells. Firefly luciferase activity was assayed as described in the Experimental section and is expressed as a percentage of that produced by the pGL3-Control vector containing the SV40 promoter/enhancer. Values were normalized for transfection efficiency on the basis of the *Renilla* luciferase value recorded in the same cells. Values are normalized means for three separate assays, each conducted in quadruplicate. The S.D. of the quadruplicates and of separate assays was less than 10%.

(lymphocytes-specific factor) [21], AP-1 (activator protein 1) [22], GATA-1 [23], Nkx-2.5 (tinman homeodomain factor) [24], CDP CR3 + HD (cut-like homeodomain protein) [25], c-Rel [26], Oct-1 (octamer factor) [27], deltaEF1 (delta-crystallin enhancer-binding protein) [28], USF (upstream stimulatory factor) [29], N-Myc [30], CREB (cAMP-response-element-binding protein) [31], STAT (signal transducers and activators of transcription) [32] and p300 (E1A-associated 300 kDa protein) [33]. Three putative binding sites for products of the Ets family were located close to the transcription initiation sites because, in a number of promoters lacking a canonical TATA box, Ets motifs are reported to be located close to the transcription initiation sites [18].

Promoter analysis of *GlcAT-I*

Next we sought to determine whether the 5'-flanking region of *GlcAT-I* could act as a functional promoter in transient cell-transfection assays. For this purpose a 1230 bp genomic fragment containing the 1198 bp sequence upstream of the transcription initiation site and the 32 bp exon I 5' sequence was cloned upstream of a firefly luciferase reporter gene in a pGL3 vector, as described in the Experimental section. This construct was transfected into four different cultured human cells along with a *Renilla* luciferase reporter gene driven by a thymidine kinase promoter from herpes simplex virus to normalize for transfection efficiency. The pGL3-Control vector, in which the firefly luciferase expression is driven by the SV40 promoter/enhancer, was used as a positive control. As shown in Figure 6, in all four cell lines examined the *GlcAT-I* promoter displayed high and comparable activity to that of the SV40 promoter, demonstrating the promoter activity of the 5'-flanking region of *GlcAT-I*. In particular, the highest promoter activity was detected in G361 melanoma cells. Expression in HepG2 hepatoma cells was approximately one-third of that detected in G361 cells. These findings provide evidence for the differential regulation of the gene expression as observed above (see Figure 4).

Further experiments with the most responsive G361 melanoma cells were performed to investigate the functional role of the *cis*-acting elements. For this purpose, stepwise deletions of the

Figure 7 Deletion analysis of the human *GlcAT-I* promoter

Schematic representation of the 5' stepwise deletion constructs used to test the functional activity of the *GlcAT-I* promoter in transient transfection assays; the relative luciferase activity of each construct is shown. Numbers to the left of each construct indicate the 5' end of the promoter fragment relative to the translation initiation site of *GlcAT-I* (see Figure 5). Firefly luciferase activity was assayed as described in the Experimental section and is expressed as a percentage of that produced by the *GlcAT-I* promoter construct containing a 1230 bp *GlcAT-I* promoter region. Values were normalized for transfection efficiency on the basis of the *Renilla* luciferase value recorded in the same cells. Values are normalized means $\pm$ S.D. for four different experiments.

GlcAT-I promoter were generated (Figure 7). The highest level of promoter activity was observed with the construct containing the sequence from -303 to $+32$ (Figure 7). Shortening of the upstream sequence from -1198 to -304 slightly increased the promoter activity, suggesting that this region might contain negative regulatory sequences. Remarkably, deletion of the sequence from -303 to -154 , which harbours one CREB, one STAT and three Ets motifs, caused a sharp decrease in functional promoter activity. The relative proximity of these three *cis*-acting elements suggests that this region might act as a functional domain.

DISCUSSION

Here we present the isolation and characterization of a functional human *GlcAT-I* and a *GlcAT-I*-related pseudogene. The findings obtained indicated that the human genome contains two related *GlcAT-I* genes, at chromosomes 11q12–q13 and 3. Only the 11q12–q13 locus is a functional human *GlcAT-I*; the other is non-functional and seems to be a processed pseudogene that remains permanently silent.

The functional *GlcAT-I* spans 7 kb and consists of five exons. GlcAT-I, like other Golgi-resident glycosyltransferases, is composed of a short N-terminal cytoplasmic domain, a transmembrane domain, a stem region and a large catalytic domain [6]. Exon I encodes the nucleotide sequence corresponding to both the cytoplasmic and transmembrane domains; exon II and exons III–V mainly encode the nucleotide sequences corresponding to the stem region and the large catalytic domain respectively. We recently determined the crystal structure of human GlcAT-I in the presence of the donor substrate product UDP, the catalytic Mn^{2+} ion and the acceptor substrate analogue Gal β 1-3Gal β 1-4Xyl [34]. The findings revealed that the enzyme is an α/β protein with two subdomains that constitute the donor and acceptor substrate-binding sites and that the active site residues lie in a cleft extending across both subdomains. From both this observation and the present analysis of the intron/exon structure of *GlcAT-I*, it became apparent that one subdomain containing the donor substrate-binding site was encoded by exon III, whereas the other subdomain containing the acceptor-

binding site and the catalytic base was encoded by exons IV and V.

As described previously [6], the amino acid sequence of GlcAT-I exhibits 42.7% identity with rat GlcAT-P involved in the biosynthesis of the carbohydrate epitope HNK-1; the highest sequence identity was found in the C-terminal catalytic domain containing the four previously identified highly conserved motifs (I–IV) [7]. It should be noted that each of the highly conserved motifs I, II and IV is divided into two exons in *GlcAT-I*. However, it is not surprising that a molecular similarity exists between the two enzymes, although the product glycan types are different (the GAG linkage region and the HNK-1 epitope on N-linked glycoproteins), because the two enzymes exhibit a certain similarity in their substrate recognition (both enzymes recognize the terminal galactose moiety of their acceptor substrates). Very recently, cDNA cloning and chromosomal mapping of human *GlcAT-P* have been reported [35]. Human GlcAT-P is highly similar (98.2% identity) to the rat counterpart; human *GlcAT-P* was mapped to chromosome 11q25. Therefore, although the genomic structure of the *GlcAT-P* gene has not yet been reported, the two human β 1,3-glucuronyltransferases GlcAT-I and GlcAT-P show molecular similarity and these genes are located on the same chromosome, 11q, suggesting that both genes might be evolutionarily related and originate from one primordial gene.

Interestingly, human *GlcAT-I* is located downstream of human *ROM1* encoding retinal outer segment membrane protein 1 [16] and separated by an intergenic region of only 194 bp. Thus the two genes are contiguous in the human genome in a tail-to-tail manner. Rom-1 is reported to be required for mammalian rod photoreceptor viability and the regulation of disk morphogenesis; it has been suggested that mutations in human *ROM1* might cause recessive photoreceptor degeneration [36]. In addition, these two genes are also contiguous in the mouse genome (Y. Tone, H. Kitagawa and K. Sugahara, unpublished work). Although the two genes seem to be functionally unrelated and are transcribed independently, they might be also contiguous in other mammalian genomes.

In addition to the functional gene, the human genome contained another homologous *GlcAT-I* gene corresponding to a processed pseudogene. Analysis of this pseudogene sequence revealed several interesting features, which are common to many previously characterized processed pseudogenes (reviewed in [37]). The pseudogene lacks completely the intervening sequences found in the functional gene. The existence of multiple nucleotide changes and a deletion event in the putative coding region indicate clearly that this gene is not functional. The pseudogene contains an insignificantly sized open reading frame, as it is capable of encoding only a 45-residue polypeptide. Although the *GlcAT-I* pseudogene contains a 3 bp insertion in the putative 5' untranslated region and has a 10 bp deletion in the putative coding region, it has retained high sequence identity (95.3%) with human *GlcAT-I* cDNA, suggesting that the pseudogene was formed recently in evolutionary time, perhaps during the descent of *Homo sapiens*. In fact, there is no *GlcAT-I* pseudogene in the mouse genome (Y. Tone, H. Kitagawa and K. Sugahara, unpublished work). In addition, the sequence similarity between the pseudogene and the functional gene ceases at the points corresponding to the beginning and the end of the transcript from the functional gene (see Figure 2). Furthermore, the pseudogene has a poly(A) sequence immediately 3' of the point at which the similarity between the pseudogene and the functional gene ceases. Lastly, the chromosomal locations of the human *GlcAT-I* pseudogene and the corresponding functional gene are not syntenic, in that the pseudogene is localized to chromosome 3 and the functional gene to chromosome 11q12–q13. All these

features clearly indicate that the pseudogene was derived from the mature *GlcAT-I* mRNA and is a processed pseudogene.

As shown in the present study, the *GlcAT-I* promoter was differentially active in the cell lines tested, exhibiting 3-fold higher activity in G361 cells, a melanoma cell line that expresses this gene highly, compared with HepG2 cells, a hepatoma cell line that expresses this gene moderately (Y. Tone, H. Kitagawa and K. Sugahara, unpublished work). These findings seem to support the differential regulation of *GlcAT-I* expression. In addition, the activity of the *GlcAT-I* promoter was comparable to that of the SV40 promoter. In this regard it should be noted that transfection of *GlcAT-I* cDNA into COS cells induced the expression of the HNK-1 epitope when the vectors were used under the control of a strong promoter, although *GlcAT-I* shows strict specificity for the GAG–protein linkage region trisaccharide Gal β 1-3Gal β 1-4Xyl, exhibiting negligible incorporation into other galactoside substrates [9,10]. It is therefore possible that in tissues that strongly express the *GlcAT-I* gene the HNK-1 epitope GlcA(3-*O*-sulphate) β 1-3Gal β 1-4GlcNAc is synthesized, because the HNK-1 3-*O*-sulphotransferase is widely distributed in various tissues [38,39]. In fact, G361 cells highly express the HNK-1 epitope (M. Taoka, H. Kitagawa and K. Sugahara, unpublished work). Hence, at certain specific developmental stages or under pathological conditions including cancers, *GlcAT-I* might participate in the synthesis of the HNK-1 epitope. Interestingly, in this context, a deletion analysis of the *GlcAT-I* promoter with G361 cells revealed that the minimum promoter region contained potential binding sites for tissue-specific transcription factors such as STAT, CREB or Ets. In particular, the Ets family of transcription factors is reported to be related to transformation, signal transduction and cell differentiation (reviewed in [40]). Moreover, Bai et al. [8] reported that transfection of Chinese hamster ovary cell mutants defective in *GlcAT-I* with *GlcAT-I* cDNA augmented GAG synthesis to levels approximately double those in wild-type cells, implying that *GlcAT-I* might be rate-limiting. It is therefore of particular importance to elucidate the tissue-specific regulatory mechanism of *GlcAT-I* expression. More detailed functional analysis of the promoter is now in progress.

This work was supported in part by the Science Research Promotion Fund of the Japan Private School Promotion Foundation (to K.S.), grants-in-aid for Encouragement of Young Scientists 11771460 (to H.K.) and for Scientific Research on Priority Areas 10178102 (to K.S.) from the Ministry of Education, Science, Culture, and Sports of Japan.

REFERENCES

- Kimata, K., Okayama, M., Oohira, A. and Suzuki, S. (1973) Cytodifferentiation and proteoglycan biosynthesis. *Mol. Cell. Biochem.* **1**, 211–228
- Perrimon, N. and Bernfield, M. (2000) Specificities of heparan sulphate proteoglycans in developmental processes. *Nature (London)* **404**, 725–728
- Sugahara, K. and Kitagawa, H. (2000) Recent advances in the study of the biosynthesis and functions of sulphated glycosaminoglycans. *Curr. Opin. Struct. Biol.* **10**, 518–527
- Salmivirta, M., Lidholt, K. and Lindahl, U. (1996) Heparan sulfate: a piece of information. *FASEB J.* **10**, 1270–1279
- Rodén, L. (1980) Structure and metabolism of connective tissue proteoglycans. In *The Biochemistry of Glycoproteins and Proteoglycans* (Lennarz, W. J., ed.), pp. 491–517, Plenum Publishing, New York
- Kitagawa, H., Tone, Y., Tamura, J., Neumann, K., Ogawa, T., Oka, S., Kawasaki, T. and Sugahara, K. (1998) Molecular cloning and expression of glucuronyltransferase I involved in the biosynthesis of the glycosaminoglycan–protein linkage region of proteoglycans. *J. Biol. Chem.* **273**, 6615–6618
- 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. U.S.A.* **94**, 6093–6098
- Bai, X., Wei, G., Sinha, A. and Esko, J. D. (1999) Chinese hamster ovary cell mutants defective in glycosaminoglycan assembly and glucuronosyltransferase I. *J. Biol. Chem.* **274**, 13017–13024
- Tone, Y., Kitagawa, H., Imiya, K., Oka, S., Kawasaki, T. and Sugahara, K. (1999) Characterization of recombinant human glucuronyltransferase I involved in the biosynthesis of the glycosaminoglycan–protein linkage region of proteoglycans. *FEBS Lett.* **459**, 415–420
- Wei, G., Bai, X., Sarkar, A. K. and Esko, J. D. (1999) Formation of HNK-1 determinants and the glycosaminoglycan tetrasaccharide linkage region by UDP-GlcUA:galactose β 1,3-glucuronosyltransferases. *J. Biol. Chem.* **274**, 7857–7864
- Kitagawa, H., Kano, Y., Shimakawa, H., Goto, F., Ogawa, T., Okabe, H. and Sugahara, K. (1999) Identification and characterization of a novel UDP-GalNAc:GlcA β -R α 1,4-N-acetylgalactosaminyltransferase from a human sarcoma cell line. *Glycobiology* **9**, 967–703
- Maruyama, K. and Sugano, S. (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138**, 171–174
- Heng, H. H., Squire, J. and Tsui, L.-C. (1992) High-resolution mapping of mammalian genes by in situ hybridization to free chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9509–9513
- Heng, H. H. and Tsui, L.-C. (1993) Modes of DAPI banding and simultaneous in situ hybridization. *Chromosoma* **102**, 325–332
- Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* **50**, 349–383
- Bascom, R. A., Schappert, K. and McInnes, R. R. (1993) Cloning of the human and murine ROM1 genes: genomic organization and sequence conservation. *Hum. Mol. Genet.* **2**, 385–391
- Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R. and Bianchi, M. E. (1994) Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO J.* **13**, 6115–6124
- Wasyluk, B. E., Hahn, S. L. and Giovane, A. (1993) The Ets family of transcription factors. *Eur. J. Biochem.* **211**, 7–18
- Morris, J. F., Hromas, R. and Rauscher, F. J. (1994) Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Mol. Cell. Biol.* **14**, 1786–1795
- Meyer, S., Downing, J. R. and Hiebert, S. W. (1993) Identification of AML-1 and the (8;21) translocation protein (AML-1/ETO) as sequence-specific DNA-binding proteins: the runt homology domain is required for DNA binding and protein–protein interactions. *Mol. Cell. Biol.* **13**, 6336–6345
- Lo, K., Landau, N. R. and Smale, S. T. (1991) LyF-1, a transcriptional regulator that interacts with a novel class of promoters for lymphocyte-specific genes. *Mol. Cell. Biol.* **11**, 5229–5243
- Lee, W., Mitchell, P. and Tjian, R. (1987) Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. *Cell* **49**, 741–752
- Merika, M. and Orkin, S. H. (1993) DNA-binding specificity of GATA family transcription factors. *Mol. Cell. Biol.* **13**, 3999–4010
- Chen, C. Y. and Schwartz, R. J. (1995) Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, *nkx-2.5*. *J. Biol. Chem.* **270**, 15628–15633
- Harada, R., Berub, 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
- Kunsch, C., Ruben, S. M. and Rosen, C. A. (1992) Selection of optimal κ B/Rel DNA-binding motifs: interaction of both subunits of NF- κ B with DNA is required for transcriptional activation. *Mol. Cell. Biol.* **12**, 4412–4421
- Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., van Leeuwen, H. C., Strating, M. J. J. and van der Vliet, P. C. (1992) The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J.* **11**, 4993–5003
- Sekido, R., Murai, K., Funahashi, J.-I., Kamachi, Y., Fujisawa-Sehara, A., Nabeshima, T.-I. and Kondoh, H. (1994) The delta-crystallin enhancer-binding protein delta EF1 is a repressor of E2-box-mediated gene activation. *Mol. Cell. Biol.* **14**, 5692–5700
- Sawadogo, M., Van Dyke, M. W., Gregor, P. D. and Roeder, R. G. (1988) Multiple forms of the human gene-specific transcription factor USF. I. Complete purification and identification of USF from HeLa cell nuclei. *J. Biol. Chem.* **263**, 11985–11993
- Alex, R., Soezeri, O., Meyer, S. and Dildrop, R. (1992) Determination of the DNA sequence recognized by the bHLH-zip domain of the N-Myc protein. *Nucleic Acids Res.* **20**, 2257–2263
- 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
- Horvath, C. M., Wen, Z. and Darnell, J. E. (1995) A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain. *Genes Dev.* **9**, 984–994

- 33 Rikitake, Y. and Moran, E. (1992) DNA-binding properties of the E1A-associated 300-kilodalton protein. *Mol. Cell. Biol.* **12**, 2826–2836
- 34 Pedersen, L. C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T. A. and Negishi, M. (2000) Heparan/chondroitin sulfate biosynthesis: structure and mechanism of human glucuronyltransferase I. *J. Biol. Chem.* **275**, 34580–34585
- 35 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
- 36 Clarke, G., Goldberg, A. F. X., Vidgen, D., Collins, L., Ploder, L., Schwarz, L., Molday, L. L., Rossant, J., Szél, Á., Molday, R. S. et al. (2000) Rom-1 is required for rod photoreceptor viability and the regulation of disk morphogenesis. *Nat. Genet.* **25**, 67–72
- 37 Vanin, E. F. (1985) Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* **19**, 253–272
- 38 Bakker, H., Friedmann, I., Oka, S., Kawasaki, T., Nifant'ev, N., Schachner, M. and Mantei, N. (1997) Expression cloning of a cDNA encoding a sulfotransferase involved in the biosynthesis of the HNK-1 carbohydrate epitope. *J. Biol. Chem.* **272**, 29942–29946
- 39 Ong, E., Yeh, J.-C., Ding, Y., Hindsgaul, O. and Fukuda, M. (1998) Expression cloning of a human sulfotransferase that directs the synthesis of the HNK-1 glycan on the neural cell adhesion molecule and glycolipids. *J. Biol. Chem.* **273**, 5190–5195
- 40 Dittmer, J. and Nordheim, A. (1998) Ets transcription factors and human disease. *Biochim. Biophys. Acta* **1377**, F1–F11

Received 6 March 2001/29 May 2001; accepted 16 July 2001