Molecular cloning, gene organization and expression of the human UDP-GalNAc:Neu5Acα2-3Galβ-R β1,4-*N*-acetylgalactosaminyltransferase responsible for the biosynthesis of the blood group Sd^a/Cad antigen: evidence for an unusual extended cytoplasmic domain

Maria-Dolores MONTIEL, Marie-Ange KRZEWINSKI-RECCHI, Philippe DELANNOY and Anne HARDUIN-LEPERS[1]

Unité de Glycobiologie Structurale et Fonctionnelle, UMR CNRS no. 8576, Laboratoire de Chimie Biologique, Université des Sciences et Technologies de Lille, F-59655 Villeneuve d'Ascq, France

The human Sd^a antigen is formed through the addition of an *N*-acetylgalactosamine residue via a β1,4-linkage to a sub-terminal galactose residue substituted with an α2,3-linked sialic acid residue. We have taken advantage of the previously cloned mouse cDNA sequence of the UDP-GalNAc:Neu5Acα2-3Galβ-R β1,4-*N*-acetylgalactosaminyltransferase (Sd^a β1,4GalNAc transferase) to screen the human EST and genomic databases and to identify the corresponding human gene. The sequence spans over 35 kb of genomic DNA on chromosome 17 and comprises at least 12 exons. As judged by reverse transcription PCR, the human gene is expressed widely since it is detected in various amounts in almost all cell types studied. Northern blot analysis indicated that five Sd^a β1,4GalNAc transferase transcripts of 8.8, 6.1, 4.7, 3.8 and 1.65 kb were highly expressed in colon and to a lesser extent in kidney, stomach, ileum and rectum. The complete coding nucleotide sequence was amplified from Caco-2 cells. Interestingly, the alternative use of two first exons, named $E1_S$ and $E1_L$, leads to the production of two transcripts. These nucleotide sequences give rise potentially to two proteins of 506 and 566 amino acid residues, identical in their sequence with the exception of their cytoplasmic tail. The short form is highly similar (74 % identity) to the mouse enzyme whereas the long form shows an unusual long cytoplasmic tail of 66 amino acid residues that is as yet not described for any other mammalian glycosyltransferase. Upon transient transfection in Cos-7 cells of the common catalytic domain, a soluble form of the protein was obtained, which catalysed the transfer of GalNAc residues to α2,3-sialylated acceptor substrates, to form the GalNAcβ1-4[Neu5Acα2-3]Galβ1-R trisaccharide common to both Sd^a and Cad antigens.

Key words: molecular cloning, UDP-GalNAc:Neu5Acα2-3Galβ-R β1,4-*N*-acetylgalactosaminyltransferase.

INTRODUCTION

The Sd^a (Sid blood group) antigen was independently discovered in 1967 by MacVie et al. [1] and Renton et al. [2] as a dominant character found in more than 90 % of Caucasian red blood cells. In addition to erythrocytes, the Sd^a antigen is also described in tissues and body fluids, particularly in urine of humans and other mammals [3]. In adult humans, the Sd^a antigen appears to be restricted to the kidney and the large intestine [4]. The structure of the Sd^a antigen was first elucidated by Donald et al. [5] as the GalNAcβ1-4[Neu5Acα2-3]Galβ1-4GlcNAcβ1-3Gal pentasaccharide carried by the N-glycans of the Tamm–Horsfall glycoprotein (THp; the major urinary glycoprotein) purified from the urine of Sd^{a+} individuals. The terminal non-reducing trisaccharide GalNAcβ1-4[Neu5Acα2-3]Galβ is also found in the Cad antigen, first described by Blanchard et al. [6] as the pentasaccharide GalNAcβ1-4[Neu5Acα2-3]Galβ1-3[Neu5Acα2-6]GalNAc, O-linked to a serine or a threonine residue of glycophorin A on red blood cells. The Cad antigen is highly expressed in normal mucins of human descending colon [7] or on normal glycolipids of the human fundic mucosa in stomach [8], but its expression decreases dramatically in colon cancer mucins and in gastric cancer, respectively [7,8].

The last step in the biosynthesis of both Sd^a and Cad antigens is catalysed by the UDP-GalNAc:Neu5Acα2-3Galβ1-R β1,4-*N*-acetylgalactosaminyltransferase (Sd^a β1,4GalNAc transferase) that transfers a GalNAc residue from UDP-GalNAc to the C_4 of the Gal residue of the Neu5Acα2-3Galβ1-R terminal sequence found on both N- and O-linked glycans of fetuin and human chorionic gonadotropin [9,10]. The Sd^a β1,4GalNAc transferase activity has been described essentially in human intestine [11], colon [4] and kidney [12]. In addition, this enzyme activity has been detected in different body fluids, such as blood plasma [13] and urine [14]. In both colon and kidney, this enzyme appears to be onco-developmentally regulated. In rat colon and guinea pig kidney, it is practically absent at birth and increases as a function of age [15,16]. The activity of the Sd^a β1,4GalNAc transferase is also dramatically reduced in human colon carcinomas compared with normal colon [11]. As suggested by enzymic activity determined *in vitro*, colon adenocarcinoma cell lines lack the Sd^a β1,4GalNAc transferase, but one cell line (Caco-2) shows significant enzymic activity, which increases with the degree

Abbreviations used: RT-PCR, reverse transcription PCR; THp, Tamm–Horsfall glycoprotein; Sd^a β1,4GalNAc transferase, UDP-GalNAc:Neu5Acα2-3Galβ1-R β1,4-*N*-acetylgalactosaminyltransferase; HTG sequence, High-Throughput Genomic sequence; 5′-RACE, 5′ rapid amplification of cDNA ends; ORF, open reading frame.

[1] To whom correspondence should be addressed (e-mail Anne.Harduin@univ-lille1.fr).

The nucleotide sequence of the short and long transcripts of β1,4-*N*-acetylgalactosaminyltransferase have been submitted to the DDBJ, EMBL, GenBank® and GSDB Nucleotide Sequence Databases under accession nos AJ517770 and AJ517771 respectively.

of enterocytic differentiation [17]. This finding supports the conclusion that expression of the Sd^a β1,4GalNAc transferase is a marker of the colonic cell differentiation. Until now, only the mouse Sd^a β1,4GalNAc transferase cDNA has been cloned and characterized [18]. In order to understand the implication of the Sd^a/Cad antigen and the regulation of its expression in normal human tissues and in cancer cells, we have used sequence information from the mouse cDNA clone to identify the corresponding human gene in the human genome databank. We have obtained and characterized a human Sd^a β1-4GalNAc transferase cDNA from Caco-2 cells that we have transiently expressed in Cos-7 cells. The recombinant enzyme efficiently transfers GalNAc residues from UDP-GalNAc on to α3-sialylated type 2 or type 3 structures. In addition, we show for the first time a preferred transfer activity on to the sialylated type 1 (Neu5Acα2-3Galβ1-3GlcNAc) structure. We also provide strong evidence that two proteins are derived from a unique gene. Finally, transcriptional expression of this gene was examined among various normal human tissues and cell lines.

EXPERIMENTAL

Materials

UDP-[^{14}C]GalNAc (2.076 GBq · mmol^{-1}), Redivue-stabilized [α-^{32}P]dCTP (110 TBq · mmol^{-1}), First Strand cDNA Synthesis Kit and Rediprime II DNA labelling system were from Amersham Biosciences (Little Chalfont, Bucks., U.K.). *Tfu* DNA polymerase was from Q Biogen (Illkirch, France) and the Dynazyme Ext Taq DNA polymerase was from Ozyme (Saint-Quentin-en-Yvelines, France). Oligonucleotides were synthesized by Eurogentec (Seraing, Belgium). Dulbecco's modified Eagle's medium with glucose and without glutamine was from BioWhittaker Europe. Eagle's α-minimal essential medium, OPTIMEM, L-glutamine and antibiotics used in cell culture were from Gibco-BRL (Cergy-Pontoise, France). LipofectAMINE™ Plus Reagent and TA Cloning kit were from Invitrogen (Carlsbad, CA, U.S.A.). Restriction endonucleases were from Eurogentec. Super-Express mRNA β-actin-normalized Northern blot of 12 human normal adult tissues (NmHAN-VII-E) was from the Biochain Institute (San Leandro, CA, U.S.A.). α_1-Acid glycoprotein, fetuin, bovine submaxillary mucin, Galβ1-3GalNAcα-*O*-benzyl, Galβ1-3GlcNAc, UDP-GalNAc, ATP and the expression vector pFLAG-CMV-1 were from Sigma (St Louis, MO, U.S.A.). Lacto-*N*-tetraose and ovine submaxillary mucin were given generously by Dr J. C. Michalski, Dr G. Strecker and Dr F. Chirat (UMR CNRS 8576, Villeneuve d'Ascq, France). Galβ1-3GlcNAcβ1-octyl, Neu5Acα2-3Galβ1-3GalNAc-*O*-benzyl, Neu5Acα2-3Galβ1-3GlcNAc, Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα-*O*-CH_3 and Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα-*O*-CH_3 were generous gifts from Dr C. Augé (URA CNRS 462, Orsay, France). $Na^+SO_3^-$-3′-Galβ1-4Glc and Na^+COO^--3′-Galβ1-4Glc were provided by Professor G. Russo (University of Milan, Milan, Italy) and synthesized by A. Rencurosi as described in [19].

Isolation of RNA and cDNA synthesis

Caco-2 cells were cultured as reported previously [20]. Total RNA was extracted using the nucleospin RNA II kit (Macherey-Nagel, Düren, Germany), 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 according to the manufacturer's protocol.

Isolation of Sd^a β1,4GalNAc transferase cDNA and homology searching

Human High-Throughput Genomic (HTG) sequences (GenBank® accession no. AC069454) with high similarities to mouse Sd^a β1,4GalNAc transferase (GenBank® accession no. L30104) were identified using the BLASTN and TBLASTN algorithms (version 2.2.1 and 2.2.4) against the human EST (dbEST) and the HTG sequence divisions of the public GenBank®/EMBL databases at the National Center for Biotechnology Information [21]. Amino acid sequence and hydropathy analyses were performed using the Internet programs transduction pour publication (http://www.infobiogen.fr/services/analyseq/cgi-bin/forpub_in.pl; Infobiogen, Villejuif, France) and TM pred (http://www.ch.embnet.org/software/TMPRED_form.html; Expasy, Swiss Institute of Bioinformatics, Switzerland) [22]. Determination of potential O-glycosylation sites was performed using the NetOGlyc 2.0 Internet program (Expasy; http://www.cbs.dtu.dk/services/NetOGlyc/) [23].

To isolate cDNA clones, a reverse transcription PCR (RT-PCR) was performed with 150 ng of first-strand cDNA synthesized from the total RNA of Caco-2 cells, using the following primer sets: sense primer-1, 5′-CTGTGGCTCCTCAAGATATTGGTC-3′ (nt 205–228 in Figure 1); antisense primer-1, 5′-AACGTTCACC-TATCAGGAGTCCAG-3′ (nt 1768–1791); sense primer-2, 5′-TTCAGGATGCCTATGGCCAGAGCG-3′ (nt 455–478) and antisense primer-2, 5′-TGAGGACCTGGACTCCAGACTCAC-3′ (nt 859–882). Each primer (0.15 μM) was combined with 0.4 units of *Tfu* DNA polymerase and 200 μM dNTPs. For the first primer pair, PCR was performed for 38 cycles at 95 °C for 45 s, 58 °C for 1 min and 72 °C for 3 min followed by 10 min at 72 °C. For the second primer pair, PCR was performed for 35 cycles at 95 °C for 45 s, 52 °C for 1 min and 72 °C for 1 min followed by 10 min at 72 °C. The PCR-amplified fragments, of 1587 and 428 bp respectively, were subcloned into pUC19 or pCR 2.1 vector (TA Cloning Kit) and their sequences were determined by the dideoxy chain-termination method [24].

5′ Rapid amplification of cDNA ends (5′-RACE)

Amplification of the 5′-end of human Sd^a β1,4GalNAc transferase was achieved with the FirstChoice™ RLM-RACE kit (Ambion, Montouge, France) according to the manufacturer's instructions. Total RNA (10 μg) from Caco-2 cells was treated with calf intestinal phosphatase and then with tobacco acid pyrophosphatase, leaving a 5′-monophosphate full-length mRNA. A 45-base RNA adapter oligonucleotide was then ligated to the RNA population using T4 RNA ligase. A random-primed reverse transcription reaction was performed followed by two consecutive PCRs with 200 μM dNTPs and 0.6 units of *Tfu* DNA polymerase using two nested sets of primers. The oligonucleotide sense-outer, 5′-GCTGATGGCGATGAATGAACACTG-3′, and the gene-specific antisense oligonucleotide 5′-TCCCTGCTC-TTTGTTGGCTTCACA-3′ (nt 421–444; Figure 1) were used in a first PCR at 94 °C for 3 min, followed by 35 cycles (94 °C for 45 s, 60 °C for 1 min and 72 °C for 2 min) and an extension step of 10 min at 72 °C. The oligonucleotide sense-inner, 5′CGCG-GATCCGAACACTGCGTTTGCTGGCTTTGATG-3′, and the gene-specific antisense oligonucleotide 5′-TGGGAGTTCTGG-CTTGGGGCTGCTGAACACTGCTT-3′ (nt 281–315) were used in a second PCR at 94 °C for 2 min, followed by 38 cycles (94 °C

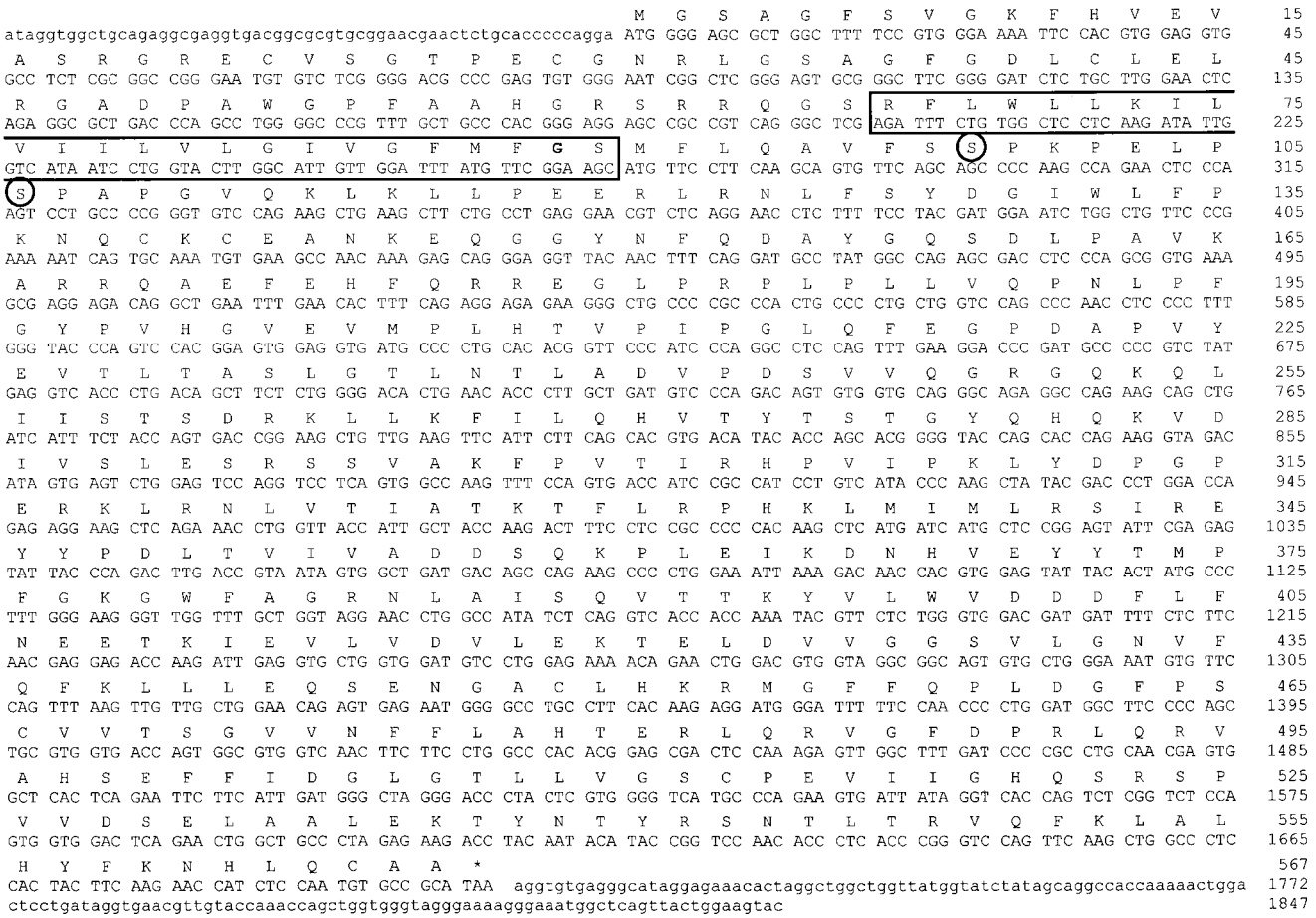


Figure 1 Nucleotide and predicted amino acid sequences of human long-form Sd^a β1,4GalNAc transferase

Numbering of the cDNA begins with the initiation codon. The amino acid sequence is shown in single-letter code. The putative N-terminal transmembrane domain is boxed. Putative O-glycosylation sites are encircled. The recombinant protein sequence starts at Gly-89, indicated in bold.

for 45 s, 65 °C for 2 min and 72 °C for 3 min) and an extension step of 10 min at 72 °C. Amplification products were analysed on a 1 % (w/v) agarose gel with ethidium bromide staining, extracted from the gel, subcloned and sequenced.

Construction and transfection of a soluble Sd^a β1,4GalNAc transferase

A truncated form of the Sd^a β1,4GalNAc transferase lacking all the cytoplasmic tail and the first 22 amino acids of the transmembrane domain was prepared by PCR amplification using a 5′ primer containing a *Not*I site (in italics), 5′-GTTGGA-*GCGGCCGC*CGGAAGCATGTTCCTTCAA-3′ (nt 250–282 in Figure 1), and a 3′ primer containing a *Xba*I site (in italics), 5′-TG-TTTC*TCTAGA*GCCCTCACACCTTTATGCGGC-3′ (nt 1693–1725). The resulting 1476 bp amplified fragment was subcloned into the pUC19 vector. The inserted fragment was cut out by digestion with *Not*I and *Xba*I, and inserted into the *Not*I and *Xba*I sites of the pFlag expression vector. Restriction-enzyme digestions and DNA sequencing confirmed the insert junctions. The resulting plasmid consisted of the signal peptide sequence, the FLAG sequence and a truncated form of the Sd^a β1,4GalNAc transferase cDNA. Qiagen-purified plasmid was transiently transfected into Cos-7 cells on a 100 mm^2 plate using LipofectAMINE™ reagent, following the instructions of the manufacturer. The cell-culture media were harvested 48 and 72 h after transfection and used as the enzyme source.

Enzymic assays

Enzyme activity was measured in 0.1 M cacodylate buffer, pH 6.2, 10 mM $MnCl_2$, 0.2 % Triton CF-54, 4.5 mM ATP and 1.017 mM UDP-[^{14}C]GalNAc (257 KBq), with one of the acceptor substrates and 16.5 μl of the enzyme source in a final volume of 50 μl. For the acceptor substrates, 2 mg · ml^{-1} glycoproteins, 1 mM arylglycosides and 0.5 mM free oligosaccharides were used, and the enzyme reaction was performed at 37 °C for 4 h. For glycoproteins, the reaction was terminated by the addition of 100 μl of water and 1 ml of 1 % phosphotungstic acid in 0.5 M HCl, followed by filtration through a fibreglass filter (GF/A; Whatman Biosystems, Maidstone, Kent, U.K.). Precipitates were washed extensively with 10 % trichloroacetic acid, distilled water and ethanol and processed for scintillation counting. For arylglycosides, the reaction was stopped with 1 ml of water. Products were applied on a C_{18} Sep-Pack cartridge (Millipore Corp., Milford, MA, U.S.A.) and eluted with acetonitrile. For free oligosaccharides, the reaction mixture was heated at 100 °C for 5 min and then centrifuged. The supernatant was subjected

to TLC (Silicagel 60; Merck) in the following solvent: ethanol/1-butanol/pyridine/water/acetic acid (100/10/10/30/3, by vol.) and the radioactive products were detected and quantified using a Personal Molecular Imager® FX (Bio-Rad, Hercules, CA, U.S.A.). The intensity of the radioactivity was converted into moles using the specific radioactivity of UDP-[^{14}C]GalNAc. After chromatography of oligosaccharides, free unlabelled oligosaccharides were stained with the orcinol/H_2SO_4 reagent [25]. For determination of kinetic parameters, incubations were performed as described above using various concentrations of substrates: 0–1 mM UDP-GalNAc, 0–500 μM Neu5Acα2-3Galβ1-3GlcNAcβ-*O*-benzyl or 0–5 mg $\cdot$ ml^{-1} fetuin. Kinetic parameters were determined by Lineweaver–Burk plots and the K_m for fetuin is expressed in mM relative to the number of O-glycosidically linked Galβ1-3GalNAc residues, as determined by GC-MS (40 nmol $\cdot$ mg^{-1}).

Northern blot analysis

A human multiple-tissue Super-Expressed mRNA Northern blot was normalized by expression of the β-actin gene and probed with a 1475 bp [α-^{32}P]dCTP-labelled Sd^a β1,4GalNAc transferase cDNA (nt 70–1545 in Figure 1). The membrane was pre-hybridized for 5 h at 42 °C, hybridized overnight at 65 °C and washed three times for 40 min at room temperature with 2 × SSC (where 1 × SSC is 0.15 M NaCl/0.015 M sodium citrate)/0.05 % SDS. This was followed by washing twice (20 min each time) at 50 °C with 0.1 × SSC/0.1 % SDS. After 72 h exposure, the membrane was analysed using a Personal Molecular Imager® FX system.

Sd^a β1,4GalNAc transferase gene expression in various human cell lines

First-strand cDNA (150 ng) synthesized from total RNA of various cultured cells was combined with 0.15 μM Sd^a β1,4GalNAc transferase-specific primers {sense primer, 5′-TTCA-GGATGCCTATGGCCAGAGCG-3′ (nt 455–478 in Figure 1), antisense primer, 5′-TGAGGACCTGGACTCCAGACTCAC-3′ (nt 859–882 in Figure 1), or GAPDH-specific primers (accession no. J04038), as described previously [26]}, 0.2 mM dNTP and 0.3 units of Dynazyme DNA polymerase. Amplification of a specific Sd^a β1,4GalNAc transferase fragment of 428 bp was obtained after 35 cycles at 45 s at 95 °C, 1 min at 52 °C and 1 min at 72 °C. PCR reactions was size-separated on a 1 % agarose gel.

RESULTS

Identification of the human gene and isolation of Sd^a β1,4GalNAc transferase cDNA

The nucleotide sequence of the mouse Sd^a β1,4GalNAc transferase cDNA was used to perform BLAST analysis of the human HTG and EST divisions of the GenBank®/EMBL database at the National Center for Biotechnology Information [21]. No human EST was found to have sufficient homology with the mouse cDNA. However, we found human genomic sequences (GenBank® accession nos AC069454, AC104587 and AC104974) as well as a partial mRNA sequence (GenBank® accession no. S83275 [27]) producing high sequence similarities to the mouse sequence. These findings enabled us to reconstitute *in silico* the nucleotide sequence of a putative human Sd^a β1,4GalNAc transferase gene. The retrieved sequence was used further to design oligonucleotides for PCR amplification. RT-PCR was performed using first-strand cDNA prepared from Caco-2 cells total RNA since this colon cancer cell line was shown to express the enzymic activity of Sd^a β1,4GalNAc transferase [28]. PCR yielded the expected 1587 and 428 bp amplified fragments, which were subcloned and sequenced. These sequences were found to be 100 % identical to the genomic sequence found in clone AC069454 and to represent almost all the open reading frame (ORF) of the human Sd^a β1,4GalNAc transferase. In order to obtain the full-length cDNA, 5′-RACE was performed from total RNA of Caco-2 cells. Two PCR amplification fragments of 400 and 300 bp were obtained, subcloned and sequenced. Sequence analysis revealed that these two cDNA fragments had a 123 bp common 3′ end but differed at their 5′ ends, suggesting that two different transcripts are present in Caco-2 cells. As shown in Figure 1, the nucleotide sequence corresponding to the longer transcript shows a very short 5′-untranslated sequence of 59 and a 1701 bp ORF beginning at a position compatible with the Kozak consensus sequence (RXXATGG) for translation initiation [29]. It encodes a putative 566-amino-acid polypeptide with a calculated molecular mass of 63.3 kDa. Sequence analysis of this polypeptide showed two putative O-glycosylation sites located in the luminal domain of the polypeptide but, in contrast to the mouse polypeptide, no N-glycosylation site was found. A hydropathy plot of this long form (results not shown) predicted a type II transmembrane protein that is described commonly for Golgi glycosyltransferases. However, this protein shows a very unusual cytosolic domain of 66 amino acid residues in its N-terminal region, a single 23-amino-acid transmembrane domain and a 477-amino-acid catalytic domain. As represented schematically in Figure 2, the nucleotide sequence of the shorter transcript was found to be identical downstream of nucleotide position +193 (numbering as in Figure 1) to the nucleotide sequence of the longer form. However, since their 5′ ends differ, these two different-sized transcripts are expected to encode two translation products with different N-termini. The shorter nucleotide sequence shows a very short 5′-untranslated region of 24 bp and contains an ORF of 1521 bp potentially encoding a 506-amino-acid polypeptide with a calculated molecular mass of 57032 Da. The primary structure of this short polypeptide was found to correspond to that of the mouse Sd^a β1,4GalNAc transferase already described [18] with 74 % amino-acid sequence identities, although the human sequence shows a deletion of four amino-acid residues in the cytoplasmic region. In addition, multiple sequence alignments indicated homologous regions highly conserved between β1,4GalNAc transferases and G_{M2} synthases from various species. It has to be noted that these glycosyltransferases belong to the same GT-12 family of the CAZy classification [30].

During the course of this study, a cDNA clone of human β1,4GalNAc transferase was deposited in GenBank® Nucleotide Sequence Database under the accession no. AF510036.

Gene organization

The gene organization of the human Sd^a β1,4GalNAc transferase was reconstituted from the genomic clones previously identified in the human databanks of GenBank®/EMBL. The gene was found to localize on chromosome 17 (17q23.1), spans over 36 kb of human genomic sequence and comprises at least 12 exons (Figure 3A). As seen in Figure 3(B), the sequences of the exon–intron splice junctions of the gene obey the GT-AG rule [31]. Exons $E1_S$ and $E1_L$ contain the 5′-untranslated regions and almost all of the cytoplasmic domains of each polypeptide isoform. Exon E2, which is common to the two transcripts, encodes the hydrophobic signal anchor and the stem region. Exons 3–11 encode the putative

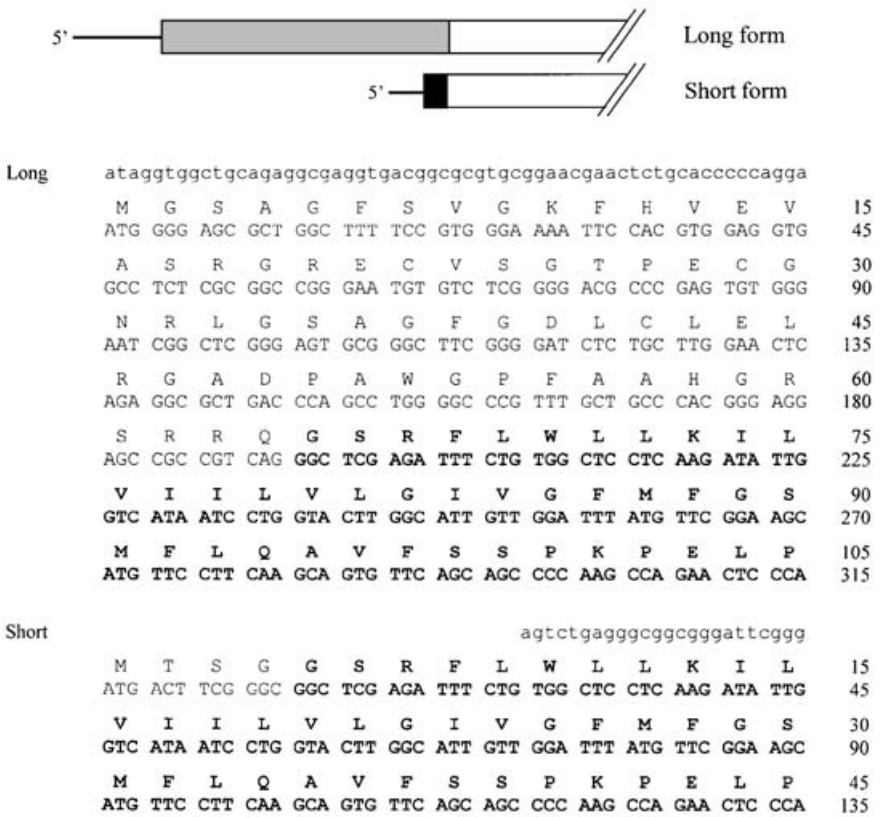


Figure 2 Comparison of the 5′ ends of the long and short forms of the Sd^a β1,4GalNAc transferase mRNAs

The 5′ ends of the Sd^a β1,4GalNAc transferase mRNAs corresponding to the long and short forms are represented schematically. The shaded grey and black boxes represent the divergent sequences. The solid lines represent the untranslated regions. The nucleotide and deduced amino acid sequences are given below and the common region is indicated in bold.

active domain of the enzyme. Exon 11 contains a translation-termination codon and a long 3′-untranslated region of the gene containing several potential polyadenylation signals.

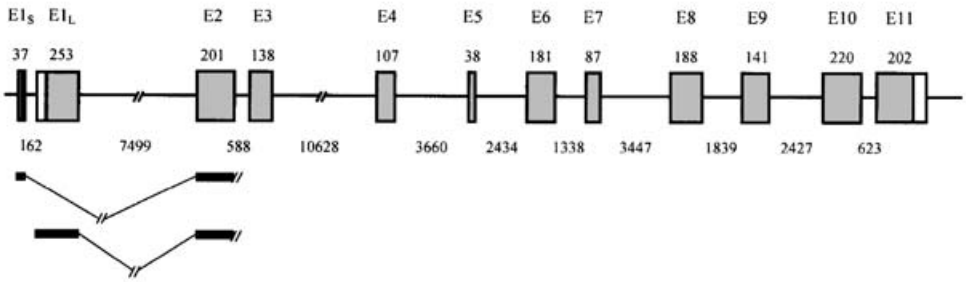
Sd^a β1,4GalNAc transferase gene expression

To determine the expression pattern and the size of Sd^a β1,4GalNAc transferase mRNAs, Northern blot analysis was performed using [α-^{32}P]dCTP-labelled β1,4GalNAc cDNA as a probe. As shown in Figure 4 (top panel), the β1,4GalNAc gene is highly expressed in colon and to a lesser extent in kidney, stomach, ileum and rectum. Almost no signal was obtained in brain, lung, spleen, jejunum, cervix, uterus or placenta. The gene was expressed as five transcripts of 8.8, 6.1, 4.7, 3.8 and 1.65 kb. Consequently, this gene appears to be expressed in a very restrictive manner. However, hybridization of poly(A)$^+$ RNA from 72 different human tissues with the same [α-^{32}P]dCTP-labelled β1,4GalNAc cDNA indicated a basal and low level of expression of the gene in each human tissue (results not shown). Finally, the expression of the Sd^a β1,4GalNAc transferase gene in several human cancer and normal cells lines was also examined by RT-PCR. As shown in Figure 4 (bottom panel), the Sd^a β1,4GalNAc transferase gene is highly expressed in some of the breast cancer lines tested (BT-20 and HBL-100), in normal breast cells (CEMN), promegakaryocytic Dami cells, lymphocytes (Jurkat), keratinocytes, promonocytic cells (HL-60 and THp) and pulmonary cancer NCI cells. It is weakly expressed in others breast cancer lines (MCF-7 and T47-D) and in the gastric cell line AGS. It is not detected in the breast cancer line MDA-MB-231 or in Neuro3 cells. Interestingly, the colon cancer cell line Caco-2 highly expressed the Sd^a β1,4GalNAc transferase gene, whereas the colon cancer cell line HT-29 showed very low levels of this mRNA.

Expression of a soluble form of Sd^a β1,4GalNAc transferase

To verify conclusively that the cloned cDNA we obtained from Caco-2 cells indeed represented human Sd^a β1,4GalNAc transferase, and to facilitate functional analysis of the enzyme, it was desirable to produce a soluble form of the enzyme that could be secreted from the cells. We constructed therefore a truncated cDNA encoding a polypeptide corresponding to the putative catalytic domain deleted of the cytoplasmic and almost the entire transmembrane domains of the protein. This construction included a preprotrypsin signal, a FLAG octapeptide (Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys) fused to the truncated N-terminus of the Sd^a β1,4GalNAc transferase protein, starting at Gly-89 as indicated in Figure 1. This recombinant protein was expressed transiently in Cos-7 cells, leading to the production of a soluble form of the enzyme in the cell-culture medium of transfected cells. As shown in Figure 5, almost no activity was detected in the mock-transfected-cell media whereas a transfer activity of a GalNAc residue on to fetuin appeared in the media of Sd^a β1,4GalNAc transferase-transfected cells collected 48 and 72 h after transfection. As reported in Table 1, the recombinant enzyme efficiently mediated the transfer of *N*-acetylgalactosamine

A



B

Exon	Splice donor	Intron	Splice acceptor	Exon
$E1_S$	CGGGATTCGGG ATG ACT TCG GGC GG **14** M T S G G	gtgagtgt ------- cccaacag	C TCG AGA TTT CTG TGG CTC CTC AAG S R F L W L L K	E2
$E1_L$	CAC GGG AGG AGC CGC CGT CAG GG **194** H G R S R R Q G	gtatgtga ------- cccaacag	C TCG AGA TTT CTG TGG CTC CTC AAG S R F L W L L K	E2
E2	CTC TTT TCC TAC GAT GGA ATC TG **395** L F S Y D G I W	gtgagaga ------- tgtttcag	G CTG TTC CCG AAA AAT CAG TGC AAA L F P K N Q C K	E3
E3	GAA TTT GAA CAC TTT CAG AGG AG **533** E F E H F Q R R	gtaatgcg ------- tcctgcag	A GAA GGG CTG CCC CGC CCA CTG CCC E G L P R P L P	E4
E4	CTG CAC ACG GTT CCC ATC CCA G **640** L H T V P I P	gtaagtac ------- tcctccag	GC CTC CAG TTT GAA GGA CCC GAT GCC G L Q F E G P D A	E5
E5	GGA CCC GAT GCC CCC GTC TAT GAG **678** G P D A P V Y E	gtgagtcc ------- caccctag	GTC ACC CTG ACA GCT TCT CTG GGG V T L T A S L G	E6
E6	CAG CAC CAG AAG GTA GAC ATA G **859** Q H Q K V D I	gtgagagc ------- ccctccag	TG AGT CTG GAG TCC AGG TCC TCA GTG V S L E S R S S V	E7
E7	AAG CTA TAC GAC CCT GGA CCA G **946** K L Y D P G P	gtaaggcc ------- ctgcccag	AG AGG AAG CTC AGA AAC CTG GTT ACC E R K L R N L V T	E8
E8	TAT TAC ACT ATG CCC TTT GGG AAG **1134** Y Y T M P F G K	gtatgtcc ------- atctacag	GGT TGG TTT GCT GGT AGG AAC CTG G W F A G R N L	E9
E9	CTG GAG AAA ACA GAA CTG GAC GTG **1275** L E K T E L D V	gtaaggga ------- ctggctag	GTA GGC GGC AGT GTG CTG GGA AAT V G G S V L G N	E10
E10	CTG CAA CGA GTG GCT CAC TCA G **1495** L Q R V A H S	gtgggaag ------- ctttgcag	AA TTC TTC ATT GAT GGG CTA GGG ACC E F F I D G L G T	E11

Figure 3 Schematic diagram of the genomic organization (A) and exon–intron junctions (B) of the human Sd^a β1,4GalNAc transferase gene

(**A**) Shaded boxes represent coding sequences and open boxes represent untranslated sequences of Sd^a β1,4GalNAc transferase gene. Exons are denoted E1 ($E1_S$ and $E1_L$) to E11 with their size (in bp) indicated above the boxes, and the intron sequences, determined from genomic databases, are shown by the solid lines with their sizes (in bp) indicated below. The length of exon 11 refers to the coding region only. A schematic representation of the 5′ end of the two transcripts is shown below the gene. (**B**) The nucleotide sequences at the intron (lower-case letters) and exon (upper-case letters) junctions are shown. The derived amino acid sequence is shown below the nucleotide sequence. Exons are numbered (bold characters) from the 5′ end starting at the initiator methionine, denoted as +1.

residues on to fetuin, which contains both N- and O-glycans, and on to α_1-acid glycoprotein, which contains N-glycans, but no significant activity was observed towards asialo-glycoproteins. Sd^a β1,4GalNAc transferase exhibited very weak activity towards the other O-glycosylproteins used as acceptor substrates (Table 1). Similarly, the Sd^a β1,4GalNAc transferase used better the sialylated arylglycosides Neu5Acα2-3Galβ1-3GlcNAcβ1-*O*-benzyl and Neu5Acα2-3Galβ1-3GalNAcα1-*O*-benzyl than the corresponding asialo-arylglycosides Galβ1-3GlcNAcβ1-*O*-octyl and Galβ1-3GalNAcα1-*O*-benzyl (Table 1). As shown in Figure 6, the sialylated oligosaccharides Neu5Acα2-3Galβ1-3GlcNAcβ1-3Galβ1-4Glc (see no. 2 in Figure 6), Neu5Acα2-3Galβ1-4Glc (no. 5), Neu5Acα2-3Galβ1-3GlcNAc (no. 8), Neu5Gcα2-3Galβ1-3GalNAc (no. 10) and Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα-*O*-CH_3 (no. 12) were found to be acceptor substrates for the recombinant Sd^a β1,4GalNAc transferase. These data indicated that Sd^a β1,4GalNAc transferase requires the minimal disaccharide structure Neu5Acα2-3Gal found on glycoproteins and oligosaccharides. This galactose residue has to be β1,3-linked to an *N*-acetylglucosamine or *N*-acetylgalactosamine residue or β1,4-linked to glucose or *N*-acetylglucosamine residues. However, it is important to note that the ganglioside G_{M3} is not an acceptor substrate for Sd^a β1,4GalNAc transferase activity (Table 1). In addition, no activity was detected using 3′-sulpholactose or 3′-*O*-methylcarboxyllactose. The kinetic parameters of Sd^a β1,4GalNAc transferase were determined using UDP-GalNAc as the donor substrate and fetuin or Neu5Acα2-3Galβ1-3GlcNAcβ1-*O*-benzyl as acceptor substrate (Table 2). The K_m value of Sd^a β1,4GalNAc transferase for UDP-GalNAc was 110 μM and the apparent K_m values of fetuin and Neu5Acα2-3Galβ1-3GlcNAcβ1-*O*-benzyl were 0.94 and 0.38 mM, respectively.

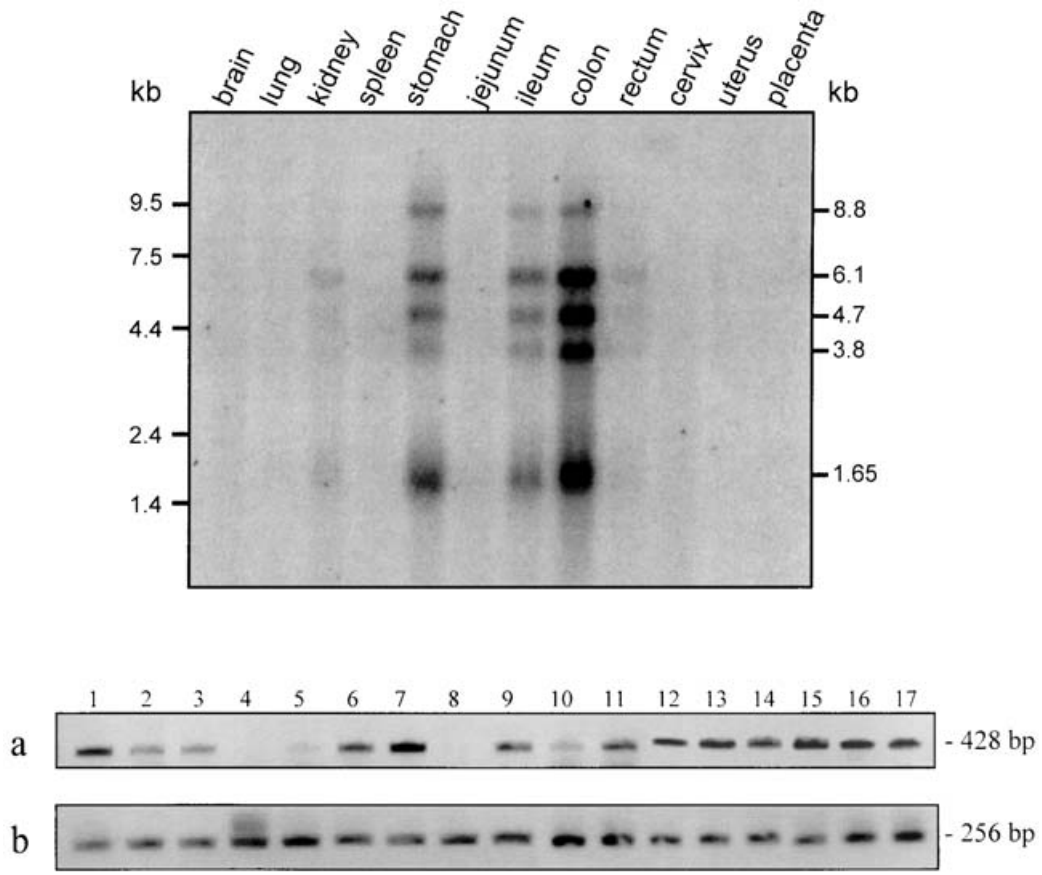


Figure 4 Northern blot analysis of Sd^a β1,4GalNAc transferase gene in various human tissues (top panel), and RT-PCR analysis of the expression of Sd^a β1,4GalNAc transferase gene in various human cell lines (bottom panel)

Top panel: commercially prepared Northern blot with 3 μg of poly(A)$^+$ RNA from various adult human tissues were probed with a 1475 bp [α-^{32}P]dCTP-labelled β1,4GalNAc transferase cDNA as a probe. RNA size markers are indicated on the left side of the blot and the size of the transcripts is indicated on the right side of the blot. Bottom panel: relative expression levels of Sd^a β1,4GalNAc transferase (**a**) and GAPDH (**b**) mRNAs were evaluated by RT-PCR, as described in the Experimental section, among various human cell lines: 1, CEMN; 2, MCF-7; 3, T47-D; 4, MDA-MB-231; 5, HT-29; 6, Caco-2; 7, NCI; 8, Neuro3; 9, Neuro6; 10, AGS; 11, BT-20; 12, Dami; 13, Jurkat; 14, HL-60; 15, THp-1; 16, keratinocytes; 17, HBL-100.

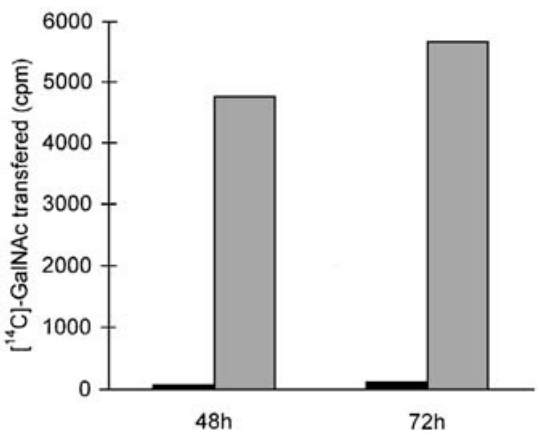


Figure 5 Expression of a soluble FLAG–Sd^a β1,4GalNAc transferase recombinant protein

Media from Cos-7 cells transfected with the expression plasmid Sd^a β1,4GalNAc transferase (grey bars) and from mock-transfected cells (black bars) and collected after 48 or 72 h post-transfection were incubated with 2 mg · ml^{-1} fetuin in enzymic assay conditions as described in the Experimental section.

DISCUSSION

Gene expression and regulation

In this study, we have identified the human Sd^a β1,4GalNAc transferase gene involved in Sd^a/Cad antigen biosynthesis. The Sd^a β1,4GalNAc transferase gene is located on chromosome 17q23.1, spans over 36 kb of genomic sequence and is distributed over at least 12 exons (Figure 3A). As observed with other glycosyltransferase genes, the Sd^a β1,4GalNAc transferase gene exhibits a unique tissue-specific pattern of expression. Northern blot analysis shows clearly that it is particularly highly expressed in colon, which is consistent with the high level of Sd^a β1,4GalNAc transferase activity reported previously in human colon [4,11] and also the high level of expression of Sd^a/Cad antigen in human colon [7]. Interestingly, the level of expression of this gene along the human intestinal tract follows exactly that of Sd^a/Cad antigen [C. Robbe, C. Capon, E. Maes, M. Rousset, A. Zweibaum, J.-P. Zanetta and J.-C. Michalski, unpublished work]. The gene is also expressed to a lesser extent in kidney, stomach, ileum and rectum. No signal for its expression is detected on this Northern blot in brain, lung, spleen, jejunum, cervix, uterus or placenta tissues. However, hybridization of a human multiple-tissue mRNA array indicated a basal level of expression of the gene in all human tissues tested (results not shown). Finally,

Table 1 Acceptor substrate specificity of the recombinant human Sd^a β1,4GalNAc transferase

Various acceptor substrates were incubated in the standard assay mixture described in the Experimental section using transfected Cos-7 cell medium as an enzyme source. Each substrate was used at a concentration of 1 mM for arylglycosides, 2 mg · ml^{-1} for glycoproteins and 0.5 mM for oligosaccharides. [^{14}C]GalNAc incorporation from recombinant Sd^a β1,4GalNAc transferase activity was estimated by subtracting the background of mock-transfected cells from the radioactivity in the presence of exogenous acceptor. Relative rates were calculated as a percentage of the incorporation of GalNAc residues on to Neu5Acα 2-3Galβ1-3GlcNAc. R in a structure indicates the remainder of the N-linked oligosaccharide chain. Bz, benzoyl.

Acceptor	Structure	Relative rate (%)
Glycoproteins		
Fetuin	Neu5Acα2-3Galβ1-3[Neu5Acα2-6]$_{0-1}$GalNAcα1-*O*-Ser/Thr Neu5Acα2-6(3)Galβ1-4GlcNAc-R	2.4
Asialo-fetuin		0.1
α_1-Acid glycoprotein	Neu5Acα2-6(3)Galβ1-4GlcNAc-R	1.4
Asialo-α_1-acid glycoprotein		0
Bovine submaxillary mucin	GlcNAcβ1-3[Neu5Acα2-6]GalNAcα1-*O*-Ser/Thr	0.1
Ovine submaxillary mucin	Neu5Acα2-6GalNAcα1-*O*-Ser/Thr	0.3
Arylglycosides		
Type I	Neu5Acα2-3Galβ1-3GlcNAc-*O*-Bz	12.7
	Galβ1-3GlcNAc-*O*-octyl	0.8
Type II	Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα1-*O*-CH$_3$	19.5
	Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-*O*-CH$_3$	0
Type III	Neu5Acα2-3Galβ1-3GalNAc-*O*-Bz	8
	Galβ1-3GalNAc-*O*-Bz	0.7
Oligosaccharides		
3′-Sialyl-lactose	Neu5Acα2-3Galβ1-4Glc	35
Lactose	Galβ1-4Glc	0
N-Acetyl-lactosamine	Galβ1-4GlcNAc	0
3′-Sulpholactose	$Na^+SO_3^-$ -3Galβ1-4Glc	0
3′-*O*-Methylcarboxyl-lactose	Na^+COO^--3Galβ1-4Glc	0
3′-Sialyl-neo-*N*-acetyl-lactosamine	Neu5Acα2-3Galβ1-3GlcNAc	100 (407)*
Neo-*N*-acetyl-lactosamine	Galβ1-3GlcNAc	0
3′-Sialyl-triose (type III)	Neu5Gcα2-3Galβ1-3GalNAc	21
3′-Sialyl-lacto-*N*-tetraose	Neu5Acα2-3Galβ1-3GlcNAcβ1-3Galβ1-4Glc	66
6′-Sialyl-lacto-*N*-tetraose	Galβ1-3[Neu5Acα2-6]GlcNAcβ1-3Galβ1-4Glc	0
Lacto-*N*-tetraose	Galβ1-3GlcNAcβ1-3Galβ1-4Glc	0
Gangliosides		
G_{M3}	Neu5Acα2-3Galβ1-4Glcβ1-Cer	0
G_{D3}	Neu5Acα2-8Neu5Acα2-3Galβ1-4Glcβ1-Cer	0
G_{D1a}	Neu5Acα2-3Galβ1-3GalNAcβ1-4Galβ1-4Glcβ1-Cer	0

* Actual activity is shown in parentheses in pmol · h^{-1} · μl^{-1}.

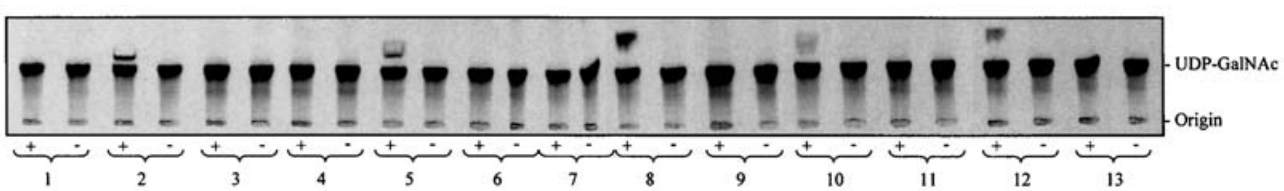


Figure 6 Sd^a β1,4GalNAc transferase activity towards free oligosaccharides

The structures of these oligosaccharides are shown in Table 1. Oligosaccharides (0.5 mM) were incubated with transfected Cos-7 cell medium as an enzyme source (+) in the standard assay mixture described in the Experimental section. Mock-transfected Cos-7 cell medium was used as controls (−). The resulting products were analysed by TLC (Silicagel 60) in ethanol/butan-1-ol/pyridine/water/acetic acid (100:10:10:30:3, by vol.). 1, Lacto-*N*-tetraose; 2, 3′-sialyl-lacto-*N*-tetraose; 3, 6′-sialyl-lacto-*N*-tetraose, 4, lactose; 5, 3′-sialyl-lactose; 6, 3′-sulpholactose; 7, 3′-*O*-methylcarboxylactose; 8, 3′-sialyl-neo-*N*-acetyl-lactosamine; 9, neo-*N*-acetyl-lactosamine; 10, Neu5Gcα2-3Galβ1-3GalNAc; 11, *N*-acetyl-lactosamine; 12, Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα1-*O*-CH$_3$; 13, Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα1-*O*-CH$_3$.

using RT-PCR we also noticed a high level of expression of the Sd^a β1,4GalNAc transferase gene in differentiated Caco-2 cells, whereas only low levels were found in non-differentiated Caco-2 cells (results not shown). These findings support the conclusion that the Sd^a β1,4GalNAc transferase gene could be a marker of colonic cell differentiation. With regards to the expression pattern of this gene and subsequently of the enzyme, previous studies have shown that CT antibodies define a cell-surface oligosaccharide expressed on murine cytotoxic T-lymphocyte subpopulations [33] identical, at least in part, to the human Sd^a/Cad antigen [34]. The murine Sd^a β1,4GalNAc transferase cDNA was isolated from a cytotoxic T-lymphocyte cell cDNA library and it has been shown that it determines the expression of the T-lymphocyte-specific CT oligosaccharide differentiation antigen [18]. Moreover, non-cytotoxic T-cell lines generally have shown low levels of Sd^a β1,4GalNAc transferase activity [35]. These studies in human lymphocytes are currently under progress in our laboratory but our preliminary data concerning the expression of the human Sd^a β1,4GalNAc transferase gene in various human lymphocytes did not point to particular overexpression of the gene in these

Table 2 Determination of the kinetic parameters of the recombinant Sd^a β1,4GalNAc transferase

Incubations were performed using various substrate concentrations: 1–1000 μM of the donor substrate (UDP-[^{14}C]GalNAc), 0.1–5 mM Neu5Acα2-3Galβ1-3GlcNAcβ1-*O*-Bz, 0.1–5 mg · ml^{-1} fetuin. The K_m of fetuin was calculated on the basis of the GalNAc content (40 nmol · mg^{-1}). Note that UDP-[^{14}C]GalNAc is the donor of GalNAc residues 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 (h^{-1} × 10^3)
Fetuin	0.94	35.43	37.65
Neu5Acα2-3Galβ1-3GlcNAc-*O*-Bz	0.38	12.40	32.63
UDP-[^{14}C]GalNAc	0.11	0.44	–

cells (results not shown). The Sd^a β1,4GalNAc transferase gene is expressed as five transcripts of 8.8, 6.1, 4.7, 3.8 and 1.65 kb. The longest ORF we obtained represents 1701 bp with a 5′-untranslated region of 59 bp, which suggests a variable-length 3′-untranslated sequence ranging from 110 to 7040 nucleotides. Analysis of the 3′-untranslated region of the human Sd^a β1,4GalNAc transferase gene showed the presence of several polyadenylation signals that could account for these multiple transcript sizes detected on the Northern blot (results not shown). Here we describe two Sd^a β1,4GalNAc transferase cDNAs with two different 5′ ends arising from two distinct exons, named $E1_S$ and $E1_L$, that are located 162 bp apart from each other on human chromosome 17. These two transcripts result from alternative splicing and alternative promoter usage. The occurrence of multiple transcripts arising from a unique gene has been described for many other glycosyltransferase genes, such as for the β1,4-galactosyltransferase β4GT-I [36,37], the α2,6-sialyltransferase ST6Gal I [38–40], the α2,3-sialyltransferase ST3Gal IV [41] and G_{M2}/G_{D2} synthase [42]. The biological significance of the alternative use of two exons in this gene remains to be investigated. Very preliminary analysis of the two upstream genomic sequences identified in the databanks indicates the absence of typical TATA and CCAAT boxes coupled with the presence of GC boxes. In addition, potential *trans*-acting-factor-binding sites, such as SP1, cAMP-response-element-binding protein (CREB) and activator protein 1 (AP1), as well as composite regulatory elements are located mainly upstream of exon $E1_S$. Future transcription-regulation studies will help to elucidate the biological significance of each transcript produced from two closely spaced promoters.

Sequence analysis of the protein

The very unusual feature of this Sd^a β1,4-GalNAc transferase gene is not only to produce two transcripts with two different transcriptional sites but also to potentially give rise to two protein isoforms exhibiting very different N-termini. The eukaryotic glycosyltransferases located in the Golgi apparatus are usually composed of a short N-terminal cytoplasmic tail, a transmembrane domain, a stem region of variable length and a large C-terminal globular catalytic domain [43]. The predicted primary sequences of this enzyme have the overall profile of a type II transmembrane protein as described previously for other glycosyltransferases. However, the long transcript encodes a protein presenting an extraordinarily long cytoplasmic tail of 66 amino acid residues, whereas translation of the short transcript predicts the biosynthesis of a protein with a very short cytoplasmic tail (six amino acid residues) homologous to the mouse sequence described previously [18]. In addition, further database searches of the mouse genome indicate the presence of a potential exon corresponding to the human exon $E1_L$ with an ATG codon lying in a favourable consensus region for translation [29]. However, translation from this ATG codon leads to three stop codons within this exon, thus preventing the synthesis of the long form of the protein in mouse tissues. The transmembrane domain and catalytic domain remain identical for the two proteins. ST3Gal IV and β4GT-I genes have been shown to give rise to various protein isoforms showing minor alterations in their cytoplasmic tail. However, to date, there is no evidence that such sequence modification in a glycosyltransferase cytoplasmic tail results in an altered cellular distribution or enzymic activity. Recent publications have pointed to unusual subcellular localization of glycosyltransferases, which normally localize to the Golgi apparatus [44], and interaction of the cytoplasmic domain of β4GT-I with the Src-suppressed C kinase substrate, a modulator of cytoarchitecture [45], provides the first example of such a connection [46]. Of interest is the lack of potential N-glycosylation site detected in the human β1,4-GalNAc transferase protein sequence, although the murine Sd^a β1,4-GalNAc transferase showed one potential N-glycosylation site at Asn-389 [18]. Multiple sequence alignments indicate that the human Sd^a β1,4GalNAc transferase polypeptide displayed 74% sequence identity with the murine Sd^a/Cad synthase (GenBank® accession no. L30104) already described [18], but also 42% identity with human G_{M2} synthase (GenBank® accession no. M83651) [47], with the highest sequence identity found in the catalytic C-terminal domain (results not shown). This finding is not surprising since this latter enzyme also transfers *N*-acetylgalactosamine residues in β1,4-linkage to the Neu5Acα2-3Galβ1-4Glc-R moiety of gangliosides. Notably, these enzymes share a common region, SQVTTKYVLWVDDDF, containing the acidic motif DXD [48,49], found to be conserved in the 13 glycosyltransferase families of the CAZy classification [30] using a UDP-sugar as the donor substrate. This DXD motif has been implicated in the binding of divalent cations [50]. The DXD motif of the G_{M2} synthase, has been shown to be required for the enzymic activity of the G_{M2} synthase but not critical for the binding of UDP [50]. Database searches and sequence alignment with other β4-*N*-acetylgalactosaminyltransferases such as the chondroitin synthases (GenBank® accession nos AB023207 and AB081516; CAZy family GT-31) [51,52] did not show significant sequence identities. In addition, the motif GWGGED found to be conserved in the β4Gal-T family, chondroitin synthase family and UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase (pp-GalNAc-T) family (CAZy family GT-27) [53–55] was not found in the peptide sequence of the members of the GT-12 CAZy family.

Substrate specificities and kinetic properties of the enzyme

In this paper, we have demonstrated that the soluble form of the human Sd^a β1,4GalNAc transferase is catalytically active. This enzyme requires the presence of the disaccharide structure Neu5Acα2-3Galβ to efficiently transfer GalNAc residues in β1,4-linkage to the galactose residue. We have shown clearly that the 3′-sialylated acceptor Neu5Acα2-3Galβ1-4GlcNAcβ1-2Manα-*O*-CH_3 was used as an acceptor by the enzyme whereas the 6′-sialylated acceptor Neu5Acα2-6Galβ1-4GlcNAcβ1-2Manα-*O*-CH_3 was not. The presence of a sialic acid residue that can be either Neu5Ac or Neu5Gc in an α2,3-linkage is a requirement whereas the charge is not, since 3′-sulpholactose or 3′-*O*-methylcarboxyl-lactose were not acceptor substrates for the enzyme (Table 1). We have shown clearly for the first time that

the best transfer activity was obtained when the free oligosaccharide Neu5Acα2-3Galβ1-3GlcNAc was used as an acceptor. Further substitution of the *N*-acetylglucosamine residue with an *O*-benzyl group or a Galβ1-4Glc group decreased the enzyme efficiency significantly. Similarly, the free trisaccharide Neu5Acα2-3Galβ1-3GalNAc and Neu5Acα2-3Galβ1-4Glc equally served as acceptors for the recombinant enzyme, but further substitution of the *N*-acetylgalactosamine with an *O*-benzyl group decreased the transfer activity of the enzyme. No activity was found on glycophorin used as a potential acceptor substrate (results not shown). Glycophorin exhibits 15 *O*-glycans bearing the Neu5Acα2-3Galβ1 structure that are clustered strictly on the peptide backbone, and thus may be hardly accessible for the Sd^a β1,4GalNAc transferase. These data are in accordance with those obtained in the human kidney [12] where the Sd^a β1,4GalNAc transferase was unable to use glycophorin as an acceptor substrate. As mentioned previously for the enzyme activity described in human gastrointestinal mucosa [27], the Sd^a β1,4-GalNAc transferase did not use gangliosides as acceptor substrates, not even G_{D1a}, which shows a peripheral Neu5Acα2-3Galβ structure. When the enzyme activity was determined as a function of UDP-GalNAc concentration using fetuin as an acceptor substrate, the K_m was 110 μM, a value in accordance with those already obtained for the human enzyme found in urine [14] or human colon [11]. Taken together, these different results indicate that the Sd^a β1,4-GalNAc transferase activity strictly requires a peripheral Neu5Acα2-3Galβ disaccharide sequence, but is also greatly influenced by the backbone (including the aglycon moiety) to which this sequence is linked.

We thank Dr Carmen Odberg-Ferragut for helpful discussion. We are also very thankful to Dr Claudine Augé (URA CNRS 462, Orsay, France) for the synthesis of Neu5Acα2-3Galβ1-4GlcNAcβ1-2Man-*O*-CH_3 and Neu5Acα2-6Galβ1-4GlcNAcβ1-2Man-*O*-CH_3, to Dr Luigi Lay and Dr Anna Rencurosi (University of Milan, Milan, Italy) for the gifts of 3′-sulpholactose and 3′-*O*-methylcarboxyl-lactose, to Dr Jean-Claude Michalski, Dr Gérard Strecker and Dr Frédéric Chirat (UMR CNRS-USTL 8576, Villeneuve d'Ascq, France) for the gift of acceptor substrates purified from human milk and urine, to Dr Guillemette Huet for the gift of Caco-2 cells, to Dr Isabelle Van Seuningen and Professor Annick Pierce for cDNAs from various cell types. We acknowledge the expert technical assistance of Anne-Marie Mir in cell culture. The University of Sciences et Technologies de Lille, the European Carbohydrate Research Network Glycotrain, the CNRS réseau "G3" and the association pour la Recherche sur le Cancer (grant n°5469) have supported this work. M.-D. M. has a fellowship from the European Glycotrain Network.

REFERENCES

1 MacVie, S. I., Morton, J. A. and Pickles, M. M. (1967) The reaction and inheritance on a new blood group antigen, Sd^a. Vox Sang. **13**, 485–492
2 Renton, P. H., Howell, P., Ikin, E., Giles, C. and Goldsmith, K. L. G. (1967) Anti-Sd^a, a new blood group antibody. Vox Sang. **13**, 493–501
3 Morton, J. A., Pickles, M. M. and Terry, A. M. (1970) The Sd^a blood group antigen in tissues and body fluids. Vox Sang. **19**, 472–482
4 Morton, J. A., Pickles, M. M. and Vanhegan, R. I. (1988) The Sd^a antigen in the human kidney and colon. Immunol. Invest. **17**, 217–224
5 Donald, A. S., Yates, A. D., Soh, C. P., Morgan, W. T. and Watkins, W. M. (1983) A blood group Sd^a-active pentasaccharide isolated from Tamm-Horsfall urinary glycoprotein. Biochem. Biophys. Res. Commun. **115**, 625–631
6 Blanchard, D., Cartron, J. P., Fournet, B., Montreuil, J., van Halbeek, H. and Vliegenthart, J. F. (1983) Primary structure of the oligosaccharide determinant of blood group Cad specificity. J. Biol. Chem. **258**, 7691–7695
7 Capon, C., Maes, E., Michalski, J. C., Leffler, H. and Kim, Y. S. (2001) Sd^a-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. Biochem. J. **358**, 657–664
8 Dohi, T., Ohta, S., Hanai, N., Yamaguchi, K. and Oshima, M. (1990) Sialylpentaosylceramide detected with anti-G_{M2} monoclonal antibody. Structural characterization and complementary expression with G_{M2} in gastric cancer and normal gastric mucosa. J. Biol. Chem. **265**, 7880–7785
9 Serafini-Cessi, F. and Dall'Olio, F. (1983) Guinea-pig kidney β-N-acetyl-galactosaminyltransferase towards Tamm-Horsfall glycoprotein. Requirement of sialic acid in the acceptor for transferase activity. Biochem. J. **215**, 483–489
10 Serafini-Cessi, F., Dall'Olio, F. and Malagolini, N. (1986) Characterization of N-acetyl-β-D-galactosaminyltransferase from guinea-pig kidney involved in the biosynthesis of Sd^a antigen associated with Tamm-Horsfall glycoprotein. Carbohydr. Res. **151**, 65–76
11 Malagolini, N., Dall'Olio, F., Di Stefano, G., Minni, F., Marrano, D. and Serafini-Cessi, F. (1989) Expression of UDP-GalNAc:NeuAcα2,3Galβ-R β1,4 (GalNAc to Gal) N-acetylgalactosaminyltransferase involved in the synthesis of Sd^a antigen in human large intestine and colorectal carcinomas. Cancer Res. **49**, 6466–6470
12 Piller, F., Blanchard, D., Huet, M. and Cartron, J. P. (1986) Identification of an α-NeuAc-(2-3)-β-D-galactopyranosyl N-acetyl-β-D-galactosaminyltransferase in human kidney. Carbohydr. Res. **149**, 171–184
13 Takeya, A., Hosomi, O. and Kogure, T. (1987) Identification and characterization of UDP-GalNAc: NeuAcα2-3Galβ1-4Glc(NAc)β1-4 (GalNAc to Gal) N-acetyl-galactosaminyltransferase in human blood plasma. J. Biochem. (Tokyo) **101**, 251–259
14 Serafini-Cessi, F., Malagolini, N. and Dall'Olio, F. (1988) Characterization and partial purification of β-N-acetylgalactosaminyltransferase from urine of Sd^{a+} individuals. Arch. Biochem. Biophys. **266**, 573–582
15 Dall'Olio, F., Malagolini, N. and Serafini-Cessi, F. (1987) Tissue distribution and age-dependent expression of β-4-N-acetylgalactosaminyl-transferase in guinea-pig. Biosci. Rep. **7**, 925–932
16 Dall'Olio, F., Malagolini, N., Di Stefano, G., Ciambella, M. and Serafini-Cessi, F. (1990) Postnatal development of rat colon epithelial cells is associated with changes in the expression of the β1,4-N-acetylgalactosaminyltransferase involved in the synthesis of Sd^a antigen of α2,6-sialyltransferase activity towards N-acetyl-lactosamine. Biochem. J. **270**, 519–524
17 Malagolini, N., Dall'Olio, F. and Serafini-Cessi, F. (1991) UDP-GalNAc:NeuAcα2,3Galβ-R (GalNAc to Gal) β1,4-N-acetylgalactosaminyltransferase responsible for the Sd^a specificity in human colon carcinoma Caco-2 cell line. Biochem. Biophys. Res. Commun. **180**, 681–686
18 Smith, P. L. and Lowe, J. B. (1994) Molecular cloning of a murine N-acetyl-galactosamine transferase cDNA that determines expression of the T lymphocyte-specific CT oligosaccharide differentiation antigen. J. Biol. Chem. **269**, 15162–15171
19 Rencurosi, A., Poletti, L., Guerrini, M., Russo, G. and Lay, L. (2002) Human milk oligosaccharides: an enzymatic protection step simplifies the synthesis of 3′- and 6′-O-sialyllactose and their analogues. Carbohydr. Res. **337**, 473–483
20 Pinto, M., Robine-Leon, S., Appay, M. D., Kedinger, M., Triadou, N., Dussaulx, E., Lacroix, B., Simon-Assmann, P., Haffen, K., Fogh, J. and Zweibaum, A. (1983) Enterocyte-like differentiation and polarization of the human colon carcinoma cell line Caco-2 in culture. Biol. Cell. **47**, 323–330
21 Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., 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
22 Hofmann, K. and Stoffel, W. (1993) A database membrane spanning protein segments. Biol. Chem. Hoppe-Seyler **347**, 166
23 Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj. J. **15**, 115–130
24 Sanger, F., Nicklen, S. and Coulsen, A. R. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
25 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-2deoxy-α-D-galactopyranoside treated HT-29 cells. Glycobiology **10**, 565–575
26 Recchi, M. A., Hebbar, M. Hornez, L., Harduin-Lepers, A., Peyrat, J. P. and Delannoy, P. (1998) Multiplex reverse transcription polymerase chain reaction assessment of sialyltransferase expression in human breast cancer. Cancer Res. **58**, 1066–4070
27 Dohi, T., Yuyama, Y., Natori, Y., Smith, P. L., Lowe, J. B. and Oshima, M. (1996) Detection of N-acetylgalactosamnyltransferase mRNA which determines expression of Sd^a blood group carbohydrate structure in human gastrointestinal mucosa and cancer. Int. J. Cancer. **67**, 626–631
28 Serafini-Cessi, F., Malagolini, N., Guerrini, S. and Turrini, I. (1995) A soluble form of Sd^a-β1,4-N-acetylgalactosaminyl-transferase is released by differentiated human colon carcinoma Caco-2 cells. Glycoconj. J. **12**, 773–779
29 Kozak, M. (1992) Regulation of translation in eukaryotic systems. Annu. Rev. Cell Biol. **8**, 197–225
30 Coutinho, P. M. and Henrissat, B. (1999) Carbohydrate-active enzymes: an integrated database approach. In Recent Advances in Carbohydrate Bioengineering (Gilbert, H. J., Davies, G., Henrissat, B. and Svensson, B., eds.), pp. 3–12, The Royal Society of Chemistry, Cambridge

31 Shapiro, M. B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. **15**, 7155–7174

32 Reference deleted

33 Lefrançois, L. and Bevan, M. J. (1985) Functional modifications of cytotoxic T-lymphocyte T200 glycoprotein recognized by monoclonal antibodies. Nature (London) **314**, 449–452

34 Conzelmann, A. and Kornfeld, S. (1984) β-Linked N-acetylgalactosamine residues present at the nonreducing termini of O-linked oligosaccharides of a cloned murine cytotoxic T lymphocyte line are absent in a *Vicia villosa* lectin-resistant mutant cell line. J. Biol. Chem. **259**, 12528–12535

35 Conzelmann, A. and Bron, C. (1987) Expression of UDP-N-acetylgalactosamine: β-galactose β1,4-N-acetylgalactosaminyltransferase in functionally defined T-cell clones. Biochem. J. **242**, 817–824

36 Shaper, N. L., Hollis, G. F., Douglas, J. G., Kirsch, I. R. and Shaper, J. H. (1988) Characterization of the full-length cDNA for murine β-1,4-galactosyltransferase. Novel features at the 5′-end predict two translational start sites at two in-frame AUGs. J. Biol. Chem. **263**, 10420–10428

37 Russo, R. N., Shaper, N. L. and Shaper, J. H. (1990) Bovine β1,4-galactosyltransferase: two sets of mRNA transcripts encode two forms of the protein with different amino-terminal domains. *In vitro* translation experiments demonstrate that both the short and the long forms of the enzyme are type II membrane-bound glycoproteins. J. Biol. Chem. **265**, 3324–3331

38 Wen, D. X., Svensson, E. C. and Paulson J. C. (1992) Tissue-specific alternative splicing of the β-galactoside α2,6-sialyltransferase gene. J. Biol. Chem. **267**, 2512–2518

39 Aasheim, H. C., Aas-Eng, D. A., Deggerdal, A., Blomhoff, H. K., Funderud, S. and Smeland, E. B. (1993) Cell-specific expression of human β-galactoside α2,6-sialyltransferase transcripts differing in the 5′ untranslated region. Eur. J. Biochem. **213**, 467–475

40 Wang, X., Vertino, A., Eddy, R. L., Byers, M. G., Jani-Sait, S. N., Shows, T. B. and Lau, J. T. (1993) Chromosome mapping and organization of the human β-galactoside α2,6-sialyltransferase gene. Differential and cell-type specific usage of upstream exon sequences in B-lymphoblastoid cells. J. Biol. Chem. **268**, 4355–4361

41 Kitagawa, H. and Paulson, J. C. (1994) Cloning of a novel α2,3-sialyltransferase that sialylates glycoprotein and glycolipid carbohydrate groups. J. Biol. Chem. **269**, 1394–1401

42 Furukawa, K., Soejima, H., Niikawa, N. and Shiku, H. (1996) Genomic organization and chromosomal assignment of the human β1,4-N-acetylgalactosaminyltransferase gene. Identification of multiple transcription units. J. Biol. Chem. **271**, 20836–20844

43 Paulson, J. C., Weinstein, J. and Schauer, A. (1989) Tissue-specific expression of sialyltransferases. J. Biol. Chem. **264**, 10931–10934

44 Berger, E. G. (2002) Ectopic localizations of Golgi glycosyltransferases. Glycobiology **12**, 29–36

45 Gelman, I. H., Lee, K., Tombler, E., Gordon, R. and Lin, X. (1998) Control of cytoskeletal architecture by src-supressed C kinase substrate, SSeCKS. Cell Motil. Cytoskel. **41**, 1–17

46 Wassler, M. J., Foote, C. I., Gelman, I. H. and Shur, B. D. (2001) Functional interaction between the SSeCKS scaffolding protein and the cytoplasmic domain of β1,4-galactosyltransferase. J. Cell Sci. **114**, 229–2300

47 Nagata, Y., Yamashiro, S., Yodoi, J., Lloyd, K. O. and Furukawa, K. (1992) Expression cloning of β1,4 N-acetylgalactosaminyltransferase cDNAs that determine the expression of G_{M2} and G_{D2} gangliosides. J. Biol. Chem. **267**, 12082–12089

48 Breton, C., Bettler, E., Joziasse, D. H., Geremia, R. A. and Imberty, A. (1998) Sequence-function relationships of prokariotic and eukariotic galactosyltransferases. J. Biochem. (Tokyo) **123**, 1000–1009

49 Wiggins, C. A. R. and Munro, S. (1998) Activity of the yeast MNN1 α-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases. Proc. Natl. Acad. Sci. U.S.A. **95**, 7945–7950

50 Gastinel, L. N., Cambillau, C. and Bourne, Y. (1999) Crystal structures of the bovine β4galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose. EMBO J. **18**, 3546–3557

51 Li, J., Rancour, D. M., Allende, M. L., Worth, C. A., Darling, D. S., Gilbert, J. B., Menon, A. K., and Young, Jr, W. W. (2001) The DXD motif is required for G_{M2} synthase activity but is not critical for nucleotide binding. Glycobiology **11**, 217–229

52 Kitagawa, H., Uyama, T. and Sugahara, K. (2001) Molecular cloning and expression of a human chondroitin synthase. J. Biol. Chem. **276**, 38721–38726

53 Uyama, T., Kitagawa, H., Tamura, Ji. J. and Sugahara, K. (2002) Molecular cloning and expression of human chondroitin N-acetylgalactosaminyltransferase: the key enzyme for chain initiation and elongation of chondroitin/dermatan sulfate on the protein linkage region tetrasaccharide shared by heparin/heparan sulfate. J. Biol. Chem. **277**, 8841–8846

54 Meurer, J. A., Naylor, J. M., Baker, C. A., Thomsen, D. R., Homa, F. L. and Elhammer, A. P. (1995) cDNA cloning, expression, and chromosomal localization of a human UDP-GalNAc:polypeptide, N-acetylgalactosaminyltransferase. J. Biochem. (Tokyo) **118**, 568–574

55 Gotoh, M., Yada, T., Sato, T., Akashima, T., Iwasaki, H., Mochizuki, H., Inaba, N., Togayachi, A., Kudo, T., Watanabe, H., Kimata, K. and Narimatsu, H. (2002) Molecular cloning and characterization of a novel chondroitin sulfate glucuronyltransferase that transfers glucuronic acid to N-acetylgalactosamine. J. Biol. Chem. **277**, 38179–38188

Received 6 December 2002/31 March 2003; accepted 4 April 2003
Published as BJ Immediate Publication 4 April 2003, DOI 10.1042/BJ20021892