Cloning, expression and gene organization of a human Neu5Ac α 2–3Gal β 1–3GalNAc α 2,6-sialyltransferase: hST6GalNAc IV¹

Anne HARDUIN-LEPERS, David C. STOKES², Wim F. A. STEELANT, Bénédicte SAMYN-PETIT, Marie-Ange KRZEWINSKI-RECCHI, Veronica VALLEJO-RUIZ, Jean-Pierre ZANETTA, Claudine AUGÉ³ and Philippe DELANNOY⁴

Unité de Glycobiologie Structurale et Fonctionnelle, UMR CNRS n°8576, Laboratoire de Chimie Biologique, Université des Sciences et Technologies de Lille, F-59655 Villeneuve d'Ascq, France

On the basis of the detection of expressed sequence tag ('EST') similar to the rat *N*-acetylgalactosamine α 2,6-sialyltransferase (ST6GalNAc) III cDNA, we have identified a novel member of the human ST6GalNAc family. We have isolated a cDNA clone containing an open reading frame that codes for a type II membrane protein of 302 amino acids with a seven-amino-acid cytoplasmic domain, an 18-amino-acid transmembrane domain and the smallest described catalytic domain of 277 amino acids. This predicted sialyltransferase sequence is similar to the rat ST6GalNAc III (46.6%), but was found to be even more similar to the recently reported mouse ST6GalNAc IV (88.1%) on the basis of amino acid sequence identity. Northern-blot analysis showed that the newly identified gene is expressed constitutively in various adult human tissues as a 2.2 kb transcript, but was also found to be expressed at lower levels in brain, heart and skeletal muscle as a 2.5 kb transcript. Expression of the hST6GalNAc IV

gene was investigated by reverse transcription PCR in various human cancer cells, and was found to be present in the majority of cell types with the exception of the carcinoma cell line T47D and pro-monocyte THP cells. The transient expression in COS-7 cells of the full-length cDNA led to the production of an active enzyme sharing the acceptor specificity of the ST6GalNAc family towards Neu5Ac α 2–3Gal β 1–3GalNAc α -O-R (where 'R' denotes H, benzyl, or a peptidic chain). Detailed analysis *in vitro* of substrate specificity revealed that the enzyme required the trisaccharide Neu5Ac α 2–3Gal β 1–3GalNAc found on O-glycans and arylglycosides. In addition, we have clarified the genomic organization of ST6GalNAc IV gene.

Key words: O-glycosylation, molecular cloning, sialyltransferases.

INTRODUCTION

Sialylated glycoconjugates exhibit remarkably diverse structures [1b,2], and their expression has been shown to change during development, differentiation, diseases and oncogenic transformation [3]. The structural diversity and regulated expression of sialylated glycoconjugates appear to correlate with their functions. Sialic acids are usually found at the non-reducing terminal position of the sugar chains of glycoconjugates, α 2,3- or α 2,6-linked to a β -D-galactopyranose residue, or in a lateral position α 2,6-linked to a GalNAc or a GlcNAc residue. These sialylated sugar chains are present at the cell surface in tissues of various animal species. Owing to this terminal position, they are thought to serve in a variety of biological functions, such as in cell–cell and cell–substrate interactions, bacterial adhesion and protein targeting [4,5]. As one of many examples, the Thomsen–Freidenreich-related antigens {including sialyl-Tn (Neu5Ac α 2–6GalNAc α -O-Ser), Tn (GalNAc α -O-Ser), T (Gal β 1–3GalNAc α -O-Ser) and sialyl6-T (Gal β 1–3[Neu5Ac α 2–6]GalNAc α -O-Ser)} constitute a group of important carbohydrate epitopes found on glycoconjugates. GalNAc α -O- and Gal β 1–3GalNAc α -O- are

ubiquitous core structures found in a cryptic manner in many O-glycosylproteins, including mucins. However, the expression of T and Tn carbohydrate epitopes is restricted to specific carrier proteins in specific cells or organs, and overexpression of T, Tn and sialyl-Tn antigens has been described in many carcinoma cells and tumours, including breast cancers [6]. The aberrant expression of the cancer-associated carbohydrate, sialyl-Tn, has been positively correlated with poor prognosis markers and with reduced overall survival of the patients [7]. The identification and characterization of sialyltransferase genes is a prerequisite both for elucidating the biological functions of the oligosaccharide epitopes and in understanding the regulatory mechanisms of their stage- and cell-type-specific expression at a molecular level.

Sialyltransferases are a family of more than 18 members that catalyse the transfer of sialic acid from CMP-Neu5Ac forming an α 2,3, an α 2,6 or an α 2,8 linkage, depending on the acceptor sugar chain. So far, the cloning of five members of the GalNAc α 2,6-sialyltransferase family (ST6GalNAc-I, -II, -III, -IV and -V) has been reported. The cDNA of ST6GalNAc I has been isolated from chick [8], mouse [9] and human [10]; ST6GalNAc II cDNA has been obtained from chick [11], mouse [12] and

Abbreviations used: Bz, benzyl; BSM, bovine submaxillary mucin; EST, expressed sequence tag; HTG, Human Throughput Genomic; mAb, monoclonal antibody; NBT, Nitro Blue Tetrazolium; ORF, open reading frame; OSM, ovine submaxillary mucin; poly(A)⁺, polyadenylated; 5'-RACE, 5'-rapid amplification of cDNA ends; RT, reverse transcription; ST6GalNAc/CMP-Neu5Ac, GalNAc- α 2,6-sialyltransferase family. The nomenclature of gangliosides is on the basis of that determined by the IUPAC-IUB Joint Commission on Biochemical Nomenclature [1]. The nomenclature of sialyltransferases follows the system of Tsuji et al. [1a].

¹ This paper is dedicated to the memory of Professor André Verbert, who showed a constant interest in this work.

² Present address: University of Cambridge, Department of Clinical Biochemistry, Wellcome Trust Centre for the Study of Molecular Mechanisms in Disease, Cambridge CB2 2XY, U.K.

³ Present address: Laboratoire de Chimie Organique Multifonctionnelle, UMR CNRS 8614, Université de Paris-Sud, Bât 420, F-91405 Orsay Cedex, France.

⁴ To whom correspondence should be addressed (e-mail pdelanno@chouia.univ-lille1.fr).

human [13]; ST6GalNAc III has been obtained from rat [14] and mouse [15]; ST6GalNAc IV has been isolated from mouse [15]; and ST6GalNAc V has been identified in mouse [16,17]. As far as can be determined from *in vitro* assays of the expressed recombinant enzymes, ST6GalNAc I exhibits the broadest specificity for the following structures: GalNAc α -O-Ser/Thr, Gal β 1-3GalNAc α -O-Ser/Thr and Neu5Ac α 2-3Gal β 1-3GalNAc α -O-Ser/Thr [8,10]. ST6GalNAc II exhibits a narrower substrate specificity, utilizing Gal β 1-3GalNAc α -O-Ser/Thr and Neu5Ac α 2-3Gal β 1-3GalNAc α -O-Ser/Thr as substrates [11-13]. ST6GalNAc III and ST6GalNAc IV exhibit the most restricted substrate specificity, only utilizing the trisaccharide sequence Neu5Ac α 2-3Gal β 1-3GalNAc, which is found on either O-glycosyl proteins or ganglioside G_{M1b}, suggesting that they do not discriminate between α - and β -linked GalNAc [14,15]. Mouse ST6GalNAc V, expressed only in the brain, seems to be specific for G_{M1b} [16,17].

As a first step towards the understanding of the molecular basis for these modifications of O-glycan biosynthesis observed in human cancer cells, we have sought after the intervening enzymes. Similarity searches in the human expressed sequence tag (EST) division of the GenBank® Database have revealed the existence of open reading frames (ORFs) encoding proteins that share conserved motifs with the sialyltransferases, the so-called sialyl 'motifs'. Among these ORFs, we have cloned the human orthologue of the mouse ST6GalNAc IV from HepG₂ cells that encodes a Neu5Ac α 2-3Gal β 1-3GalNAc (sialyl to GalNAc) α 2,6-sialyltransferase. Expression of the gene was analysed in various human tissues by Northern-blot analysis and in human cultured cancer cells by reverse transcription (RT)-PCR. When transiently transfected in COS-7 cells, this human full-length cDNA directed the expression of an active enzyme capable of transferring a sialic acid residue on to the GalNAc residue of the trisaccharide, Neu5Ac α 2-3Gal β 1-3GalNAc, found predominantly on O-glycans, but also on arylglycosides. Finally, the ST6GalNAc IV gene was found to be localized on human chromosome 9, band 9q31.

EXPERIMENTAL

Materials

CMP-[¹⁴C]Neu5Ac (10.5 GBq·mmol⁻¹), [α -³⁵S]dATP (37 GBq·mmol⁻¹) and [α -³²P]dCTP (110 TBq·mmol⁻¹) were purchased from Amersham Pharmacia Biotech. (Little Chalfont, Bucks., U.K.). Hi-Taq DNA polymerase was from Bioprobe Systems (Montreuil sur Seine, France). Dulbecco's modified Eagle's medium, Optimem, LIPOFECTAMINE[®] and antibiotics used in cell culture were obtained from Gibco BRL (Cergy-Pontoise, France). Oligonucleotides were synthesized by Gibco BRL, Eurogentec (Seraing, Belgium) or Cybergène (Evry, France). Restriction endonucleases were from Eurogentec. α -Acid glycoprotein, asialo-G_{M1}, G_{M1a}, G_{D1a}, bovine submaxillary mucin (BSM), CMP-Neu5Ac, CMP-agarose, fetuin, GalNAc α 1-O-benzyl (-Bz), GlcNAc α 1-O-Bz and Gal β 1-3GalNAc α 1-O-Bz were from Sigma (St Louis, MO, U.S.A.). Gal β 1-3GlcNAc β -O-Bz and Gal β 1-4GlcNAc β -O-Bz were synthesized by the method of Augé and Veyrières [18]; Neu5Ac α 2-3Gal β 1-3GlcNAc, Neu5Gc α 2-3Gal β 1-3GalNAc, Neu5Ac α 2-3Gal β 1-3GlcNAc β -O-p-methoxy-Bz and Neu5Gc α 2-3Gal β 1-3GalNAc β -O-Bz were prepared by the method of Lubineau et al. [19,20]. Neu5Ac α 2-3Gal β 1-3GalNAc α -O-Bz was chemoenzymically synthesized as described in [21]. The expression vector pFLAG-CMV-1 and monoclonal antibody (mAb) BioM2 were from Sigma. Ovine submaxillary mucin (OSM) was generously given by Dr J. C. Michalski (UMR CNRS 8576, Villeneuve d'Ascq,

France), and ganglioside mixtures were prepared from rat brain [22].

Lacto-N-tetraose and Neu5Ac α 2-3Lacto-N-tetraose were prepared from human milk. Asialo-glycoproteins were prepared as described previously [13]. Briefly, BSM, fetuin, OSM or α -acid glycoprotein (10 mg·ml⁻¹) were incubated for 1 h in 0.05 M sulphuric acid at 80 °C. The asialo-products were then neutralized, dialysed, freeze-dried and checked for desialylation by MS before use.

G_{M1b} was generated using recombinant soluble ST3Gal I as follows: 200 nmol of asialo-G_{M1} and 0.3 % (v/v) Triton CF-54 were dried under nitrogen, and then resuspended in 200 μ l of 20 mM cacodylate buffer, pH 6.2, 2 mM CMP-Neu5Ac (9.17 kBq) and 80 μ l of ST3Gal I-transfected COS-7 cell culture medium. The reaction was allowed to proceed for 48 h, the mixture was passed over a C₁₈ Sep-Pak cartridge (Waters Millipore Corp, Milford, MA, U.S.A.), and the G_{M1b} and asialo-G_{M1} products were eluted from the column with 2 ml of methanol and 2 ml of chloroform/methanol (1:1, v/v). The ganglioside mixture was resolved by HPTLC and orcinol staining [22].

Cell culture, isolation of RNA and cDNA synthesis

The cells used in the present study and the cell culture conditions were the same as those described in [13]. Total RNA was extracted using the guanidium thiocyanate method described by Sambrook et al. [23], and cellular RNA was quantified by spectrophotometry at 260 nm.

For subsequent PCR amplifications, first-strand cDNA was synthesized from total RNA using the First Strand cDNA Synthesis kit (Amersham Pharmacia Biotech, Piscataway, NJ, U.S.A.) according to the manufacturer's protocol.

Isolation of ST6GalNAc IV cDNA and homology searching

Human ESTs (GenBank® accession numbers AA864984, AA557535 and AA830319) with high similarities to rat ST6GalNAc III (GenBank® accession number L29554) and the human gene were identified using the BLASTN and TBLASTN algorithms (version 2.0.8 and 2.0.11) against the human EST division (dbEST) and the human High Throughput Genomic (HTG) Sequences division of the public GenBank®/EMBL databases at the National Centre for Biotechnology Information [24]. All these EST cDNAs contained one or two of the sialyl 'motifs' described for all mammalian sialyltransferases.

Amino acid sequence and hydropathy analyses were performed using the Internet programs Translation-2 (Infobiogen, Villejuif, France) and TMPred (Expasy, Swiss Institute of Bioinformatics, Switzerland). Determination of potential O-glycosylation sites was performed using the Internet program NetOGlyc 2.0 (Expasy, Switzerland).

To isolate cDNA clones, the RT-PCR method was performed using 150 ng of first-strand cDNA synthesized from the total RNA of HepG₂ cells. Two sense primers containing an *Eco*RI site, 5'-CGCGAGGAATTCGCGAGCTGTGCCGTGGTG-3' (designed in the sialyl 'motif' L) and 5'-CAGTCGGAATTCTTCCTCAGCACCGGCTG-3' (designed in the sialyl 'motif' S) and one antisense primer also containing an *Eco*RI site, 5'-GACCGCCTTCTCAGTCTTAAGCGGTG-3' (designed in the sialyl 'motif' VS), were constructed (here, and subsequently, the restriction enzyme sites are underlined). Each primer (2 μ M) was combined with 0.625 unit of Hi-Taq polymerase and 150 μ M dNTPs. PCR was performed as follows: 35 cycles at 95 °C for 40 s, 54 °C for 1 min and 72 °C for 1 min, and 72 °C for

10 min. The RT-amplified fragments (640 bp and 230 bp) were subcloned into Bluescript SKII *Eco*RI-cut vector, and their sequences were determined by the dideoxy termination method [25].

5'-Amplification of cDNA ends (5'-RACE)

Amplification of the 5'-end of human ST6GalNAc IV was achieved with the 5'-RACE kit (Gibco BRL, Life Technologies) according to the manufacturer's instructions. cDNA was reverse-transcribed from 1 μ g of HepG₂ total RNA with 2.5 pmol of the antisense oligonucleotide 5'-GGCGGCTGTCTAGACGACGGAAGCTACTCAGTCCTC-3' (complementary to the hST6GalNAc IV coding strand, nt 897–932) and 200 units of superscript II reverse transcriptase, and then was C-tailed. Two consecutive PCR were performed with 150 μ M dNTPs, 1.5 units of Hi-Taq polymerase and two nested sets of primers. Oligonucleotide sense Abridged Anchor Primer, 5'-GGCCACGCGTCTGACTAGTAC-3', and the specific antisense oligonucleotide 5'-CCTCCAGGACGGATGGCGAACACGATGGGCCTCTT-3' (nt 865–900) were used in a first PCR of 95 °C for 2 min, followed by 40 cycles (95 °C for 45 s, 72 °C for 3 min) and by an extension step of 10 min at 72 °C. The oligonucleotide sense Abridged Universal Amplification Primer, 5'-GGCCACGCGTCTGACTAGTAC-3', and the specific antisense oligonucleotide 5'-TGATGAGCACGAGCCGACCC-3' (nt 12–31) were used in a second PCR of 95 °C for 5 min, 40 cycles (95 °C for 45 s, 72 °C for 3 min) and a final elongation step at 72 °C for 10 min. Amplification products were analysed on a 1% (w/v) agarose gel and sequenced.

Design of expression constructs of hST6GalNAc IV

A truncated form of hST6GalNAc IV lacking the first 36 amino acids of the ORF was prepared by PCR amplification using a 5'-primer containing an *Eco*RI site, 5'-CTCTGCCTGAATTCCTGCCTGGACCACCACTTC-3' (nt 93–126), and a 3'-primer containing a *Bam*HI site, 5'-TGCGAGGACGACGGATCC-TACTCAGTCCTCC-3' (nt 896–926). PCR amplification conditions were as follows: 95 °C for 1 min, 35 cycles of 40 s at 95 °C, 40 s at 67 °C and 1 min at 72 °C, followed by 10 min at 72 °C. The resulting 850 bp amplified fragment was digested with *Eco*RI and *Bam*HI, and inserted into the corresponding restriction sites of the pFLAG expression vector pFLAG-CMV-1, which contains the preprotrypsin leader sequence. This resulted in the transient expression of a secretory N-terminal FLAG fusion protein in COS-7 cells. However, the soluble form possessed very weak sialyltransferase activity. Therefore a PCR fragment encoding the ST6GalNAc coding region in its entirety was prepared by PCR using a 5'-primer sense oligonucleotide containing a *Hind*III site (5'-AGGGACAGCAAGCTTATGAGGCTCCGGGTGCGCTC-3'; nt –15–21) and a 3'-primer containing a *Xba*I site 5'-GGCGGCTGTCTAGACGACGGAAGCTACTCAGTCCTC-3' (nt 897–932). A 1 kb PCR fragment was obtained under the following conditions: 1 min at 95 °C and 35 cycles (95 °C for 45 s and 72 °C for 90 s), with an extension step of 10 min at 72 °C. This PCR fragment was subsequently subcloned into the *Hind*III/*Xba*I sites of pFLAG-CMV-1 vector, which resulted in the transient expression of an N-terminal FLAG fusion protein in COS-7 cells. All PCR-derived sequences and cloning sites were completely re-sequenced to verify that no random PCR-induced mutation or frame-shift artefacts existed in the expression constructs.

A sample (10 μ g) of each Qiagen-purified plasmid, in addition to 10 μ g of a β -galactosidase plasmid used as a positive control

of the transfection, was transiently transfected into COS-7 cells on a 100 mm² tissue-culture dish using LIPOFECTAMINE[®] Plus reagent, following the manufacturer's instructions. The transfected cells were harvested 48 h after transfection by scraping into PBS, and were pelleted by low-speed centrifugation. The sought-after protein was found to be expressed within the cells and in the culture media that was thus collected, ultrafiltered 10-fold against 20 mM Mes, pH 6.2/50 mM NaCl/1 mM dithioerythritol, and used as the enzyme source.

Immunofluorescence microscopy

For detection of hST6GalNAc IV, 6×10^4 COS-7 cells plated on glass coverslips in 24-well plates were transfected as described above. After transfection (48 h), cells were washed three times with PBS and fixed in 3% (v/v) *p*-formaldehyde in PBS at room temperature for 20 min. For confocal laser scanning using a Leica DM IRBE fluorescence microscope and a Leica TCS NT confocal unit, cells were incubated in 50 mM NH₄Cl in PBS and permeabilized with 0.2% (v/v) Triton X-100 for 5 min. Fixed cells were washed three times with PBS, and incubated for 30 min at room temperature in blocking buffer [3% (w/v) BSA in Tris-buffered saline]. Cells were then incubated for 60 min at room temperature in Tris-buffered saline containing a 1:200 dilution of the anti-FLAG BioM2 mAb (Eastman Kodak Company, Rochester, NY, U.S.A.). After washing in Tris-buffered saline, cells were incubated with Texas Red conjugated secondary antibody sheep anti-mouse (dilution 1:50) from Amersham Pharmacia Biotech. Cells were thoroughly washed, and the coverslips were mounted in Glycergel before observation.

Sialyltransferase assays

Sialyltransferase assays were performed as described previously [13]. In brief, enzyme activity was measured in 0.1 M cacodylate buffer, pH 6.2, 10 mM MnCl₂, 0.2% Triton CF-54, 50 μ M CMP-[¹⁴C]Neu5Ac (1.85 kBq), with one acceptor substrate and 20 μ l of the enzyme source in a final volume of 50 μ l. As acceptor substrate, 2 mg·ml⁻¹ glycoproteins, or 1 mM arylglycosides or free oligosaccharides, was used and the enzyme reaction was performed at 33 °C for 3 h. Reaction products were separated from CMP-[¹⁴C]Neu5Ac, depending on the acceptor substrate. For glycoproteins, the reaction was terminated by precipitation with 1% (w/v) phosphotungstic acid in 0.5 M HCl, followed by filtration on glass microfibre filters (GF/A; Whatman Biosystems Ltd., Maidstone, Kent, U.K.). For arylglycosides, the reaction was stopped with 450 μ l of H₂O and heating at 100 °C for 5 min. Products were applied on to a C₁₈ Sep-Pak cartridge (Waters Millipore Corp.) and eluted with 30% (v/v) acetonitrile in water. For glycolipids, reactions were stopped by adding 1 ml of methanol, samples were centrifuged at 3000 *g* for 5 min, diluted with 1 ml of water and applied on to a Sep-Pak C₁₈ cartridge. Glycolipids were eluted with 5 ml of CHCl₃/CH₃OH (1:1, v/v). For free oligosaccharides, the reaction mixture was heated at 100 °C for 5 min and then centrifuged. The supernatant was subjected to Whatman 3 paper chromatography, and then developed with pyridine/ethyl acetate/acetic acid/water (5/5/1/3, by vol.) for 26 h. For kinetic analysis, incubations were performed using various concentrations of acceptor substrates: 5–500 μ M CMP-Neu5Ac, or 0–5 mM Neu5Ac α 2-3Gal β 1-3GalNAc-*O*-Bz or Neu5Ac α 2-3Gal β 1-3GalNAc. Concentrations of fetuin and asialofetuin were calculated on the basis of the number of O-glycosidically linked Gal β 1-3GalNAc residues determined by gas chromatography (61 nmol·mg⁻¹).

Linkage analysis and identification of sialylated products

For linkage analysis of sialic acids, 1 mM Neu5Ac α 2-3Gal β 1-3GalNAc-*O*-Bz sialylated with soluble hST6GalNAc IV and 50 μ M CMP-[14 C]Neu5Ac (1.83 kBq) was purified on a Sep-Pak C $_{18}$ cartridge. The disialylated product was analysed by TLC (Silicagel 60) in methyl acetate/chloroform/methanol/1-propanol/0.25 % aqueous KCl (25/20/20/20/17.5, by vol.) [22]. Radiolabelled disialylated product formed was detected by autoradiography, whereas the unlabelled arylglycoside mixture was revealed by orcinol/sulphuric acid staining [26].

Western-blot analysis

COS-7 cells transfected with the expression plasmid pFLAG/hST6GalNAc IV and mock-transfected cells were lysed in 1 % (v/v) Triton X-100, 0.5 mM dithiothreitol, 5 mM MnCl $_2$ and 20 % (v/v) glycerol, and the corresponding media were dialysed against 20 mM Mes buffer, pH 7.0, 1 mM dithioerythritol, 50 mM NaCl. Cell extracts and dialysed media were incubated overnight with gentle stirring with CMP-agarose beads (1 μ mol $\cdot$ ml $^{-1}$ CMP). After washing, the supernatants were discarded, and beads were boiled for 5 min in SDS/PAGE loading buffer, centrifuged and then loaded on polyacrylamide gels (4–20 % gradient) under reducing conditions. After Western blotting, the nitrocellulose membrane was incubated with 10 μ g $\cdot$ ml $^{-1}$ anti-FLAG BioM2 mAb according to the manufacturer's instructions. Alkaline-phosphatase-labelled anti-mouse IgG (BioSys, Compiègne, France) was used as a secondary antibody, and proteins were revealed by Nitro Blue Tetrazolium (NBT)/5-bromo-4-chloroindol-3-yl phosphate staining (Boehringer Mannheim, Mannheim, Germany).

Northern-blot analysis

The human ST6GalNAc IV mRNAs were detected by using Northern-blot analysis with commercially available human multiple-tissue polyadenylated [poly(A) $^{+}$] RNA blot (Clontech, Palo Alto, CA, U.S.A.). A BamHI/EcoRI 830 bp fragment from the human ST6GalNAc IV cDNA was labelled with [α - 32 P]dCTP by random priming (2 $\times$ 10 4 c.p.m./ μ l) and hybridized to the poly(A) $^{+}$ RNA blot. A 1.8 kb human β -actin cDNA (Clontech; 5 $\times$ 10 4 c.p.m./ μ l) was used as a positive control. The membrane was washed three times (10 min each time) at room temperature with 2 $\times$ SSC (where 1 $\times$ SSC is 0.15 M NaCl/0.015 M sodium citrate)/0.05 % (w/v) SDS; twice (20 min each time) at 50 $^{\circ}$ C with 0.1 $\times$ SSC/0.1 % SDS, and the membrane was then exposed for 14 days between intensifying screens at -80° C.

Analysis of hST6GalNAc IV gene expression in various human cultured cells

Analysis of the expression of hST6GalNAc IV in various human cultured cells was conducted exactly as described previously [13].

RESULTS

Identification and isolation of human ST6GalNAc IV cDNA

The nucleotide sequence of rat ST6GalNAc III (GenBank $^{\circ}$ accession number L29554) was used to perform a BLAST analysis (Version 2.0.8; [24]) of the human EST division of the GenBank $^{\circ}$ /EMBL database. All the EST sequences showing high sequence similarities were conceptually translated, and three clones containing elements of the sialyl 'motifs' L, S or VS were treated as the putative orthologue of the rat enzyme. Three ESTs bearing the GenBank $^{\circ}$ accession numbers AA864984, AA557535 and R06989 were identified as being significantly

Figure 1 Schematic alignment of EST sequences with the rat ST6GalNAc III cDNA

ESTs that were similar to rat ST6GalNAc III cDNA (L29554) were retrieved using a BLAST search of the human EST division of the GenBank $^{\circ}$ /EMBL DataBank. The EST sequences were aligned with the corresponding rST6GalNAc III sequence, and those human sequences that overlapped and presented sialyl 'motifs' L, S or VS were assembled to form a larger potential ORF. The protein-coding region of rST6GalNAc III is represented by a rectangle on which the sialyl 'motifs' are shown. Sequence identification numbers are provided to the left of the lines representing the EST fragments.

similar to the query sequence. The retrieved sequences were aligned with the query sequence, and partially overlapping sequences were assembled to build a larger ORF (Figure 1) and were used to generate oligonucleotides for PCR. PCR was conducted from HepG2 cDNAs (see the Experimental section). The unique PCR-amplified fragment obtained (640 bp) was subcloned, sequenced in its entirety and found to be 60 % identical to the rat ST6GalNAc III sequence at the nucleotide level. Subsequently, this new human cDNA sequence has been used as a query to perform a BLAST analysis (release 56; June 22, 1998) of the mouse EST division of the GenBank $^{\circ}$ /EMBL database, and a 3635 bp cDNA sequence described as a putative sialyltransferase (accession number AJ007310) has been found. This mouse cDNA was highly similar (> 80 %) to our partial human sequence, and was used to probe back the human EST division of the GenBank $^{\circ}$ /EMBL database with the TBLASTN algorithm (release January 5, 1999). This procedure led to the identification of numerous other ESTs (235 BLAST hits), and allowed us to reconstitute the entire ORF of a putative human ST6GalNAc sialyltransferase that was 85 % identical to the mouse sequence. Additional 5'-RACE was performed in order to confirm the identity of this sequence as shown in Figure 1. While this work was in progress, two new mouse ST6GalNAc cDNAs were cloned {ST6GalNAc III (accession number Y11342) and ST6GalNAc IV (accession number Y15779; [15])} and their enzymic activities were characterized. The latter of these corresponded to the mouse sequence AJ007310, described previously in the databanks. Recently, while searching the GenBank $^{\circ}$ /EMBL database for newly deposited ST6GalNAc sequences (release June 3, 1999), we found additional human cDNA sequences. Two cDNA sequences under the accession numbers Y17461 and Y17460 were found, corresponding to the last exon and part of the penultimate exon of a protein that had been named SIAT3C [27], and two unpublished cDNA sequences were deposited in the databases under the accession numbers AF127142 and AB035172. These results suggested that our human sequence belonged to the ST6GalNAc family and could be the human counterpart of the recently published mST6GalNAc IV protein.

Figure 3 Northern-blot analysis of hST6GalNAc IV in various human tissues (A), and RT-PCR analysis of the expression of hST6GalNAc IV in various human cancer lines (B)

(A) Commercially prepared Northern blots with 1 μ g of poly(A)⁺ RNA from various adult human tissues were probed (a) with a 1 kb full-length hST6GalNAc IV cDNA labelled at random with ³²P, as described in the Experimental section, or (b) the same filter was probed with a human β -actin cDNA control probe (1.8 kb). RNA size-marker bands are indicated on the left side of the blot. sk. muscle, skeletal muscle; p. bl. leukocyte, peripheral blood leukocyte. (B) Relative expression levels of hST6GalNAc IV mRNA were measured by RT-PCR, as described in the Experimental section, among various human cell lines. 1, HBL100; 2, MCF-7AZ; 3, MDA-MB-231; 4, BT-20; 5, T47D; 6, L131; 7, ECV; 8, HepG₂; 9, Caco-2; 10, HT-29; 11, THP; 12, HL-60; 13, Raji; 14, Jurkat; 15, Dami; 16, SKNSH; 17, CH8; 18, Kelly; 19, IC8; 20, T1C3; 21, keratinocytes.

using the complete ORF cDNA of hST6GalNAc IV as a probe. We have found this gene to be expressed at low levels as a 2.2 kb transcript in almost all human tissues examined, indicating a possible constitutive expression (Figure 3A). In addition to this constitutive expression, a 2.5 kb transcript was detected in heart, brain and skeletal muscle. Northern blot hybridization yielded only faint signals, and could not be used to examine precisely the specific expression of the hST6GalNAc IV gene in human cancer cell lines. Therefore the expression of the hST6GalNAc IV gene was also determined in several human cancer cell lines by RT-PCR. As shown in Figure 3(B), all of the breast cancer cell lines tested (HBL-100, MCF-7, MDA-MB-231, BT-20, with the notable exception of T47D cells), as well as the hepatocarcinoma HepG₂ cells, colon cancer cells (HT-29, Caco2), neuroblastoma SKNSH cells, embryonic cell lines (L131, ECV), pro-monocytic cell line HL-60, promegakaryocytic Dami cells and lymphocytes (Jurkat and Raji) highly expressed a hST6GalNAc IV mRNA. However, the expression of the hST6GalNAc IV gene was weaker in neuroblastoma Kelly cells, brain-derived CH8 cells and melanoma cells (IC8, T1C3, keratinocytes). The hST6GalNAc IV gene does not appear to be expressed at all in T47D cells or pro-monocytic THP cells.

Expression of a recombinant hST6GalNAc IV protein in COS-7

In order to verify conclusively that the cloned cDNA we had obtained from HepG₂ cells indeed represented human ST6GalNAc IV, and to facilitate functional analysis of the enzyme, it was desirable to produce a soluble form of the enzyme that

Figure 4 Immunofluorescence staining of COS-7/ST6GalNAc IV cells (A), and immunoblotting of ST6GalNAc IV protein from transfected cells and cell medium (B)

(A) COS-7 cells were transiently transfected with pFLAG/ST6GalNAc IV or pFLAG by means of LIPOFECTAMINE[®]. After transfection (48 h), cells were fixed with *p*-formaldehyde, trypsin-treated and stained with mAb BioM2, as described in the Experimental section. (a) Control COS-7 cells; (b) COS-7/ST6GalNAc IV cells. (B) Mock-transfected cells and pFLAG/hST6GalNAc IV-transfected cells (48 h after transfection), as well as their corresponding cell-culture medium, were incubated with CMP-agarose beads. The beads were washed, boiled and subjected to SDS/PAGE under reduced conditions and Western blotting using the Bio M2 anti-FLAG mAb. The position of the high-range pre-stained SDS/PAGE standards (Bio-Rad Laboratories, Hercules, CA, U.S.A.) are indicated (in kDa) on the left side of the Figure. Lane 1, pFLAG/hST6GalNAc IV-transfected cells; lane 2, mock-transfected cells; lanes 3 and 4, 30 or 50 μ l respectively from the pFLAG/hST6GalNAc IV cell-culture medium.

could be secreted from the cells. We therefore constructed a truncated cDNA lacking the first 36 amino acids of the N-terminus region (i.e. cytoplasmic, transmembrane and stem region) and N-terminally fused it to the FLAG octapeptide (Asp-Tyr-Lys-Asp-Asp-Asp-Lys). This truncated form of hST6GalNAc IV has been transiently transfected into COS-7 cells, but exhibited enzyme activity that was too low (10 pmol/h per ml) to determine kinetic parameters. Therefore the full-length ST6GalNAc IV was subcloned into the pFLAG expression vector and transiently expressed in COS-7. The presence of the recombinant protein FLAG-hST6GalNAc IV within the cell was then examined by immunofluorescence (Figure 4A), and hST6GalNAc IV was found to be expressed within transfected cells. Recombinant FLAG-hST6GalNAc was also detected in the cell-culture medium after incubation with CMP-agarose beads, SDS/PAGE and Western blotting using the BioM2 anti-FLAG mAb. A band of approx. 35 kDa was observed, corresponding to the expected molecular mass of the polypeptide (34.9 kDa) (Figure 4B). Since COS-7 cell lysates intrinsically exhibit strong

Table 1 Comparison of the acceptor substrate specificity of human ST6GalNAc IV with the mouse enzyme and human ST6GalNAc II

Various acceptor substrates were incubated in the standard assay mixture using transfected concentrated COS-7 cell medium as an enzyme source. Each substrate was used at a concentration of 1 mM for arylglycosides and 2 mg · ml⁻¹ for glycoproteins. The specific [¹⁴C]Neu5Ac incorporation resulting from recombinant hST6GalNAc IV activity was estimated by subtracting the background incorporation into endogenous substrate and the background sialyltransferase activities of mock-transfected cells from the radioactivity in the presence of exogenous acceptor. Relative rates are calculated as a percentage of the incorporation of sialic acid on to Gal β 1–3GalNAc α 1-*O*-Ser/Thr found in asialofetuin. A value of 0 indicates < 1%. nd, not determined; -R, the remainder of the N-linked oligosaccharide chain.

Acceptor	Structures†	Relative rate (%)*		
		hST6GalNAc IV	mST6GalNAc IV‡	hST6GalNAc II§
Glycoproteins				
Fetuin	Neu5Ac α 2–3Gal/ β 1–3GalNAc α 1- <i>O</i> -Ser/Thr	100 (0.39)*	100 (0.02)	100 (7.75)
	Neu5Ac α 2–3Gal/ β 1–3[Neu5Ac α 2–6]GalNAc α 1- <i>O</i> -Ser/Thr			
	Neu5Ac α 2–6(3)Gal/ β 1–4GlcNAc-R			
Asialofetuin	Gal/ β 1–3GalNAc α 1- <i>O</i> -Ser/Thr Gal/ β 1–4GlcNAc-R	0.85	0.5	131
α ₁ -Acid glycoprotein	Neu5Ac α 2–6Gal/ β 1–4GlcNAc-R	8.7	0	nd
Asialo- α ₁ -acid glycoprotein	Gal/ β 1–4GlcNAc-R	0	0	0
BSM	GlcNAc/ β 1–3[Neu5Ac α 2–6]GalNAc α 1- <i>O</i> -Ser/Thr	0	3.3	1.3
	Neu5Ac α 2–6GalNAc α 1- <i>O</i> -Ser/Thr			
Asialo-BSM	GlcNAc/ β 1–3GalNAc α 1- <i>O</i> -Ser/Thr	0	0	19.7
	GalNAc α 1- <i>O</i> -Ser/Thr			
OSM	Neu5Ac α 2–6GalNAc α 1- <i>O</i> -Ser/Thr	0	nd	0
Asialo-OSM	GalNAc α 1- <i>O</i> -Ser/Thr	0	nd	0
Arylglycosides	GalNAc α 1- <i>O</i> -Bz	3.2	nd	0
	Gal/ β 1–3GalNAc α 1- <i>O</i> -Bz	0	nd	0
	Neu5Ac α 2–3Gal/ β 1–3GalNAc α 1- <i>O</i> -Bz	224	nd	nd
	Neu5Gc α 2–3Gal/ β 1–3GalNAc α 1- <i>O</i> -Bz	196	nd	nd
	GlcNAc α 1- <i>O</i> -Bz	7.6	nd	nd
	Gal/ β 1–3GlcNAc/ β - <i>O</i> -Bz	0.6	nd	nd
	Neu5Ac α 2–3Gal/ β 1–3GlcNAc α 1- <i>O</i> -Bz	0	nd	nd
	Gal/ β 1–4GlcNAc/ β - <i>O</i> -Bz	3.6	nd	nd
Oligosaccharides				
	Neu5Ac α 2–3Gal/ β 1–3GlcNAc/ β 1–4Gal/ β 1–4Glc	0	nd	nd
	Gal/ β 1–3GlcNAc/ β 1–4Gal/ β 1–4Glc	0	nd	nd
	Neu5Gc α 2–3Gal/ β 1–3GalNAc	60.3	290	nd
	Neu5Ac α 2–3Gal/ β 1–3GlcNAc	0	nd	nd
Gangliosides				
G _{M3}	Neu5Ac α 2–3Gal/ β 1–4Glc-Cer	nd	0	nd
G _{M1b}	Neu5Ac α 2–3Gal/ β 1–3GalNAc/ β 1–4Gal/ β 1–4Glc-Cer	0	11	nd
G _{M1a}	Gal/ β 1–3GalNAc/ β 1–4[Neu5Ac α 2–3]Gal/ β 1–4Glc-Cer	0	nd	nd
Asialo-G _{M1}	Gal/ β 1–3GalNAc/ β 1–4Gal/ β 1–4Glc-Cer	0	nd	0
G _{D1a}	Neu5Ac α 2–3Gal/ β 1–3GalNAc/ β 1–4[Neu5Ac α 2–3]Gal/ β 1–4Glc-Cer	0	0	nd
Gangliosides mixture		4	nd	0

* Actual activities are shown in parentheses in pmol · h⁻¹ · μ l⁻¹.

† Data are from [47] (for fetuin) and from [48] (for BSM).

‡ Data are from [15].

§ Data are from [13].

Table 2 Determination of the kinetic parameters of hST6GalNAc IV

Incubations were performed using various concentrations of acceptor substrates: 5–500 μ M CMP-[¹⁴C]Neu5Ac; 0.5–10 mM Neu5Gc α 2–3Gal β 1–3GalNAc or Neu5Ac α 2–3Gal β 1–3GalNAc α -*O*-Bz. Note that CMP-Neu5Ac is the donor of sialic acid, and the $V_{\max}/K_m$ ratio cannot be compared with other values.

Substrate	K_m (mM)	$V_{\max}$ (nmol · h ⁻¹ · ml ⁻¹)	$V_{\max}/K_m$ (10 ³ · h ⁻¹)
Neu5Gc α 2–3Gal β 1–3GalNAc	1.10	0.41	0.38
Neu5Ac α 2–3Gal β 1–3GalNAc α - <i>O</i> -Bz	2.21	1.42	0.64
CMP-Neu5Ac	0.07	8.86	–

ST3Gal I and ST3Gal II, as well as ST6Gal I, enzyme activities ([15], and A. Harduin-Lepers, B. Samyn-Petit, M.-A. Krzewinski-Recchi, V. Vallejo-Ruiz and P. Delannoy, unpublished work), this enzyme fraction was not suitable for analysis of the substrate specificity. Thus concentrated cell culture medium was used to assess the substrate specificity of this enzyme. In order to monitor the activity of this form of hST6GalNAc IV,

cell-culture media from cells transfected with either pFLAG-ST6GalNAc IV or the control plasmid pFLAG were collected and assayed for sialyltransferase activity with each of the various acceptor substrates. Using fetuin, almost no activity was detected in the control media (160 d.p.m.), whereas a transfer of sialic acid residues (8800 d.p.m.) appeared using the media of pFLAG-ST6GalNAc IV-transfected cells as the enzyme source (results

Figure 5 Analysis of linkage specificity of hST6GalNAc IV

Radiolabelled sialylated products (Neu5Ac α 2-3Gal β 1-3[Neu5Ac α 2-6]GalNAc-*O*-Bz) resulting from the hST6GalNAc IV activity were purified on Sep-Pak C₁₈ columns and separated further on HPTLC, as described in the Experimental section, and further detected by autoradiography (lane 2). Orcinol/sulphuric acid staining of arylglycoside controls is shown on each side of the autoradiogram. Lane 1, Neu5Ac α 2-3Gal β 1-3GalNAc β -*O*-Bz; lane 3, mixture of GalNAc α -*O*-Bz derivatives extracted from HT-29 G⁻ cells, as described in [26]; #5 is Neu5Ac α 2-3Gal β 1-3GalNAc-*O*-Bz, and #9-a is Neu5Ac α 2-3Gal β 1-3[Neu5Ac α 2-6]GalNAc-*O*-Bz.

not shown). Taken together, these data indicated that a soluble form of the recombinant enzyme is secreted in the media of pFLAG-ST6GalNAc IV-transfected COS-7 cells.

In addition, as has been observed previously for rat brain ST6(*O*)-II enzymic activity [32], enzyme assays performed at a range of incubation temperatures revealed that the temperature optimum was approximately 33 °C for the enzyme produced in the cell-culture medium, while the recombinant hST6GalNAc IV found within the cell had a temperature optimum of 28 °C (results not shown).

The recombinant enzyme efficiently mediated the transfer of sialic acid residues to O-linked oligosaccharides of fetuin (Table

1), which is known to contain three core 1 O-linked oligosaccharides, one of them being a disialylated tetrasaccharide, Neu5Ac α 2-3Gal β 1-3[Neu5Ac α 2-6]GalNAc, and the other two the monosialylated trisaccharide Neu5Ac α 2-3Gal β 1-3GalNAc [33]. The two GalNAc residues of native fetuin unsubstituted in position-6 can therefore serve as an acceptor for hST6GalNAc IV. No significant activity was observed towards asialofetuin (Gal β 1-3GalNAc α 1-*O*-Ser/Thr), asialo-BSM or asialo-OSM (GalNAc α 1-*O*-Ser/Thr), or asialo- α 1-acid-glycoprotein. Free oligosaccharides, glycolipids and arylglycosides (Table 1) did not serve as acceptor substrates for hST6GalNAc IV, with the exception of the structures bearing the sialyl α 2-3Gal β 1-3GalNAc trisaccharide (i.e. Neu5Ac α 2-3Gal β 1-3GalNAc α 1-*O*-Bz, Neu5Gc α 2-3Gal β 1-3GalNAc α 1-*O*-Bz) and the free oligosaccharide Neu5Gc α 2-3Gal β 1-3GalNAc. However, it is important to note that G_{M1b} and G_{D1a}, which also present the structure Neu5Ac α 2-3Gal β 1-3GalNAc, are not acceptor substrates for hST6GalNAc IV (Table 1). The internal α 2-3-linked sialic acid attached to the internal galactose of G_{D1a} abolishes the ability of ST6GalNAc IV to transfer sialic acid to this sequence. From these results, which are comparable with those that have been obtained for mouse ST6GalNAc IV [15] (Table 1), the acceptor substrate specificity of hST6GalNAc IV appears to be restricted to Neu5Ac α 2-3Gal β 1-3GalNAc oligosaccharide, in which the GalNAc is either α -linked to a serine or threonine residue or β -linked to Bz.

Kinetic properties of the recombinant hST6GalNAc IV protein

The kinetic parameters of hST6GalNAc IV were determined using CMP-Neu5Ac as the donor substrate, and Neu5Ac α 2-3Gal β 1-3GalNAc α -*O*-Bz or Neu5Ac α 2-3Gal β 1-3GalNAc as acceptor substrates (Table 2). The K_m value of hST6GalNAc IV for CMP-Neu5Ac (68 μ M) was very close to those described previously for purified or recombinant sialyltransferases, which range from 30 to 100 μ M. The relative substrate activity of hST6GalNAc IV towards Neu5Ac α 2-3Gal β 1-3GalNAc-*O*-Bz

Figure 6 Schematic diagram of the human ST6GalNAc IV genomic, cDNA and protein sequences

(a) Hatched-shaded boxes represent coding sequences, and open boxes represent untranslated sequences of hST6GalNAc IV gene. Exons are denoted E1 to E7, and the intron sequences are shown by the solid lines between the exons, with their sizes indicated below. (b) Representation of hST6GalNAc IV mRNA. The relative position of the exon is indicated by the splicing site. (c) The domain structure of hST6GalNAc IV protein is represented schematically by a rectangle subdivided to show the major structural elements [from left to right, cytoplasmic domain (open box), transmembrane domain (filled box) and catalytic domain (hatched-shaded box)]. Sialyl 'motifs' L, S and VS are also indicated.

and Neu5Ac α 2-3Gal β 1-3GalNAc was $0.64 \times 10^{-3} \text{ h}^{-1}$ and $0.38 \times 10^{-3} \text{ h}^{-1}$ respectively.

Analysis of linkage specificity of hST6GalNAc IV

On the basis of acceptor specificity of the enzyme described above for ST6GalNAc IV, and also of the sialyltransferase activities described previously, we deduced that hST6GalNAc IV synthesized either Neu5Ac α 2-3Gal β 1-3[Neu5Ac α 2-6]GalNAc, commonly found on O-glycans as well as on G_{D1a}, or Neu5Ac α 2-8Neu5Ac α 2-3Gal β 1-3GalNAc, which is found on the ganglioside G_{D1c}. The radiolabelled products resulting from the sialylation of the trisaccharide Neu5Ac α 2-3Gal β 1-3GalNAc-O-Bz were purified on a SepPak C₁₈ column and separated further by TLC, with controls as described previously [22]. Labelled products were detected by autoradiography, and unlabelled controls stained with the orcinol/sulphuric acid reagent (Figure 5). Data obtained clearly demonstrated that hST6GalNAc IV transferred sialic acid to an α 2,6 position of the GalNAc residue.

Genomic organization of hST6GalNAc IV gene

As a first step towards gaining information about the organization and the regulation of this human ST6GalNAc gene in human tissues, we sought to find the genomic sequences of the gene in the human genome databanks. The full-length cDNA was used as a probe to screen the human HTG division of the National Centre for Biotechnology Information [24]. The gene was found to be located on chromosome 9, and its organization is schematically shown in Figure 6. This analysis suggests that the entire ST6GalNAc IV gene spans over at least 10 kb of human genomic DNA. The sialyl 'motifs' L, S and VS are found to be located on one exon each, which is different from what has been described previously for rST6Gal I [34], hST3Gal IV [35], mST6GalNAc II [12], mST8Sia II [36], mST8Sia IV [37,38] and ST3Gal I [39], in which the sialyl 'motifs' are spliced on to two exons each, but is nearer to the situation described previously for mST8Sia III [40].

DISCUSSION

We have previously cloned the cDNAs encoding hST6GalNAc I from HT-29 colon carcinoma cells (M.-A. Krzewinski-Recchi, A. Harduin-Lepers, W. F. A. Steelant, B. Samyn-Petit and P. Delannoy, unpublished work), which is regarded as being the best candidate for the *in vitro* synthesis of sialyl-Tn antigen [9] and hST6GalNAc II [13] from MDA-MB-231 breast cancer cells. In this paper, we describe for the first time the identification, isolation and enzymic characterization of a human cDNA encoding a CMP-Neu5Ac:Neu5Ac α 2-3Gal β 1-3GalNAc α 2,6-sialyltransferase of the ST6GalNAc family by means of an EST-screening procedure.

Sequence analysis of the enzymic protein

As has been observed for other glycosyltransferases, two in-frame ATG codons are present in the mouse ST6GalNAc IV gene that are located 5' of the transmembrane domain, and which potentially lead to the synthesis of two enzymic isoforms [15]. Only one ATG codon has been found in the human cDNA sequence, giving rise to a protein that corresponds to the shorter mouse isoform. Moreover, we noticed that the smaller size of hST6GalNAc IV (302 amino acids) reflects the shortest stem region ever found, giving rise to the smallest sialyltransferase described to date, whereas the size of the catalytic domain is

conserved among the different ST6GalNAc enzymes. The presence of three potential O-glycans in this short stem region could be interpreted as providing a means of preventing the action of proteases that convert a transmembrane protein into a soluble circulating enzyme. Alternatively, another hypothesis would be that this retains the catalytic domain of the enzyme further into the Golgi lumen.

It is interesting to note that the protein is very rich in cysteine residues (17 in total) that are found mainly at the N-terminus of the protein. Twelve cysteine residues are conserved between the human and mouse ST6GalNAc IV enzymes, supporting the formation of up to six disulphide bridges. This might be an indication of their implication in biological functions such as conferring an active structure to the enzymic protein, and would account for why the truncated form of the enzyme we have generated is not active.

Kinetic properties and substrate specificity of the enzyme

In the present paper, we have shown that the acceptor specificity of hST6GalNAc IV is restricted to the trisaccharide Neu5Ac α 2-3Gal β 1-3GalNAc α - or β -linked to an aglycon fraction. A number of studies have attempted to define the acceptor substrate specificity and the kinetic properties of these sialyltransferases. Assays *in vitro* using the recombinant enzymes with various glycoprotein and glycolipid acceptors have shown that ST6GalNAc I sialylates GalNAc residues found in the following structures: GalNAc α 1-O-Ser/Thr, Gal β 1-3GalNAc α 1-O-Ser/Thr and Neu5Ac α 2-3Gal β 1-3GalNAc α 1-O-Ser/Thr. ST6GalNAc II exhibits a narrower substrate specificity, using Gal β 1-3GalNAc α 1-O-Ser/Thr and Neu5Ac α 2-3Gal β 1-3GalNAc α 1-O-Ser/Thr as substrates, whereas ST6GalNAc III uses only Neu5Ac α 2-3Gal β 1-3GalNAc α - or β -linked and, to a lesser extent, the ganglioside G_{M1b}. However, recent evidence obtained from the cloning of a fifth mouse ST6GalNAc enzyme, mST6GalNAc V (specific for G_{M1b}), clearly indicates that this sialylation step of glycoproteins and glycolipids is considerably more complex than was previously believed. It is now clear that an entire family of ST6GalNAcs, with at least five and possibly more members, is involved in this process. These enzymes display subtle differences in substrate specificity, and their expression is differentially regulated in different organs or cells. This level of complex regulation suggests that a high level of functional significance accompanies these particular glycosidic linkages. Future studies should be aimed at analysing precisely the acceptor substrate specificity of each of the enzymes by way of *in vitro* studies, as well as using *in vivo* model systems with gene knock-outs or the introduction *de novo* of specific sialyltransferases.

Existence of several additional genes in the ST6GalNAc family may be predicted

The exponential rise in the number of sequences accumulated in public databases via the systematic sequencing of the human genome, along with the development of powerful computational tools, has given rise to the new era of computer cloning. This EST cloning strategy has been used previously to identify several different families of glycosyltransferases, such as the β 1,4-galactosyltransferases [41,42], β 1,3-galactosyltransferases [43,44] and core 2/4 N-acetylglucosaminyltransferases [45]. An estimate from the studies in our laboratory suggests the presence of at least six or seven genes in the human genome that bear 15–48% amino acid identity with the human ST6GalNAc IV sequence, when comparing with the full-length protein. The presence of several ST6GalNAc isoenzymes underlines the complexity in

O-linked oligosaccharide biosynthesis. One can hypothesize that the different ST6GalNAc enzymes would have distinct substrate-recognition profiles, and that this would provide a mechanism for the cell to retain a greater potential for sialylating different related acceptor structures. However, only relatively minor variations in substrate preference have been reported so far for the ST6GalNAc III, ST6GalNAc IV and ST6GalNAc V enzymes. The possibility remains that some isoforms serve more general housekeeping functions, while others act on specific molecules only. Most of these data have been obtained with recombinant enzymes in *in vitro* assays, which assess the specificity of a single sialyltransferase for a given exogenous substrate only, whereas studies *in vivo* would represent the accumulated activities of the entire repertoire of sialyltransferases expressed in the chosen cell type, i.e. data concerning the specificity of sialyltransferases assessed *in vitro* might fail to reflect their *in vivo* activity.

Chromosomal location and gene organization

Gilley and Fried [27] have recently cloned seven *Fugu rubripes* (*Fugu*) genes closely linked to the *Surfeit* genes in two regions of the *Fugu* genome, and they have found by homology searching of the public databases that one of these genes was similar to a *Rattus norvegicus* cDNA encoding ST6GalNAc III (L29554). This *Fugu* gene was named provisionally SIAT3C (GenBank® Y17466), and the authors mapped to chromosome band 9q34 the human homologue of the *Fugu* SIAT3C gene, by both PCR analysis of a panel of human chromosomal somatic cell-hybrid DNAs (Genebridge 4 RH panel) and fluorescence *in situ* hybridization. In addition, Southern blotting, subcloning and sequencing of representative cosmids of a human chromosome-9-specific cosmid library enabled them to determine the exon-intron boundaries of the last two exons (accession numbers Y17461 and Y17460) of the human SIAT3C-related gene. We have determined that these two human exons belong to the human ST6GalNAc IV gene, and we have been able to elucidate the genomic organization of this gene (Figure 6) through homology searches in the HTG division of the public databases. This allows us to conclude that the human ST6GalNAc IV and the *Fugu* SIAT3C genes are orthologues, since the boundaries of these exons are conserved.

The five ST6GalNAc proteins described to date in the mouse and human species are very different in terms of size, with hST6GalNAc IV representing the smallest sialyltransferase cloned to date (302 amino acids), and hST6GalNAc I representing the longest. When comparing the primary protein structures, genomic organization and phylogenetic trees of the different ST6GalNAc sialyltransferases, two subfamilies appear: one featuring ST6GalNAc I and ST6GalNAc II, the other grouping together ST6GalNAc-III, -IV and -V. These groupings suggest a divergence arising from gene duplication in one organism resulting in functionally distinct proteins.

Expression pattern of ST6GalNAc genes in cells

Human ST6GalNAc IV mRNA appears to be expressed at a low level in all cell types as a 2.2 kb mRNA, and as a 2.5 kb mRNA in brain, heart and skeletal muscle. This constitutive expression was also observed in more than 20 different cancer cell lines for hST6GalNAc II [13]. The identification of all the members of the human ST6GalNAc family will allow us to precisely determine the specific implication of these different sialyltransferases in the biosynthesis of sialylated O-glycans, especially in cancerous cells and tissues, and to evaluate the putative correlation between ST6GalNAc expression and tumour development.

We acknowledge the expert technical assistance of Anne-Marie Mir in preparation of cell cultures. We are grateful to Dr. L. Oomen for skilled assistance with the immunofluorescence microscopy studies. This work was supported in part by the University of Sciences and Technologies of Lille, the European Carbohydrate Research Network Carenet 2, the CNRS 'Programme Interdisciplinaire Physique et Chimie du Vivant, Réseau G-Trec' and the 'Association pour la Recherche sur le Cancer' (grant no. 5469). W.S. has a fellowship of the European Carbohydrate Research Network Carenet 2.

REFERENCES

- 1 IUPAC-IUB Joint Commission on Biochemical Nomenclature 1998 Nomenclature of glycolipids. *Eur. J. Biochem* **257**, 293–298
- 1a Tsuji, S., Datta, A. K. and Paulson, J. C. (1996) Systematic nomenclature for sialyltransferases. *Glycobiology* **6**, v–xiv
- 1b Harduin-Lepers, A., Recchi, M. A. and Delannoy, P. (1995) 1994, the year of sialyltransferases. *Glycobiology* **5**, 741–758
- 2 Varki, A. (1994) Selectin ligands. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7390–7397
- 3 Hakomori, S. I. (1991) Possible functions of tumor-associated carbohydrate antigens. *Curr. Opin. Immunol.* **3**, 646–653
- 4 Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130
- 5 Kelm, S. and Schauer, R. (1997) Sialic acids in molecular and cellular interactions. *Int. Rev. Cytol.* **175**, 137–240
- 6 Kjeldsen, T., Clausen, H., Hirohashi, S., Ogawa, T., Iijima, H. and Hakomori, S. I. (1988) Preparation and characterization of monoclonal antibodies directed to the tumor-associated O-linked sialosyl-2-6- α -N-acetylglactosaminyl (sialosyl-Tn) epitope. *Cancer Res.* **48**, 2214–2220
- 7 Fukutomi, T., Hirohashi, S., Tsuda, H., Nanasawa, T., Yamamoto, H., Itabashi, M. and Shimoto, Y. (1991) The prognostic value of tumor-associated carbohydrate structures correlated with gene amplifications in human breast carcinomas. *Jpn. J. Surg.* **21**, 499–507
- 8 Kurosawa, N., Hamamoto, T., Lee, Y. C., Nakaoka, T., Kojima, N. and Tsuji, S. (1994) Molecular cloning and expression of GalNAc α 2,6-sialyltransferase. *J. Biol. Chem.* **269**, 1402–1409
- 9 Kurosawa, N., Takashima, S., Kono, M., Ikehara, Y., Inoue, M., Tachida, Y., Narimatsu, H. and Tsuji, S. (2000) Molecular cloning and genomic analysis of mouse GalNAc α 2,6-sialyltransferase (ST6GalNAc I). *J. Biochem. (Tokyo)* **127**, 845–854
- 10 Ikehara, Y., Kojima, N., Kurosawa, N., Kudo, T., Kono, M., Nishihara, S., Issiki, S., Morozumi, K., Itzkowitz, S., Tsuda, T. et al. (1999) Cloning and expression of a human gene encoding an N-acetylglactosamine- α 2,6-sialyltransferase (ST6GalNAc I): a candidate for synthesis of cancer-associated sialyl-Tn antigens. *Glycobiology* **9**, 1213–1224
- 11 Kurosawa, N., Kojima, N., Inoue, M., Hamamoto, T. and Tsuji, S. (1994) Cloning and expression of Gal β 1,3GalNAc-specific GalNAc α 2,6-sialyltransferase. *J. Biol. Chem.* **269**, 19048–19053
- 12 Kurosawa, N., Inoue, M., Yoshida, Y. and Tsuji, S. (1996) Molecular cloning and genomic analysis of mouse Gal β 1,3GalNAc-specific GalNAc α 2,6-sialyltransferase. *J. Biol. Chem.* **271**, 15109–15116
- 13 Samyn-Petit, B., Krzewinski-Recchi, M. A., Steelant, W. F. A., Delannoy, P. and Harduin-Lepers, A. (2000) Molecular cloning and functional expression of human ST6GalNAc II. Molecular expression in various human cultured cells. *Biochim. Biophys. Acta* **1474**, 201–211
- 14 Sjöberg, E. R., Kitagawa, H., Glushka, J., Van Halbeek, H. and Paulson, J. C. (1996) Molecular cloning of a developmentally regulated N-acetylglactosamine α 2,6-sialyltransferase specific for sialylated glycoconjugates. *J. Biol. Chem.* **271**, 7450–7459
- 15 Lee, Y. C., Kaufmann, M., Kitazume-Kawaguchi, S., Kono, M., Takashima, S., Kurosawa, N., Liu, H., Pircher, H. and Tsuji, S. (1999) Molecular cloning and functional expression of two members of mouse Neu5Ac α 2,3Gal β 1,3GalNAc GalNAc α 2,6-sialyltransferase family, ST6GalNAc III and IV. *J. Biol. Chem.* **274**, 11958–11967
- 16 Ikehara, Y., Shimizu, N., Kono, M., Nishihara, S., Nakanishi, H., Kitamura, T., Narimatsu, H., Tsuji, S. and Tatematsu, M. (1999) A novel glycosyltransferase with a polyglutamine repeat; a new candidate for G_{D12} synthase (ST6GalNAc V). *FEBS Lett.* **463**, 92–96
- 17 Okajima, T., Fukumoto, S., Ito, H., Kiso, M., Hirabayashi, Y., Urano, T. and Furukawa, K. (1999) Molecular cloning of brain-specific G_{D12} synthase (ST6GalNAc V) containing CAG/Glutamine repeats. *J. Biol. Chem.* **274**, 30557–30562
- 18 Augé, C. and Veyrières, A. (1976) Synthesis of a disaccharide oxazoline: 2-methyl-[4,6-di-O-acetyl-1,2-dideoxy-3-O-(2,3,4,6-tetra-O-acetyl- β -D-galactopyranosyl)- α -D-glucopyranosyl][2',1':4,5]-2-oxazoline. *Carbohydr. Res.* **46**, 293–298

- 19 Lubineau, A., Augé, C. and François, P. (1992) The use of porcine liver (2-3)- α -sialyltransferase in the large-scale synthesis of α -Neup5Ac-(2-3)- β -D-Galp-(1-3)-D-GlcNAc, the epitope of the tumor-associated carbohydrate antigen CA 50. *Carbohydr. Res.* **228**, 137-144
- 20 Lubineau, A., Augé, C., Gautheron-Le Narvor, C. and Ginot, J.-C. (1994) Combined chemical and enzymatic synthesis of the sialylated non reducing terminal sequence of G_{M1b} glycolylated ganglioside, a potential human tumor marker. *Bioorg. Med. Chem.* **2**, 669-674
- 21 Lubineau, A., Basset-Carpentier, K. and Augé, C. (1997) Porcine liver (2-3)- α -sialyltransferase: substrate specificity studies and application of the immobilized enzyme to the synthesis of various sialylated oligosaccharide sequences. *Carbohydr. Res.* **300**, 161-167
- 22 Zanetta, J. P., Vitiello, F. and Vincendon, G. (1980) Gangliosides from rat cerebellum: demonstration of considerable heterogeneity using a new solvent for thin layer chromatography. *Lipids* **15**, 1055-1061
- 23 Sambrook, J., Fritsch, E. J. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- 24 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402
- 25 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463-5467
- 26 Zanetta, J. P., Gouyer, V., Maes, E., Pons, A., Hemon, B., Zweibaum, A., Delannoy, P. and Huet, G. (2000) Massive *in vitro* synthesis of tagged oligosaccharides in 1-benzyl-2-acetamido-2-deoxy- α -D-galactopyranoside treated HT-29 cells. *Glycobiology* **10**, 565-575
- 27 Gilley, J. and Fried, M. (1999) Extensive gene order differences within regions of conserved synten between the *Fugu* and human genomes: implications for chromosomal evolution and the cloning of disease genes. *Hum. Mol. Genet.* **8**, 1313-1320
- 28 Kozak, M. (1992) Regulation of translation in eukaryotic systems. *Annu. Rev. Cell Biol.* **8**, 197-225
- 29 Datta, A. K. and Paulson, J. C. (1997) Sialyl motifs of sialyltransferases. *Indian J. Biochem. Biophys.* **34**, 157-165
- 30 Drickamer, K. (1993) A conserved disulphide bond in sialyltransferases. *Glycobiology* **3**, ii-iii
- 31 Geremia, A. R., Harduin-Lepers, A. and Delannoy, P. (1997) Identification of two novel conserved amino acid residues in eukaryotic sialyltransferases: implications for their mechanism of action. *Glycobiology* **7**, v-xi
- 32 Baubichon-Cortay, H., Serres-Guillaumond, M., Louisot, P. and Broquet, P. (1986) A brain sialyltransferase having a narrow specificity for O-glycosyl-linked oligosaccharide chains. *Carbohydr. Res.* **149**, 209-223
- 33 Nilsson, B., Norden, N. E. and Svensson, S. (1979) Structural studies on the carbohydrate portion of fetuin. *J. Biol. Chem.* **254**, 4545-4553
- 34 Svensson, E. C., Soreghan, B. and Paulson, J. C. (1990) Organization of the β -galactoside α 2,6-sialyltransferase gene. Evidence for the transcriptional regulation of terminal glycosylation. *J. Biol. Chem.* **265**, 20863-20868
- 35 Kitagawa, H., Mattei, M. G. and Paulson, J. C. (1996) Genomic organization and chromosomal mapping of the Gal β 1,3GalNAc/Gal β 1,4GlcNAc α 2,3-sialyltransferase. *J. Biol. Chem.* **271**, 931-938
- 36 Yoshida, Y., Kurosawa, N., Kanematsu, T., Kojima, N. and Tsuji, S. (1996) Genomic structure and promoter activity of the mouse polysialic acid synthase gene (mST8SialI). Brain-specific expression from a TATA-less GC-rich sequence. *J. Biol. Chem.* **271**, 30167-30173
- 37 Takashima, S., Yoshida, Y., Kanematsu, T., Kojima, N. and Tsuji, S. (1998) Genomic structure and promoter activity of the mouse polysialic acid synthase (mST8SialV/PST) gene. *J. Biol. Chem.* **273**, 7675-7683
- 38 Eckhardt, M. and Gerardy-Schahn, R. (1998) Genomic organization of the murine polysialyltransferase gene ST8SialV (PST-1). *Glycobiology* **8**, 1165-1172
- 39 Chang, M. L., Eddy, R. L., Shows, T. B. and Lau, J. T. Y. (1995) Three genes that encode human β -galactoside α 2,3-sialyltransferases. Structural analysis and chromosomal mapping studies. *Glycobiology* **5**, 319-325
- 40 Yoshida, Y., Kurosawa, N., Kanematsu, T., Taguchi, A., Arita, M., Kojima, N. and Tsuji, S. (1996) Unique genomic structure and expression of the mouse α 2,8-sialyltransferase (ST8Sia III) gene. *Glycobiology* **6**, 573-580
- 41 Almeida, R., Amado, M., David, L., Levery, S. B., Holmes, E. H., Merckx, G., van Kessel, A. G., Rygaard, E., Hassan, H., Bennett, E. P. and Clausen, H. (1997) A family of human β 4-galactosyltransferases. Cloning and expression of two novel UDP-galactose: β -N-acetylglucosamine β 1,4-galactosyltransferases, β 4Gal-T2 and β 4Gal-T3. *J. Biol. Chem.* **272**, 31979-31991
- 42 Lo, N. W., Shaper, J. H., Pevsner, J. and Shaper, N. L. (1998) The expanding β 4-galactosyltransferase gene family: messages from the databanks. *Glycobiology* **8**, 517-526
- 43 Hennet, T., Dinter, A., Kuhnert, P., Mattu, T. S., Rudd, P. M. and Berger, E. G. (1998) Genomic cloning and expression of three murine UDP-galactose: β -N-acetylglucosamine β 1,3-galactosyltransferase genes. *J. Biol. Chem.* **273**, 58-65
- 44 Amado, M., Almeida, R., Carneiro, F., Levery, S. B., Holmes, E. H., Nomoto, M., Hollingsworth, M. A., Hassan, H., Schwientek, T., Nielsen, P. A. et al. (1998) A family of human β 3-galactosyltransferases. Characterisation of four members of a UDP-galactose: β -N-acetylglucosamine/ β -N-acetylglactosamine β -1,3-galactosyltransferase family. *J. Biol. Chem.* **273**, 12770-12778
- 45 Schwientek, T., Nomoto, M., Levery, S. B., Merckx, G., van Kessel, A. G., Bennett, E. P., Hollingsworth, M. A. and Clausen, H. (1999) Control of O-glycan branch formation. Molecular cloning of human cDNA encoding a novel β 1,6-N-acetylglucosaminyltransferase forming core 2 and core 4. *J. Biol. Chem.* **274**, 4504-4512
- 46 Hofmann, K. and Stoffel, W. (1993) A database membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **347**, 166
- 47 Spiro, R. G. and Bhoyroo, V. D. (1974) Structure of the O-glycosidically linked carbohydrate units of fetuin. *J. Biol. Chem.* **249**, 5703-5717
- 48 Tsuji, T. and Osawa, T. (1986) Carbohydrate structures of bovine submaxillary mucin. *Carbohydr. Res.* **151**, 391-402

Received 11 April 2000/5 July 2000; accepted 30 August 2000