Molecular Medicine 8(1): 42–55, 2002

Identification and Functional Characterization of a Human GalNAc α2,6-Sialyltransferase with Altered Expression in Breast Cancer

Georgia Sotiropoulou,[1]* Mari Kono,[3] Anthony Anisowicz,[2] Goran Stenman,[4] Shuichi Tsuji,[3] and Ruth Sager[2†]

[1]Department of Pharmacy, School of Health Sciences, University of Patras, Rion, Patras, Greece
[2]Division of Cancer Genetics, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA
[3]Laboratory for Molecular Glycobiology, Frontier Research Program, The Institute of Physical and Chemical Research (RIKEN), Wako, Saitama, Japan
[4]Laboratory of Cancer Genetics, Department of Pathology, Goteborg University, Goteborg, Sweden

Contributed by A. B. Pardee. Accepted December 5, 2001

Abstract

Background: We sought to identify genes with altered expression during human breast cancer progression by applying mRNA comparisons of normal and tumor mammary cell lines with increasingly malignant phenotypes. The gene encoding a new sialyltransferase (STM) was found to be down-regulated in tumor cells. Abnormal expression and enzymatic activities of sialyltransferases in tumor cells result in the formation of tumor-associated carbohydrate antigens that can be used for the better understanding of the disease process and are applied for tumor diagnosis and immunotherapy. Altered glycosylation patterns of the MUC1 mucin, in particular, is a target antigen for immunotherapy of breast and other cancers.
Materials and Methods: Total RNAs from multiple normal mammary epithelial cell strains and tumor cell lines were compared by differential display and the differential expression of selected cDNAs was confirmed by Northern analyses. Recombinant STM was expressed in COS-7 cells. The substrate and linkage specificity of STM was examined using various oligosaccharides and *O*-glycosylated proteins as acceptor substrates. The chromosomal localization of the SIATL1 gene was assigned by somatic cell hybrid analysis.
Results: A human sialyltransferase gene was identified by differential display as being down-regulated in breast tumor cell lines as compared to normal mammary epithelial cell strains, and the corresponding full-length cDNA (*stm*) was cloned. The encoded protein of 374 amino acid residues contained the *L*- and *S*-sialylmotifs, two catalytic regions conserved in all functional sialyltransferases. Recombinant STM is an active GalNAc α2,6-sialyltransferase with Galβ1,3 GalNAc-*O*-Ser/Thr and (+/−Neu5Acα2,3) Galβ1,3GalNAc-*O*-Ser/Thr acceptor specificity. The SIATL1 gene, encoding STM, was mapped to the long arm of human chromosome 17 at q23-qter, a region that is nonrandomly deleted in human breast cancers. However, Southern analyses indicated that SIATL1 is usually not grossly rearranged in breast tumors. Northern analyses showed that the gene was widely expressed in normal human tissues, as well as in normal breast and prostate epithelial cell lines, but significantly down-regulated or absent in corresponding tumor cell lines.
Conclusions: Our findings suggest that aberrant expression of STM sialyltransferase in tumors could be a feature of the malignant phenotype. In breast cancers, the MUC1 mucin is overexpressed and contains shorter *O*-glycans as compared to the normal mucin. Because STM catalyzes the synthesis of *O*-glycans, cloning and characterization of its substrate specificity will contribute to the understanding of the molecular mechanisms underlying the aberrant glycosylation patterns of *O*-glycans and the formation of mucin-related antigens in human breast cancers.

Introduction

The growth and dissemination of human tumors is a multistage process involving aberrant functions of the tumor cell, which result from changes in the expression of multiple genes. The concept of RNA genetics in cancer (1) underscores that the identification of specific changes in gene expression, underlying the switch of a normal cell to a malignant phenotype, is essential for the prognosis and staging of malignant tumors, as well as for the design of effective anticancer therapies. Subtractive hybridization and differential display (2) methods of differential expression cloning were used to identify genes with aberrant expression during human mammary carcinogenesis. Genes with various cellular functions and altered expression in normal mammary epithelial cells versus corresponding tumor cells were cloned and functionally characterized (3–5). The gene encoding STM, a new human sialyltransferase, was identified among the differentially expressed

*Address correspondence and reprint requests to: G. Sotiropoulou, Department of Pharmacy, University of Patras, Rion, 26500 Patras, Greece. Phone: +30-610-996384; fax: +30-610-997658; e-mail: G.Sotiropoulou@upatras.gr
†Deceased.

This manuscript is dedicated (by G.S., A.A., and G.S.) to the memory of our mentor and friend Professor Ruth Sager.

genes based on its decreased or absent mRNA expression in breast tumor cell lines.

Sialyltransferases constitute a family of glycosyltransferases that catalyze the posttranslational transfer of sialic acid (*N*-acetylneuraminic acid) to acceptor oligosaccharide substrates at terminal positions on glycoproteins and glycolipids (6). It is estimated that the human genome encodes more than 20 different sialyltransferases required to synthesize all known sialo-oligosaccharide structures present in mammalian cells, but only 16 distinct human sialyltransferase cDNAs have been cloned (7–9). Originally, sialyltransferases were biochemically purified and their cDNAs were cloned using N-terminal sequences (10,11). Comparison of the obtained cDNA sequences revealed two highly conserved regions, termed the *L*- and *S*-sialylmotifs, that participate in substrate binding (12,13). Subsequently, several sialyltransferases were cloned by PCR using degenerate primers designed within the sialylmotifs (14,15) or by expression cloning (16–18). Cloning of the gene encoding STM by differential display adds an entirely different approach to the identification of novel sialytransferases with putative functional significance in disease-related processes.

Sialyltransferases differ in their substrate specificity and tissue distribution, and they are classified into four families according to the carbohydrate linkages they synthesize: the ST3Gal-, ST6Gal-, ST6GalNAc-, and ST8Sia-families. The members of each family exhibit strong activity toward certain acceptor groups, although the substrate specificities of these enzymes overlap; one linkage can be synthesized by multiple enzymes (8,9).

Specific terminal glycosylation sequences are important recognition determinants in cell–cell interactions, protein targeting, and host–pathogen interactions (19,20), as well as important regulators of cell growth, cell maturation, and differentiation (21,22). Pathologic conditions are often associated with modified terminal sugar structures induced by changes in expression of the glycosyltransferases responsible for their synthesis. Thus, tumor cells are characterized by aberrant expression and enzymatic activities of sialyltransferases (23–27). The appearance of new glycoconjugate structures on the surface of tumor cells and the accumulation of precursors due to blocked synthesis result in the formation of tumor-associated carbohydrate antigens (24–27). In particular, sialic acid has a masking effect on tumor antigens, because hypersialylation of the tumor cell surface results in decreased susceptibility of tumor cells to natural killer cells with a compromising effect on the host immune system. In addition, sialic acids are involved in a variety of biological processes (28,29). Specific recognition of sialoglycoconjugates by adhesion molecules underlies their involvement in determining the metastatic potential of malignant cells (30,31).

We describe the identification and cDNA cloning of the gene encoding the STM sialyltransferase on the basis of its down-regulated expression in mammary tumor cells, as well as the biochemical characterization of the encoded enzyme. Cloning and characterization of human sialytransferases will enable studies on their normal cellular role(s), as well as on the mechanism of their aberrant regulation in malignant cells and its functional consequences.

Materials and Methods

Cell Lines and Media

Normal human mammary epithelial cell strains (81N, 76N, and 70N) derived from reduction mammoplasty specimens, as well as primary (21NT, 21PT), and metastatic (21MT-1, 21MT-2) tumor cell lines were established in long-term culture and characterized as described previously (32). Metastatic mammary tumor cell lines MCF-7, BT-474, BT-549, T-47D, ZR-75-1, MDA-MB-157, MDA-MB-231, MDA-MB-361, MDA-MB-435, and MDA-MB-436 were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA) and were routinely grown in DFCI-1 medium (33). Normal-immortalized prostate epithelial cell lines: CF3 (HPV-immortalized), CF91, and MLC (SV40-immortalized) were provided by Dr. Johng Rhim and were cultured in KGM medium (Difco, Beckton Dickinson, Sparks, MD, USA). Prostate metastatic tumor cell lines DU145, LNCaP, and PC3 were obtained from the ATCC. All cells were plated at a density of 1×10^6 cells per 100-mm diameter plate, cultured for 3–5 days, and harvested when the cultures were about 75% confluent for RNA isolation and near confluent for DNA isolation. Tissue culture medium components were purchased from Life Technologies, Inc., Gaithersburg, MD, USA or Hyclone, Logan, UT, USA.

Differential Display of mRNAs

Total cellular RNAs (50 ng) from exponentially growing cells were treated with DNase I in the presence of RNasin ribonuclease inhibitor, in order to remove residual DNA contamination (4). RNAs were extracted with phenol/chloroform, precipitated with ethanol, and redissolved in DEPC-treated water. Subsequently, RNAs were reverse-transcribed with a 3′-anchoring oligo dT primer T_{12}MA (where M is degenerate for G, C, or A). The resultant cDNAs were amplified by polymerase chain reaction (PCR) using T_{12}MA and an arbitrary 10-mer (OPA3:AGTCAGCCAC) as 5′-primer, and compared side by side on a sequencing gel as ^{35}S-labelled partial cDNA fragments corresponding to the 3′-end of the mRNAs (4).

Cloning, Sequencing of cDNAs, and Database Analysis

The partial *stm* cDNA obtained from differential display was reamplified by PCR, cloned into the PCRII vector using the TA cloning system (Invitrogen, Carlsbad, CA, USA), and sequenced on both strands with T7 and SP6 primers. Full-length cDNA clones were isolated from a 76N cDNA library constructed in λZap (Stratagene, La Jolla, CA, USA), which was screened

using the cloned PCR product as a probe. A full-length cDNA was sequenced on both strands using multiple internal primers (data not shown). Sequencing was performed with an ABI automated sequencer (model 373A). Oligonucleotides were synthesized by Amitof, Cambridge, MA, USA. The BLAST algorithm was used for nucleic acid sequence comparisons (34). Protein sequence comparisons were performed on GCG with final alignments on PILEUP and PRETTYPLOT.

Northern and Southern Blots

Total cellular RNAs were purified by standard guanidinium isothiocyanate and cesium chloride centrifugation as described elsewhere (4). Equal amounts of 20 μg RNA per lane were loaded and transferred onto nylon membrane. Following hybridization, RNA blots were washed in 2 $\times$ SSC containing 0.1% SDS at 65°C for 1 hr, dried, and exposed to Kodak films. After stripping at high stringency the blots were rehybridized with a ^{32}P-labeled 36B4 probe as a loading control. The 36B4 gene encodes a ribosomal protein, whose expression is not affected by growth conditions or estrogen receptor expression (35). Genomic DNA was isolated and hybridized by standard methods (4). Densitometric scans of autoradiographs were obtained with an imaging densitometer (Bio-Rad GS-700, BioRad Laboratories, Hercules, CA, USA) using the Molecular Analyst software.

Somatic Cell Hybrid and Southern Blot Analyses

Two panels of human–rodent somatic cell hybrids, the monochromosome hybrid mapping panel 2 and the regional mapping panel for chromosome 17 (obtained from the National Institute of General Medical Sciences [NIGMS]), were used to map the SIATL1 gene. The chromosome 17 content of the six hybrids in the regional mapping panel is shown in Fig. 6. DNA isolated from hybrid and parental cell lines were digested to completion with *Hin*dIII, separated electrophoretically on 0.8% agarose gels, and transferred onto nylon filters. Blots were hybridized with a ^{32}P-labeled 2.0-kb full-length *stm* cDNA probe.

Expression of Recombinant Protein and Sialyltransferase Assay

The cDNA encoding the catalytic domain of STM (Arg^{42}-Arg^{374}) was amplified by PCR, using the primers A, 5′-GGAGCCTCGAGGGACACCACATCATTTG-3′ (nt 157–181) and B, 5′-GCCGCAACTCGAGAAGAAGCAAAGCG-3′ (nt 1180–1205), with the *Xho*I restriction site being underlined. The amplified and *Xho*I-digested 1.05-kb fragment was inserted into the *Xho*I site of pcDSA vector (36). The single insertion in the correct orientation was analyzed by restriction digests and DNA sequencing, and the resulting expression construct was designated pcDSA-STM, which consisted of the IgM signal peptide sequence, a protein A IgG binding domain, and the catalytic domain of STM. COS-7 cells (5×10^6) were transiently transfected with 10 μg of pcDSA-STM using the DEAE-dextran procedure and cultured as described previously (37). Following a 48-hr transfection, the culture medium was collected and the A-STM fusion protein secreted into the medium was adsorbed on IgG-Sepharose (60 μL of resin/30 ml of medium) at 4°C for 16 hr. The resin was collected by centrifugation, washed three times with PBS, suspended in a final volume of 50 μL of Dulbecco's Modified Eagle medium without fetal bovine serum, and used as immobilized enzyme.

Enzymatic Assays

Enzyme activity assays were performed as described previously (36,37). Each reaction mixture contained 0.1 M of MES buffer (pH 6.4), 10 mM of $MgCl_2$, 2 mM of $CaCl_2$, 0.3% Triton CF-54, 0.1 M of CMP-[^{14}C]NeuAc (3.6 kBq), 0.15 mM of acceptor substrate, and 5 μL of enzyme preparation, in a total volume of 20 μL. After a 4-hr incubation at 37°C, the reaction was terminated by adding SDS-polyacrylamide gel electrophoresis loading buffer (10 μL) and boiling at 100°C for 15 min. Subsequently, the incubation mixture was subjected to SDS-PAGE for glycoprotein acceptors. For glycolipid acceptors, the incubation mixtures were applied on a C-18 column (Sep-Pak Vac, 100 mg; Waters, Milford, MA, USA) (37). The radioactive materials in glycoproteins or glycolipids were visualized with a BAS2000 radio image analyzer (Fuji Film, Tokyo, Japan), and the radioactivity incorporated into the acceptor was counted. The NDV sialidase was purchased from Oxford, Abingdon, Oxon, UK, NANase III from Glyco, Novato, CA, USA, and *N*-Glycanase from Genzyme, Tokyo, Japan. For linkage analysis of sialic acids, [^{14}C]NeuAc-incorporated fetuin and asialofetuin were synthesized with STM. After incubation with STM, [^{14}C]sialylated glycoproteins were collected by ethanol precipitation and washed 10 times with 70% ethanol to remove the substrate CMP-[^{14}C]NeuAc. To obtain the oligosaccharide portion of [^{14}C]sialylated fetuin and asialofetuin, the [^{14}C]sialylated glycoproteins were treated with 0.1 N NaOH/1 M of $NaBH_4$ at 37°C for 48 hr, and neutralized by the gradual addition of acetic acid in an ice-bath. Samples were desalted by gel filtration on Sephadex G-25 (1.3 $\times$ 25 cm). Then, [^{14}C]sialylated oligosaccharide alditols were purified by preparative TLC. The purified [^{14}C]sialylated oligosaccharide alditols were treated with various sialidases (NANase I, specific for α2,3-linked sialic acids; Newcastle Disease Virus sialidase, specific for α2,3- and α2,8-linked sialic acids; *Vibrio cholerae* sialidase, specific for α2,3-, α2,6-, and α2,8-linked sialic acids). Samples were subjected to HPTLC with silica gel sorbant and a solvent system of 1-propanol: aqueous ammonia:water = 6:1:2.5. The chromatogram was visualized with a BAS2000 radio image analyzer. The reference oligosaccharide alditol [Galβ1,3(NeuAcα2,6) GalNAc-ol (Oligo 1) and NeuAcα2,3 Galβ1,3 (NeuAcα2,6)GalNAc-ol (Oligo 2)] were detected with the resorcinol reagent.

Results

Identification of the STM Sialyltransferase by Differential Display

Differential display allows simultaneous comparisons of gene expression between multiple, closely related cell populations. In an attempt to isolate genes whose expression is reduced or lost during human mammary tumor progression, total RNA from normal mammary epithelial cell strains (76N, 70N), which senesce in culture, were displayed and compared with primary (21PT, 21NT), and metastatic ER+ (MCF-7) and ER− (21MT-1, 21MT-2, MDA-MB-435) malignant tumor cell lines (Fig. 1A), which all replicate indefinitely in culture. The 21T tumor progression series of cell lines (21PT→21NT→21MT-2→21MT-1) was established from one patient with infiltrating and intraductal breast carcinoma (32). A partial cDNA of 125 bp was identified as being down-regulated or absent in tumor cells (Fig. 1A) and was named *stm*. The differential expression of *stm* was confirmed by Northern hybridizations using the cloned and ^{32}P-labeled partial *stm* cDNA as a probe (Fig. 1B). The two differentially expressed cDNAs of ~125 bp shown in Figure 1A were cloned and sequenced independently and both corresponded to *stm*. The *stm* cDNA displayed no homology to any nucleotide sequence in the Genbank/EMBL databases.

Cloning and Sequencing of a Full-Length cDNA

The *stm* partial cDNA was used as a probe to screen 300,000 plaques of a cDNA library constructed from 76N normal mammary epithelial cells. Positive clones were selected and their differential expression was confirmed on Northern blots (Fig. 1B). A full-length cDNA (1930 bp) was sequenced on both strands, and contained a single open reading frame (40–1161 bp) with a consensus AUG start codon, a 5′-untranslated region (39 bp), and a 3′-untranslated region (765 bp) containing a polyadenylation signal AATAAA (nt 1890–1895) and a polyA tail. The nucleotide sequence as well as the deduced primary sequence are shown in Fig. 2. The encoded protein of 374 amino acid residues was named STM (locus name SIATL1), because it displayed homology to known sialyltransferases, and represents a new member of

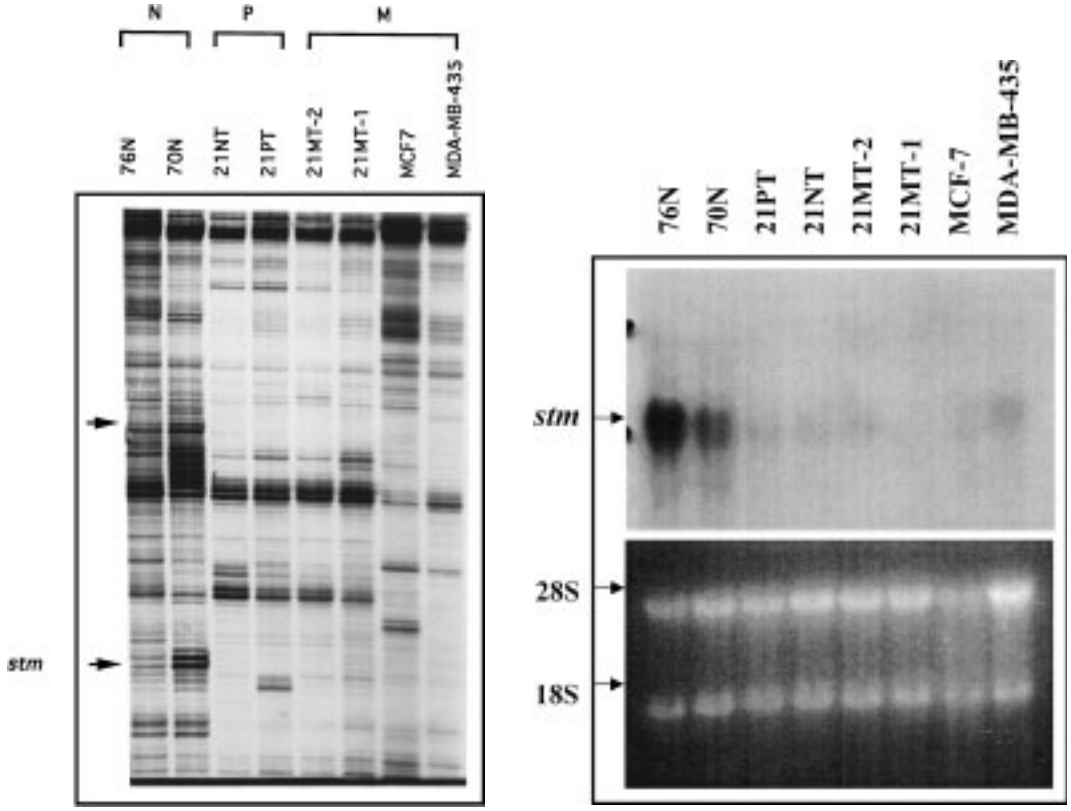

Fig. 1. Discovery of STM sialyltransferase by multiway differential display. (A) Total RNAs were isolated from normal human mammary epithelial cell strains (N, 76N, 70N; lanes 1 and 2), primary (P, 21NT, 21PT; lanes 3 and 4) and metastatic mammary tumor cell lines (M, 21MT-1, 21MT-2, MCF-7, and MDA-MB-435; lanes 5, 6, 7, and 8). The RNAs were reverse-transcribed with the T_{12}MA 3′-end anchored primer and amplified with T_{12}MA and OPA3 (AGTCAGCCAC), a 5′-end random primer. The position of the *stm* partial cDNA (125 bp), and another example of a differentially displayed *c*DNA, are indicated with arrows. **(B)** Northern blot analysis of differential *stm* gene expression. A cloned and ^{32}P-labeled partial *stm c*DNA obtained from differential display was used as a probe (upper). Ethidium bromide-staining of the gel used for the Northern blot (lower). Each lane contains 20 μg of total RNA. The 28S and 18S subunits of ribosomal RNA are shown.

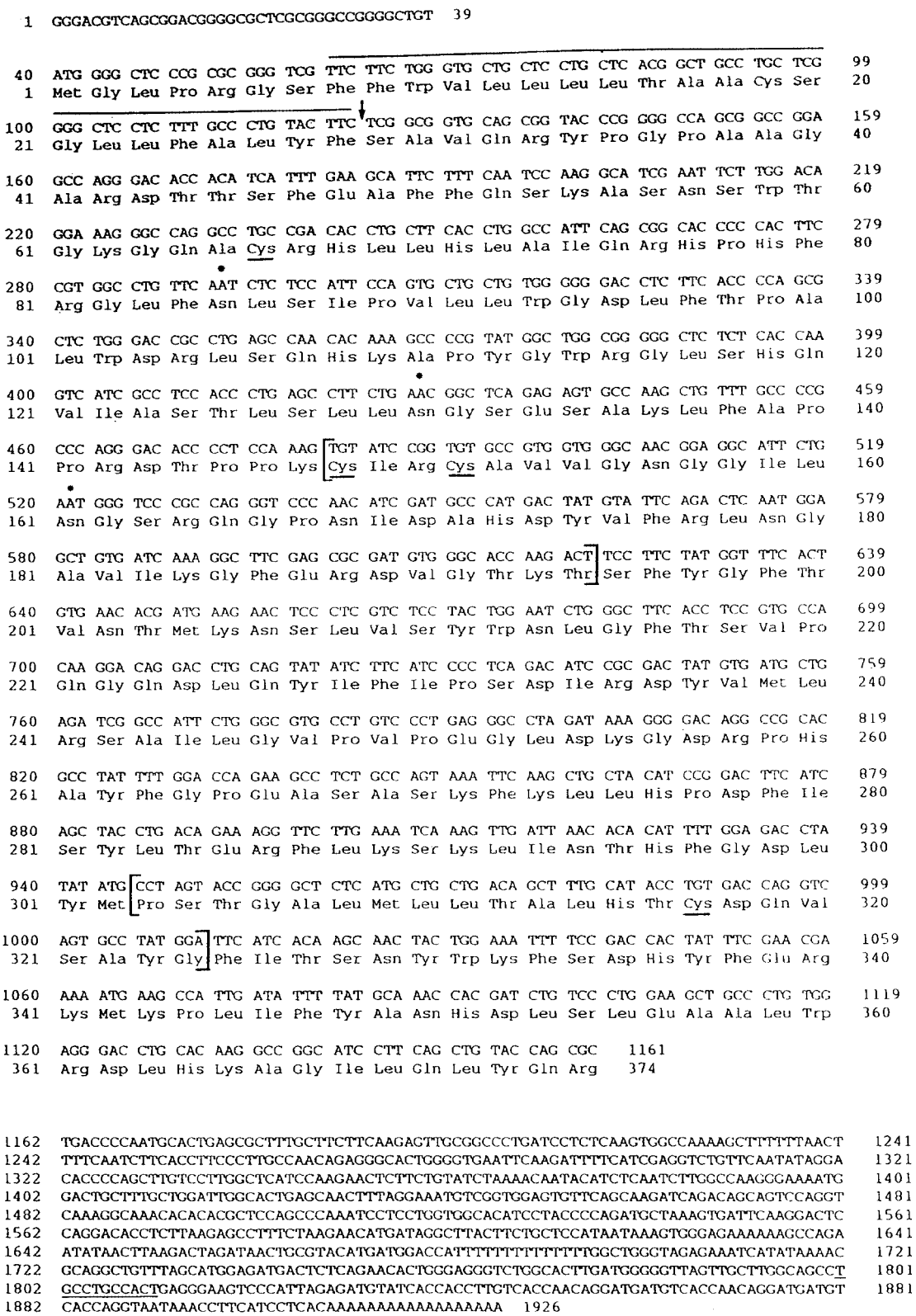

Fig. 2. Sequence and structural features of STM. Primary amino acid sequence of the STM sialyltransferase inferred from the cDNA sequence. The *L*-sialylmotif (aa 148–194) and the *S*-sialylmotif (aa 303–324) are defined in brackets. Overlined sequence indicates a potential membrane anchor domain. Conserved cysteines are underlined. Potential *N*-linked glycosylation sites are marked with an asterisk. Upper numbers on the right and left represent nucleotides and lower numbers represent amino acid residues. The OPA3 priming position (nt 1801–1810) is underlined. A polyadenylation signal is present (nt 1890–1896).

the sialyltransferase gene family. The initiator ATG indicated in Fig. 2 lies within a relatively strong Kozak consensus (38) and is probably the translation start site, because the resulting protein sequence aligns optimally with other sialyltransferases (data not shown). The downstream sequence contains no AUG codons in proximity to the initiation site, and no AUG was found in the additional 288-bp upstream of the predicted start codon when a longer cDNA was sequenced (data not shown). In Figure 2, the sequence of the 5′-arbitrary OPA3 primer was underlined (nt 1801–1811) and contained two mismatches. The original partial cDNA obtained from differential display corresponds to bases 1801–1926 (Fig. 2), where sialyltransferases do not display any significant homology and, therefore, it gave no clues to the identity of the differentially displayed gene.

Structure of STM and Sequence Comparison to Other Sialyltransferases

A Hoop and Woods hydrophilicity plot revealed a hydrophobic region Phe^{8}-Phe^{28} at the N-terminus, which could serve as a noncleavable anchor domain, a common structural feature of sialyltransferases (6). A potential membrane-spanning region suggests that STM has a type II membrane orientation and a predicted molecular weight of 42 kDa. A potential site for casein kinase II (CK-2) phosphorylation is present in STM (Thr^{45}-Ser-Phe-Glu^{48}), and three consensus motifs for *N*-glycosylation at positions Asn^{85}, Asn^{130}, and Asn^{161} (see Fig. 2). Glycosylation as well as phosphorylation/dephosphorylation are plausible mechanisms that regulate the activity of glycosyltransferases. A primary sequence alignment of STM with related sialyltransferases revealed that STM contains two catalytic domains, the *L*-sialylmotif (aa 148–194) and *S*-sialylmotif (aa 303–324), of 47 and 23 amino acid residues, respectively. All sialyltransferases share these functional motifs, which are unique to the sialyltransferase gene family, are indispensable for sialyltransferase activity, and likely contain binding sites for donor and acceptor sugars. In both sialylmotifs, about 60% of the amino acids are conserved, which are mostly charged or polar. A third homology domain shared by sialyltransferases is the *VS*-sialylmotif, a short stretch of four highly conserved residues (His^{336}-Tyr-Phe-Glu^{339}) and Leu^{363} located near the C-terminus (39). Also present in STM are the cysteine residues (Cys^{151} and Cys^{317}) involved in the formation of a disulfide bond, which is essential for the active conformation of sialyltransferases (40). Recently, it was shown by site-directed mutagenesis that the invariant residues Cys^{181} or Cys^{332} in the *L*- and *S*-sialylmotifs participate in the formation of an additional intradisulfide linkage that is essential for proper conformation and activity of ST6Gal I (13). STM displayed the highest sequence similarity to mouse ST6GalNAc II (41) with 229 identical and 274 conserved amino acids, whereas 229 amino acids were identical and 242 amino acids were conserved between STM and chicken ST6GalNAc II (42). As shown in Fig. 3, sequence similarity between STM, mouse and chicken ST6GalNAc II is not restricted to sialylmotifs. It should be noted that, although the overall amino acid identity between mouse and human ST6GalNAc II is only 61%, it is 82.3% in the region from the N-terminus of the *L*-sialylmotif to the C-terminus of the enzyme, which is considered to contain the catalytic site and substrate binding sites. The corresponding region displays a 75.2% identity between chicken and human ST6GalNAc II. Therefore, these three enzymes will probably display similar acceptor specificities.

Down-Regulated Expression of stm in Human Cancer Cell Lines

The differential expression of *stm* was confirmed by Northern blot analysis employing a series of normal and tumor mammary epithelial cell lines. Two mRNA species with approximate sizes of 2.5 and 1.8 kb were detected in 81N, 76N, and 70N normal mammary epithelial cell strains. Northern blots were hybridized to a full-length *stm* cDNA probe. The same expression pattern was obtained when blots were hybridized to probes corresponding to different fragments of the cDNA (data not shown), indicating that the two transcripts more likely represent alternatively processed messages of the same gene, as reported for the liver α2,6-sialyltransferase (43). Expression of *stm* was markedly down-regulated in the T-47D, BT-474, and MDA-MB-435 tumor cell lines, while traces or no *stm* message was detected in all other tumor cell lines (Fig. 4A). However, the MDA-MB-361 metastatic tumor cell line expressed *stm* mRNA levels which were as high as in 76N and higher than in 81N and 70N normal cell strains (Fig. 4A). MDA-MB-361 differs from other metastatic breast tumor cell lines in karyology in that it was isolated from a brain metastasis and not from pleural effusions, and has a lower malignant potential (44). In addition, down-regulation of *stm* expression was observed in prostate malignant tumor cells. As shown in Fig. 4B, *stm* was expressed at similar levels in all three normal-immortalized prostate epithelial cell lines (CF3, CF91, MLC) but down-regulated in PC3 and completely absent in the DU145 and LNCaP prostate metastatic tumor cell lines. These results indicate that *stm* could be down-regulated in human tumors, although this should be confirmed by studies employing matched normal and tumor tissue specimens.

Southern Analysis

Restriction digests of total genomic DNAs isolated from normal and tumor mammary cell lines were hybridized to a full-length *stm* cDNA probe. The *Eco*RI digest revealed two major hybridizing fragments of 7.8 and 4.4 kb, respectively, and a minor fragment of 6.2 kb that, upon prolonged exposure, was detected in all lanes (Fig. 5). However, the 4.4-kb band was absent in MDA-MB-231 and

```
cST6GalNAc II   1:  MGSPRWKRFCFLLLAAFTSSLLLYGHYYATV-DVRSGPRVVTSLLQPELLFLVRPDTPHP  59
hST6GalNAc II   1:  MGLPRGSFFWVLLLLTAACSGL-LFALYFSAVQR-------Y------PGPAAGA-------  41
mST6GalNAc II   1:  MDLPRRWLFRMLLLVATSSG-ILLM-LYSSAGQQ-------S------PETQVPA-------  40

cST6GalNAc II  60:  DNSHHKELRGTVKSREFFSQPSSELEKPKPSGKQPTPCPRSVAATAKADPTFGELFQFDI  119
hST6GalNAc II  42:  ----RDTTSFE-----AFFQSKASNSWTG----KGQA-CRHLLHLAIQRHPHFRGLFNLSI  88
mST6GalNAc II  41:  ----RNMA-YPR---AFFDPKPPNSE-N---RKSRL-CQHSLSLAIQKDRRFRSLFDLST  87

cST6GalNAc II 120:  PVLMWDQHFNPETWDRLKARRVPYGWQGLSQAAVGSTLRLLNTSSNTRLFDRHLF-PGGC  178
hST6GalNAc II  89:  PVLLWGDLFTPALWDRLSQHKAPYGWRGLSHQVIASTLSLLNGSESAKLFAPPRDTPPKC  148
mST6GalNAc II  88:  PVLLWEGLFTQELWNNLSQHKVPYGWQGLSHEVIASTLRLLKSPESGELFGAPRKLPLSC  147

cST6GalNAc II 179:  IRCAVVGNGGILNGSRQGRAIDAHDLVFRLNGAITKGFEEDVGSKVSFYGFTVNTMKNSL  238
hST6GalNAc II 149:  IRCAVVGNGGILNGSRQGPNIDAHDYVFRLNGAVIKGFERDVGTKTSFYGFTVNTMKNSL  208
mST6GalNAc II 148:  IRCAVVGNGGILNGSRQGQKIDAHDYVFRLNGAITEAFERDVGTKTSFYGFTVNTMKNSL  207

cST6GalNAc II 239:  IAYEAYGFTRTPQGKDLKYIFIPSDARDYIMLRSAIQGSPVPEGLDKGDEPQKYFGLEAS  298
hST6GalNAc II 209:  VSYWNLGFTSVPQGQDLQYIFIPSDIRDYVMLRSAILGVPVPEGLDKGDRPHAYFGPEAS  268
mST6GalNAc II 208:  ISYAKLGFTSVPQGQNLRYIFIPSSIRDYLMLRSAILGVPVPEGPDKGDRPHTYFGPETS  267

cST6GalNAc II 299:  AEKFKLLHPDFLHYLTTRFLRSELLDMQYGHLYMPSTGALMLLTALHTCDQVSAYGFITA  358
hST6GalNAc II 269:  ASKFKLLHPDFISYLTERFLKSKLINTHFGDLYMPSTGALMLLTALHTCDQVSAYGFITS  328
mST6GalNAc II 268:  ASKFKLLHPDFISYLTERFLKSKLINTRFGDMYMPSTGALMLLTALHTCDQVSAYGFITN  327

cST6GalNAc II 359:  NYEQFSDHYYEPEKKPLVFYANHDMLLEAELWRSLHRAGIMELYQR               404
hST6GalNAc II 329:  NYWKFSDHYFERKMKPLIFYANHDLSLEAALWRDLHKAGILQLYQR               374
mST6GalNAc II 328:  NYQKYSDHYFEREKKPLIFYANHDLSLEASLWRDLHNAGILWLYQR               373
```

Fig. 3. Primary sequence alignment of STM with mouse and chicken GalNAc II α2,6-sialyltransferases. Sequences appear in Genbank/EMBL with the accession numbers STM, hST6GalNAc II (U14550); mouse, mST6GalNAc II (X93999); and chicken, cST6GalNAc II (X77775). Amino acid sequences are boxed according to the 3/3 score. Amino acids are identified by the single-letter code.

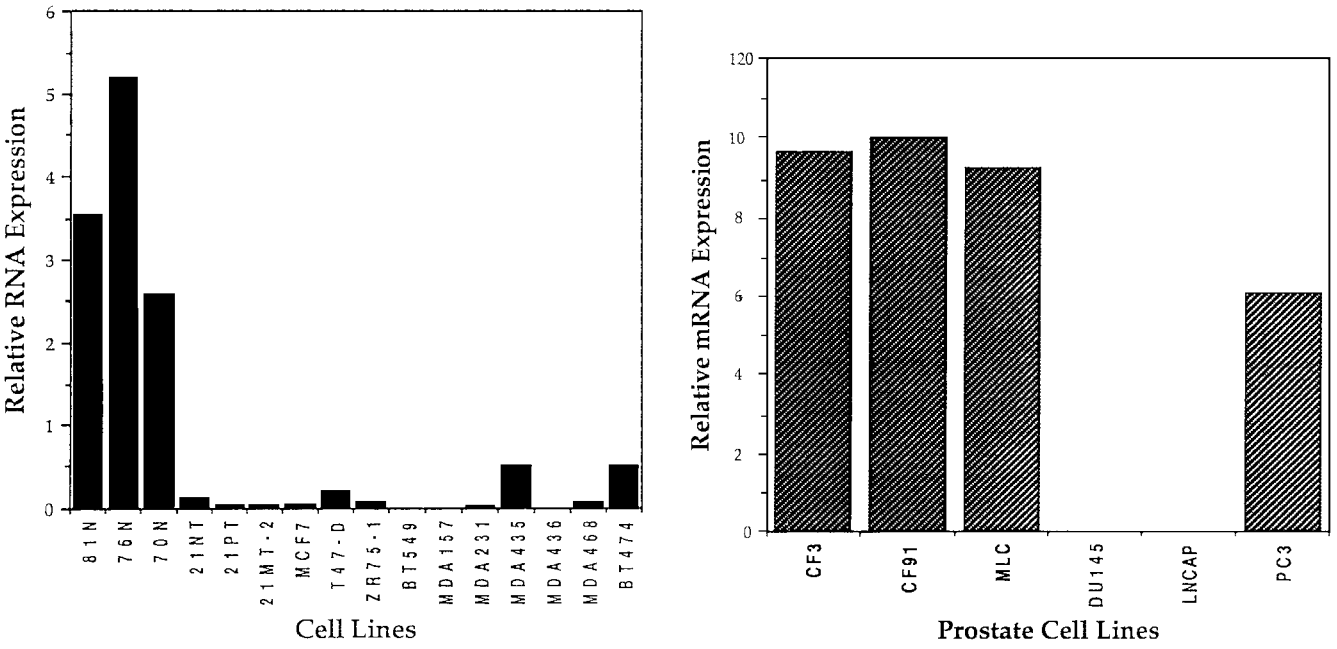

Fig. 4. Differential expression of STM sialyltransferase in human normal and tumor cell lines. Densitometric scans of Northern blot autoradiographs of left, human mammary epithelial normal cell strains, primary, and metastatic tumor cell lines. Right, human prostate epithelial normal-immortalized and metastatic tumor cell lines. Northern blots contained 20 μg of total RNA in each lane and were hybridized to a ^{32}P-labeled full-length *stm* *c*DNA probe. For an internal standard, blots were hybridized to a ^{32}P-labeled 36B4 *c*DNA probe (35). Ratios of STM/36B4 RNA levels are presented. Details on cell lines are given in Materials and Methods.

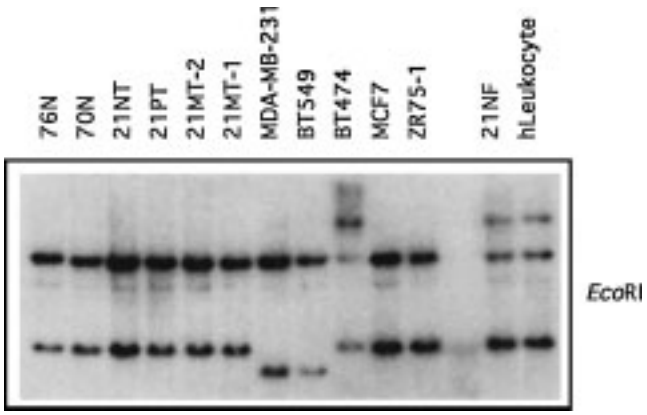

Fig. 5. Southern blot analysis of the gene encoding STM sialyltransferase in normal and tumor human mammary epithelial cell lines. A full-length *stm* cDNA was ^{32}P-labeled and used as a probe. Each lane contained 20 μg of total *Eco*RI-restricted genomic DNA. The sizes of hybridizing fragments were 7.8, 6.2, and 4.4 kb. The size of the higher band present only in BT-474, 21NF, and leukocyte was 9.9 kb, and that of the lower band present only in MDA-MB-231 and BT-549 was 4.1 kb.

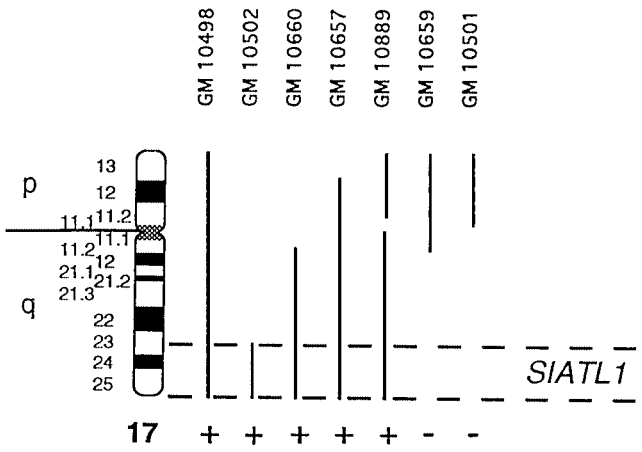

Fig. 6. Chromosomal localization of the SIATL1 gene. Ideogram showing the chromosome 17 content of the regional mapping panel used for sublocalization of the SIATL1 gene. This panel allows assignment of genes to at least six different intervals on chromosome 17. Southern blot analysis showed that the presence of SIATL1-specific sequences correlated with the presence of the region 17q23-qter. The presence (+) or absence (−) of SIATL1 sequences in the hybrid clones are indicated.

BT-549 tumor cell lines, which instead contained a lower band of 4.1-kb. In the BT-474 tumor cell line, as well as in the 21NF, normal fibroblasts and leukocytes, a band of approximately 9.9 kb was detected. In *Hin*dIII digests of the same DNAs, uniform bands of 6.0, 5.5, and 4.3 kb were detected (data not shown). The uniform restriction patterns obtained on Southern blots suggest that the SIATL1 gene does not appear to be consistently rearranged in breast tumor cell lines, although genomic alterations might be present in a subset of breast tumors. Down-regulation of *stm* expression could be regulated at the transcriptional level in some tumors or by the presence of genomic mutations in others. Alternatively, the differences detected in the patterns of hybridizing restriction fragments could arise from polymorphisms.

Mapping of the SIATL1 Gene to Human Chromosome 17q23-qter

The chromosomal localization of SIATL1 was initially determined by analysis of its segregation in a monochromosome hybrid mapping panel. The human SIATL1-specific *Hin*dIII fragments, which could clearly be resolved from the hamster and mouse fragments, segregated with human chromosome 17 (data not shown). There were no discordancies for SIATL1 localization to this chromosome. By analysis of a regional mapping panel for chromosome 17, SIATL1 was subsequently regionally mapped to an interval between 17q23 and 17qter, defined by the breakpoint in hybrid GM10502 (Fig. 6).

Tissue Distribution of STM

A Northern blot containing polyA+ RNA from different normal human tissues was hybridized to a *stm* cDNA probe corresponding to the 3′-untranslated sequence. This probe is expected to specifically detect only *stm*-specific transcripts because sialyltransferases display no significant homologies in the 3′-untranslated regions of their cDNAs. The signal intensities are roughly proportional to the abundancy of the transcripts. The expression of *stm* displayed a wide tissue distribution. As shown in Fig. 7, abundant expression of *stm* was detected in heart, lung, skeletal muscle, kidney, and pancreas, whereas lower *stm* levels were detected in placenta. The liver contained trace amounts of *stm* mRNA, which was not detected in brain tissue. The same transcripts were detected when this tissue blot was hybridized to a full-length probe.

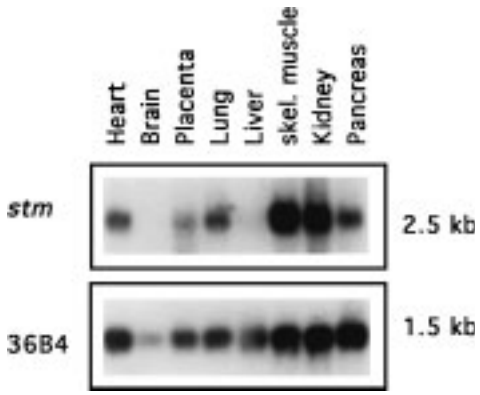

Fig. 7. Expression of STM sialyltransferase in normal human tissues. The Northern blot was probed with the 0.7-kb *Eco*RI + *Xho*I restriction fragment that corresponds to the 3′-untranslated end of the *stm* cDNA (upper) and 36B4 as an internal standard (lower). Each lane contained approximately 2 μg of pure polyA$^+$ RNA. The RNAs were run on a denaturing formaldehyde/1.2% agarose gel and blotted onto a nylon membrane (Human MTN Blot, Clontech, #7760-1).

Table 1. Acceptor substrate specificity of recombinant STM sialyltransferase

Acceptor	Representative Structures of Carbohydrates	STM	mST6GalNAc II	mST6GalNAc I
			(%)	
Fetuin	NeuAcα2-3Galβ1-3GalNAc-Ser/Thr NeuAcα2-3Galβ1-3(NeuAcα2-6)GalNAc-Ser/Thr NeuAcα2-6(3)Galβ1-4GlcNAc-R	83.3	76.0	84.0
Asialofetuin		100.0	100.0	100.0
Asialo-agalacto-fetuin		9.12	12.0	91.0
BSM	NeuAcα2-3Galβ1-3GalNAc-Ser/Thr NeuAcα2-6GalNAc-Ser/Thr	4.73	7.32	19.0
Asialo-BSM		9.12	15.0	169.0
α1 acid glycoprotein	NeuAcα2-6(3)Galβ1-4GlcNAc-R	0	0	14.0
Asialo-α1 acid glycoprotein		0	0	8.0
Ovomucoid		3.38	9.41	n.t.
Galβ1-3GalNAc-benzyl		0	0	0
NeuAcα2-3Galβ1-3GalNAc-benzyl		0	2.50	n.t.
Asialo-GM1	Galβ1-3GalNAcβ1-4Galβ1-4Glc1-1Cer	0	0	0
GM1b	NeuAcα2-3Galβ1-3GalNAcβ1-4Galβ1-4Glc1-1Cer	0	0	n.t.
Paragloboside	Galβ1-4GlcNAcβ1-3Galβ1-4Glc1-1Cer	0	0	n.t.

Comparison of the substrate specificity of STM (hST6GalNAc II), mST6GalNAc I (46) and mST6GAlNAc II (41) sialyltransferases. The relative activity of incorporation of sialic acids into asialofetuin as a substrate is shown. Each substrate was used at the concentration of 0.15 mM. A value of 0 indicates less than 0.1%. R represents the reminder of the *N*-linked oligosaccharide chain.

GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; NeuAc, *N*-acetylneuraminic acid.

Acceptor Substrate Specificity

The cDNA sequence that encodes the putative enzymatically active domain of STM (Arg^{42}-Arg^{374}) was fused to the IgG binding domain of protein A and transiently expressed in COS-7 cells. The fusion protein efficiently mediated the transfer of sialic acids to glycosidically *O*-linked oligosaccharides of fetuin and asialofetuin because sialic acids were incorporated into *N*-glycanase resistant materials. The relative activity of incorporation of sialic acids into the asialofetuin substrate is shown in Table 1. Native fetuin contains three glycosidically *O*-linked oligosaccharides, two of which are NeuAcα2,3Galβ1,3GalNAc and NeuAcα2,3Galβ1,3 (NeuAcα2,6)GalNAc. Therefore, in native fetuin, the GalNAc residues in two of the three *O*-linked oligosaccharides can serve as acceptors. As shown in Table 1, the incorporation of NeuAc residues for the asialofetuin was increased about 20% from that of native fetuin. The incorporation of NeuAc residues for the agalactoasialofetuin (galactosidase-treated asialofetuin) was dramatically decreased. The activity toward asialo-BSM, in which only 5% of the total carbohydrate chains contain the Galβ1,3GalNAc sequence, was almost negligible. No significant activity was observed toward α1 acid glycoprotein having only glycosidically *N*-linked oligosaccharides. In addition, glycosphingolipids could not serve as acceptors for this enzyme. The acceptor specificity of STM was similar to that observed for mouse and chicken ST6GalNAc II (41,42), but significantly different from that of chicken and mouse ST6GalNAc I (45,46). The incorporated NeuAc residues are resistant to treatment with NDV sialidase, but sensitive to treatment with NANase III (Table 2).

Table 2. Linkage analysis of incorporated sialic acids

Treated by	Relative Count (%)
NANase III	0.0
NDV sialidase	85.0
N-glycanase	96.3
Nontreated fetuin	100.0

[^{14}C] sialylated fetuin was prepared as described in Materials and Methods. Subsequent to sialidase and *N*-glycanase treatment, each sample was subjected to SDS-PAGE. The radioactive materials in fetuin were visualized with a BAS2000 radio image analyzer and the incorporated radioactivity was counted.

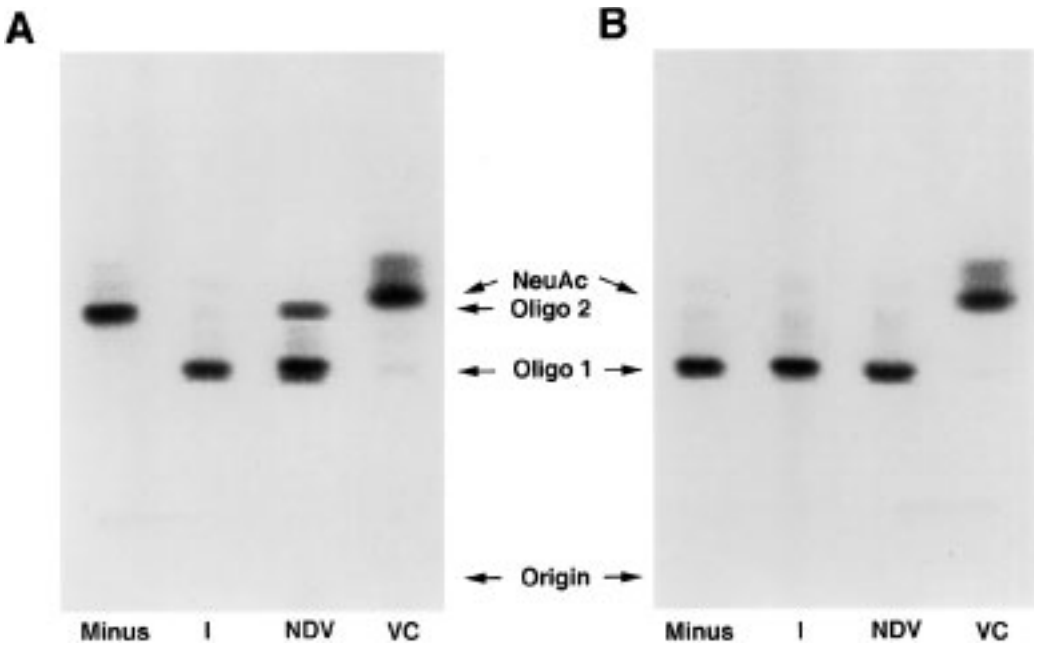

Fig. 8. Linkage analysis of incorporated sialic acids. Oligosaccharide alditols derived on 0.1 N NaOH/1 M NaBH4 treatment of [14C]sialylated **(A)** fetuin, and **(B)** asialofetuin, were treated in the absence of sialidase (minus) or the presence of NANase I (I), Newcastle Disease Virus sialidase (NDV), and Vibrio cholerae sialidase (VC). Oligo 2, NeuAca2,3 Galb1,3(Neu Aca2,6)GalNAc-ol); Oligo 1, Galb1,3(NeuAca2,6)GalNAc-ol.

The former sialidase cleaves α2,3- and α2,8-linkage but not α2,6-sialyl-linkage, and the latter cleaves α2,3-, α2,6- and α2,8-sialyl linkage. These results show that the cloned gene encodes an enzymatically active GalNAc α2,6-sialyltransferase. To determine the linkage specificity of STM, [^{14}C]sialylated oligosaccharide alditols were prepared by β-elimination of [^{14}C]sialylated fetuin and asialofetuin (Fig. 8). Subsequently, a desalted sample was subjected to HPTLC. All of the radioactive product migrated as a low molecular compound, that is, no radioactivity remained at the origin, suggesting that [^{14}C]sialylation occurred exclusively on *O*-linked glycan chains of fetuin (not shown). In Fig. 8A, NDV lane shows two bands resulting from partial digestion with NDV sialidase. The upper and lower bands correspond to Oligo 2 (NeuAcα2,3Galβ1,3(NeuAcα2,6)GalNAc-ol) and Oligo 1 (Galβ1,3(NeuAcα2,6)GalNAc-ol), respectively. Linkage analysis using linkage-specific sialidases revealed that the [^{14}C]sialylated oligosaccharide alditols contained α2,6-linked sialic acids.

Discussion

We sought to identify genes with aberrant expression in human breast tumors by applying differential display comparisons of multiple normal mammary cell strains and tumor cell lines with increasingly malignant phenotypes. A gene encoding a new sialyltransferase (named STM) was identified as being down-regulated in breast tumor cells and the corresponding full-length cDNA was cloned. STM has a putative type II transmembrane topology and contains the *L*- and *S*-sialylmotifs, two functional domains present in all enzymatically active sialyltransferases. The catalytic domain of STM was expressed in COS-7 cells as a fusion protein with the IgG binding domain of protein A, and its enzymatic activity was characterized. STM could efficiently mediate the transfer of sialic acids to *O*-linked oligosaccharides in native fetuin and asialofetuin but not in agalactoasialofetuin, whereas *N*-linked oligosaccharides and glycosphingolipids could not serve as acceptors. Therefore, STM is an active GalNAc α2,6-sialyltransferase with Galβ1,3GalNAc-*O*-Ser/Thr and (+/−Neu5Ac α2,3) Galβ1,3GalNAc-*O*-Ser/Thr acceptor specificity and represents the human ST6GalNAc II. Nonetheless, STM displayed the highest primary sequence similarity to mouse (41) and chicken (42) ST6GalNAc II, which also form the linkage Neu5Acα2,6GalNAc on the Galβ1,3GalNAc acceptor. Given that sequence homologies of STM to mouse and chicken ST6GalNAc II are 75% and 77%, respectively, STM is likely encoded by a unique sialyltransferase gene. Unique sialyltransferase genes from one species typically exhibit <50% primary sequence homology, and the same gene exhibits >95% homology in different species (7). Overall primary sequence homologies of STM to known sialyltransferases range from 15–23% for amino acid identities and 20–30% for conservative replacements.

Analysis of two independent mapping panels allowed us to assign the SIATL1 gene encoding STM

to the long arm of human chromosome 17 at q23-qter. This localization differs from those described for mapped sialyltransferases (e.g., the Galβ1,4GlcNAc α2,6-sialyltransferase gene was mapped to 3q27–28 [47], Galβ1,3 GalNAc/Galβ1,4 GlcNAc α2,3-sialyltransferase to 11q23–24 [47], Galβ1,3(4) GlcNAc α2,3-sialyltransferase to 1p33–34 (47), and NeuAcα2,3Galβ1,4Glcβ1-1/B4Cer α2,8-sialyltransferase to 12p11.2–12.1 [48]). The region of 17q to which SIATL1 maps is syntenic with mouse chromosome 11. At least 28 genes on human 17q11.2-qter are known to map on mouse chromosome 11 (49). Localization of SIATL1 to this region suggests that SIATL1 might be part of this linkage group and thus resides on mouse chromosome 11. Chromosome 17q is frequently affected in human breast cancers. Three breast cancer-associated genes have been mapped to 17q: *BRCA1* at 17q21, *NME1* (*NM23*) at 17q21–22, and *ERBB2* at 17q11.2–12 (50,51). In addition, three separate regions on 17q are nonrandomly deleted in breast cancer: a region surrounding *BRCA1* at 17q21, a central region flanked by markers D17S86 and D17S21, and a distal region including markers D17S4 and D17S24 at 17q23–q25 (52,53). Interestingly, the latter region coincides with the region to which SIATL1 was mapped. Whether or not SIATL1 is deleted in a subgroup of breast cancers remains to be established.

The SIATL1 gene specifies two transcripts of approximately 2.5 and 1.8 kb, probably generated through alternative splicing or alternative promoter utilization, as shown for other sialyltransferases (54,55). Northern analyses of normal and tumor cell lines showed that both transcripts are significantly down-regulated and mostly absent in mammary and prostate tumor cell lines (see Fig. 4A, 4B). Of all breast tumor cell lines tested, only MDA-MB-361 expressed *stm* at levels similar (and even higher) than normal mammary epithelial cells. The MDA-MB-361 cell line was isolated from a brain metastasis; it has a lower malignant potential than cell lines isolated from pleural effusions (e.g., MDA-MB-435, MDA-MB-231, and MDA-MB-468 [44]), which express much less or no *stm* message. Down-regulation of *stm* expression in tumor cells could arise from genomic alterations. Because Southern analysis did not reveal consistent gross genomic rearrangements in mammary tumor cells, we speculate that down-regulation of *stm* in tumor cells occurs at the level of transcription, although deletions or point mutations cannot be excluded in a subset of breast tumors. Regulation of sialyltransferases at the level of transcription was described previously (56,57). It should be mentioned that the expression of *stm* could not be induced by cycloheximide in MDA-MB-231 breast tumor cells, suggesting that negative regulation of *stm* transcription in tumor cells does not involve protein factors with short half-life times. In 76N normal breast cells, *stm* mRNA was enhanced 1.5- to 2.0-fold, probably due to the stabilizing effect of cycloheximide. Furthermore, down-regulation of *stm* in MDA-MB-231 tumor cells does not result from altered RNA stability (data not shown).

The putative functional implication(s) of STM inactivation in tumor cells have not been established. Normally, surface sialic acids play important roles as receptors for molecules that regulate cellular growth, differentiation, cell–cell communication, and adhesion. Nonetheless, aberrant glycosylation of glycoproteins and glycolipids in human cancers has been associated with an invasive phenotype (23,28,29). Neoplastic cells often express a larger amount of multiantennary and hypersialylated *N*-linked glycans than their normal counterparts, due to incomplete synthesis of carbohydrate chains and accumulation of their precursors (58). Hypersialylation results in increased surface charge that affects the interactions of tumor cells with host cells, and increases their metastatic potential (23), as well as tumor cell immunogenicity (59). Thus, aberrant sialic acid structures are recognized as tumor-associated surface antigens and result, at least in part, from the aberrant regulation of glycosyltransferases (25–27). For example, elevated α2,3-sialyltransferase activity is involved in the accumulation of the α2,3-sialylated Lec antigen (NeuAcα2,3 Galβ1,3GlcNAc), precursor for the CA19-9 antigen that is elevated in pancreatic and colorectal malignant tumors (25). In breast cancer patients, the expression of ST3Gal III is often highly increased and positively correlates with the expression of ST6Gal I and ST3Gal IV (60), leading to the formation of the Sialyl-Lewisa and Sialyl-Lewisx determinants. In human breast tumors, high expression of ST3Gal III and ST6Gal I is associated with a poor prognosis. In addition, the Tn antigen (GalNAc α1-*O*-Ser/Thr) and Sialyl-Tn antigen (NeuAcα2,6GalNAc α1-*O*-Ser/Thr) are mucin-type (*O*-linked) carbohydrate epitopes expressed in many human tumors. The Sialyl-Tn antigen is synthesized by ST6GalNAc I (61) and is related to poor prognosis for colorectal and breast cancer patients (26). A long cDNA of 2.46 kb encoding an active ST6GalNAc I enzyme, as well as a shorter splice variant of 2.23 kb encoding an inactive enzyme, have been identified (61). Interestingly, we have cloned different splice variants of STM lacking the *S*-sialylmotif due to premature stop codons (data not shown). The long form of the human ST6GalNAc I gene encodes the candidate synthase for the sialyl-Tn antigen (61). It should be noted that mouse ST6GalNAc II also exhibits activity toward GalNAc-*O*-Ser/Thr (62), but chick and human ST6GalNAc II have a very weak activity and no activity, respectively, toward GalNAc-*O*-Ser/Thr (61,63).

Two linkage patterns are common in *O*-linked oligosaccharides: the α2,3 (Neu5Acα2,3 Gal-R and NeuAcα2,3GlcNAc-R) and the α2,6 (NeuAcα2,6GalNAc-*O*-Ser/Thr). Of all known sialyltransferases, the synthesis of *O*-linked sialylated structures is

catalyzed by GalNAc α2,6-sialyltransferases, namely ST6GalNAc I (43,46), ST6GalNAc II (41,42), ST6GalNAc III (15,64), ST6GalNAc IV (64), and ST6GalNAc V (64,65), as well as by Galβ1,3GalNAc α2,3-sialyltransferases, namely ST3Gal I (11,66) and ST3Gal II (67,68). In breast tumor cells, aberrant glycosyltransferase activities engaged in assembling, elongating, and terminating *O*-glycan core 1 (Galβ1,3GalNAcα-R) and core 2 (Galβ1,3 (GlcNAcβ1,6)GalNAcα-R) result in altered glycosylation of the polymorphic epithelial mucin (product of the MUC1 gene) and the selective exposure of a cryptic epitope in the tandem repeat (69,70). In breast and ovarian carcinomas, MUC1 is overexpressed and contains a large amount of *O*-linked glycans, which in normal mammary epithelial cells consist of core 2–based structures. The mucin expressed by breast carcinomas has shorter side chains, often consisting of sialylated core 1 structures (Galβ1,3GalNAc) synthesized by ST3Gal I, which adds sialic acid to core 1 and terminates chain extension (71). ST3Gal I expression is elevated in primary breast carcinomas when compared to normal or benign tissue (70). Although Galβ1,3GalNAc α2,3-sialyltransferase activity is increased by several fold in breast tumor cells (69), the Galβ1,3 GalNAc(GlcNAc to GalNAc) β1,6-GlcNAc-transferase activity, which is responsible for core 2 synthesis, is absent (69,70). Loss of core 2 branching in tumor cells leads to shorter, sialylated *O*-glycans, adjacent to peptide epitopes. Thus, aberrant glycosylation of MUC1 in breast cancer results, at least in part, from increased activity of the ST3Gal I sialyltransferase (70). Increased α2,3 sialylation of core 1 inhibits core 2 branching. Recently, it was shown that it is primarily the elevation of ST3Gal I in breast tumors that leads to the dominant expression of core 1 *O*-glycans on MUC1, because ST3Gal I can compete effectively with C2GnT1 β6Glc NAc-transferase *in vivo*, and inhibit the formation of core 2 structures on MUC1 (72).

In normal cells, sialyltransferases that use the same substrate but add different sugars can compete for the common substrate providing that they overlap in the Golgi. Down-regulation or inactivation of certain sialyltransferases, like STM, in breast tumor cells could enhance the activity of other sialyltransferases that catalyze linkages associated with hypersialylated cell surfaces. Understanding the sialylation mechanisms will be facilitated by the identification of all sialyltransferases present in human cells, and by the elucidation of the mechanism of their aberrant regulation in tumor cells. Cloning of STM and determination of its differential expression and substrate specificity should contribute to the elucidation of pathways involved in biosynthesis of *O*-glycan chains of mucins in normal versus tumor cells, and should enable a complete analysis, as well as clinical applications of mammary cell mucins and mucin-related antigenicity.

Acknowledgments

The nucleotide sequence reported in this manuscript appears in Genbank/EMBL with the accession number U14550. The HUGO/GDB designation for the gene symbol is SIATL1.

This work was supported by grants from the National Institutes of Health (CA61253) (R.S.), the Greek Ministry of Education (G.S.), Grants-in-Aid for Scientific Research on Priority Areas (# 09254268, 09260231, 10152263) and Grants-in-Aid for Scientific Research (# 09680639) (S.T.). The gene mapping work was supported by a grant from the Swedish Cancer Society (G.S.). A Senior Research Award from the Fulbright Program/Council for International Exchange of Scholars to Georgia Sotiropoulou is acknowledged. We especially thank Dr Arthur B. Pardee for critical reading of the manuscript.

References

1. Sager R. (1997) Expression genetics in cancer: shifting the focus from DNA to RNA. *Proc. Natl. Acad. Sci. USA* **94:** 952–955.
2. Liang P, Pardee AB. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257:** 967–970.
3. Sager R, Sheng S, Anisowicz A, et al. (1994) RNA genetics of breast cancer: maspin as paradigm. *Cold Spring Harbor Quant. Biol.* **59:** 537–546.
4. Sager R, Anisowicz A, Neveu M, Liang P, Sotiropoulou G. (1993) Identification by differential display of alpha 6 integrin as a candidate tumor suppressor gene. *FASEB J.* **7:** 964–970.
5. Zou Z, Anisowicz A, Hendrix MJC, et al. (1994) Maspin, a serpin with tumor-suppressing activity in human mammary epithelial cells. *Science* **263:** 526–529.
6. Paulson JC, Colley KJ. (1989) Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. *J. Biol. Chem.* **264:** 17615–17618.
7. Tsuji S, Datta AK, Paulson JC. (1996) Systematic nomenclature for sialyltransferases. *Glycobiology* **6:** 5–7.
8. Tsuji S. (1996) Molecular cloning and functional analysis of sialyltransferases. *J. Biochem.* **120:** 1–13.
9. Harduin-Lepers A, Vallejo-Ruiz V, Krzewinski-Recchi M, Samyn-Petit B, Julien S, Delannoy P. (2001) The human sialyltransferase family. *Biochimie* **83:** 727–737.
10. Weinstein J, de Souza-e-Silva U, Paulson JC. (1982) Purification of a Galβ1,4GlcNAc α2,6 sialyltransferase and a Galβ1,3(4)GlcNAc α2,3 sialyltransferase to homogeneity from rat liver. *J. Biol. Chem.* **257:** 13835–13844.
11. Gillespie W, Kelm S, Paulson JC. (1992) Cloning and expression of the Galβ1,3GalNAc α2,3-sialyltransferase. *J. Biol. Chem.* **267:** 21004–21010.
12. Wen DX, Livingston BD, Medzihradszky KF, Kelm S, Burlingame AL, Paulson, JC. (1992) Primary structure of Galβ1,3(4)GlcNAc α2,3-sialyltransferase determined by mass spectrometry sequence analysis and molecular cloning: evidence for a protein motif in the sialyltransferase gene family. *J. Biol. Chem.* **267:** 21011–21019.
13. Datta AK, Chammas R, Paulson JC. (2001) Conserved cysteines in the sialyltransferase sialylmotifs form an essential disulfide bond. *J. Biol. Chem.* **276:** 15200–15007.
14. Livingston BD, Paulson JC. (1993) Polymerase chain reaction cloning of a developmentally regulated member of the sialyltransferase gene family. *J. Biol. Chem.* **268:** 11504–11507.
15. Sjoberg E, Kitagawa H, Glushka J, van Halbeek H, Paulson JC. (1996) Molecular cloning of a developmentally regulated *N*-acetylgalactosamine α2,6-sialyltransferase specific for sialylated glycoconjugates. *J. Biol. Chem.* **271:** 7450–7459.
16. Nara K, Watanabe Y, Maruyama K, Nagai Y, Sanai Y. (1994) Expression cloning of a CMP-NeuAc:NeuAcα2-3Galβ1-

4Glcβ1-1′Cer α2,8-sialyltransferase (GD3 synthase) from human melanoma cells. *Proc. Natl. Acad. Sci. USA* **91:** 7952–7956.
17. Nakayama J, Fukuda MN, Hirabayashi Y, et al. (1996) Expression cloning of a human G_{T3} synthase-G_{D3} and G_{T3} are synthesized by a single enzyme. *J. Biol. Chem.* **271:** 3684–3691.
18. Nakayama J, Fukuda MN, Fredette B, Ranscht B, Fukuda M. (1995) Expression cloning of a human polysialyltransferase that forms the polysialylated neural cell adhesion molecule present in embryonic brain. *Proc. Natl. Acad. Sci. USA* **92:** 7031–7035.
19. Sharon NL, Lis H. (1989) Lectins as cell recognition molecules. *Science* **246:** 227–234.
20. Stoolman LM. (1989) Adhesion molecules controlling lymphocyte migration. *Cell* **56:** 907–910.
21. Feizi, T. (1985) Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens. *Nature* **314:** 53–57.
22. Newman RA, Delia D. (1983) Analysis of the binding of peanut agglutinin (PNA) to leukemic cells and its relationship to T-cell differentiation. *Immunology* **49:** 147–152.
23. Yogeeswaaran G, Salk PL. (1981) Metastatic potential is positively correlated with cell surface sialylation of cultured murine tumor cell lines. *Science* **212:** 1514–1516.
24. Dennis JW, Laferte S, Waghorne C, Breitman ML, Kerbel RS. (1987) β1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science* **236:** 582–585.
25. Akamatsu S, Yasawa S, Tachikawa T, et al. (1996) α2,3-Sialyltransferase associated with the synthesis of CA19-9 in colorectal tumors. *Cancer* **77:** 1694–1700.
26. Itzkowitz SH, Bloom EJ, Kokal WA, Modin G, Hakomori SI, Kim YS. (1990) Sialosyl-Tn: a novel mucin antigen associated with poor prognosis in colorectal cancer patients. *Cancer* **66:** 1960–1966.
27. Gendler SJ, Lancaster CA, Taylor-Papadimitriou J, et al. (1990) Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *J. Biol. Chem.* **265:** 15286–15293.
28. Collard JG, Schijven JF, Bikker A, La Riviere G, Bolscher JGM, Roos E. (1986) Cell surface sialic acid and the invasive and metastatic potential of T-cell hybridomas. *Cancer Res.* **46:** 3521–3527.
29. Dennis J, Waller C, Timpl R, Schirrmacher V. (1982) Surface sialic acid reduces attachment of metastatic tumour cells to collagen type IV and fibronectin. *Nature* **300:** 274–276.
30. Brandley BK, Swiedler SJ, Robbins PW. (1990) Carbohydrate ligands of the LEC cell adhesion molecules. *Cell* **63:** 861–863.
31. Springer TA, Laskey LA. (1991) Cell adhesion. Sticky sugars for selectins. *Nature* **349:** 196–197.
32. Band V, Zajchowski D, Swisshelm K, et al. (1990)) Tumor progression in four mammary epithelial cell lines from the same patient. *Cancer Res.* **50:** 7351–7357.
33. Band V, Sager R. (1989) Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types. *Proc. Natl. Acad. Sci. USA* **86:** 1249–1253.
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.
35. Masiakowski P, Breathnach R, Bloch J, Gannon F, Krust A, Chambon P. (1982) Cloning of cDNA sequences of hormone-regulated genes from the MCF-7 human breast cancer cell line. *Nucleic Acids Res.* **10:** 7895–7903.
36. Kojima N, Yoshida Y, Kurosawa N, Lee Y-C, Tsuji S. (1995) Enzymatic activity of a developmentally regulated member of the sialyltransferase family (STX): evidence for alpha 2, 8-sialyltransferase activity toward *N*-linked oligosaccharides. *FEBS Lett.* **360:** 1–4.
37. Yoshida Y, Kojima N, Kurosawa N, Hamamoto T, Tsuji S. (1995) Molecular cloning of Sia α2,3 Galβ1,4GlcNAc α2,8-sialyltransferase from mouse brain. *J. Biol. Chem.* **270:** 14628–14633.
38. Kozak M. (1984) Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preinsulin in vivo. *Nature* **308:** 241–246.
39. Geremia RA, Harduin-Lepers A, Delannoy P. (1997) Identification of two novel conserved amino acid residues in eukaryotic sialyltransferases: implications for their mechanism of action. *Glycobiology* **7:** 5–7.
40. Drickamer K. (1993) A conserved disulfide bond in sialyltransferases. *Glycobiology* **3:** 2–3.
41. Kurosawa N, Inoue M, Yoshida Y, Tsuji S. (1996) Molecular cloning and genomic analysis of mouse Galβ1,3GalNAc-specific GalNAc α2,6-sialyltransferase. *J. Biol. Chem.* **271:** 15109–15116.
42. Kurosawa N, Kojima N, Inoue M, Hamamoto T, Tsuji S. (1994) Cloning and expression of Galβ1,3 GalNAc-specific GalNAc α2,6-sialyltransferase. *J. Biol. Chem.* **269:** 19048–19053.
43. Wang XC, Vertino A, Eddy RL, et al. (1993) Chromosomal mapping and organization of the human β-galactoside α6-sialyltransferase gene. *J. Biol. Chem.* **268:** 4355–4361.
44. Zhang RD, Fidler IJ, Price JE. (1991) Relative malignant potential of human breast carcinoma cell lines established from pleural effusions and a brain metastasis. *Invasion Metast.* **11:** 204–215.
45. Kurosawa N, Hamamoto T, Lee Y-C, Nakaoka T, Kojima N, Tsuji S. (1994) Molecular cloning and expression of GalNAc α2,6-sialyltransferase. *J. Biol. Chem.* **269:** 1402–1409.
46. Kurosawa N, Takashima S, Kono M, et al. (2000) Molecular cloning and genomic analysis of mouse GalNAc alpha2,6-sialyltransferase (ST6GalNAc I). *J. Biochem.* **127:** 845–854.
47. Kitagawa H, Mattei M-G, Paulson JC. (1996) Genomic organization and chromosomal mapping of the Galβ1,3GalNAc/Galβ1,4GlcNAc α2,3-sialyltransferase. *J. Biol. Chem.* **271:** 931–938.
48. Matsuda Y, Nara K, Watanabe Y, Saito T, Sanai Y. (1996) Chromosome mapping of the G_{D3} synthase gene (SIAT8) in human and mouse. *Genomics* **32:** 137–139.
49. O'Brien SJ, Peters J, Searle A, Womack J, Marshall-Graves JA. (1992) Report of the committee on comparative gene mapping. Chromosome Coordinating Meeting, pp. 758–809; Cuticchia AJ, Pearson PL, Klinger HP (eds). (1993) *Genome priority reports,* Vol 1, Karger, Basel, Switzerland.
50. Solomon E, Ledbetter D, Fain P. (1992) Report of the committee on the genetic constitution of chromosome 17. Chromosome Coordinating Meeting, pp. 473–450; Cuticchia AJ, Pearson PL, Klinger HP (eds). (1993) *Genome priority reports,* Vol 1, Karger, Basel, Switzerland.
51. Futreal PA, Liu Q, Shattuck-Eidens D, et al. (1994) BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266:** 120–122.
52. Kirchweger R, Zeillinger R, Schneeberger C, Speiser P, Louason G, Theillet C. (1994) Patterns of allele losses suggest the existence of five distinct regions of LOH on chromosome 17 in breast cancer. *Int. J. Cancer* **56:** 193–199.
53. Cropp CS, Champeme M-H, Liderau R, Callahan R. (1993) Identification of three regions on 17q in primary human breast carcinomas which are frequently deleted. *Cancer Res.* **53:** 5617–5619.
54. Wen DX, Svensson EC, Paulson JC. (1992) Tissue-specific alternative splicing of the β-galactoside α2,6-sialyltransferase gene. *J. Biol. Chem.* **267:** 2512–2518.
55. Wang XC, O'Harlon TP, Lau JTY. (1989) Rat β-galactoside α-2,6-sialyltransferase genomic organization: alternate promoters direct the synthesis of liver and kidney transcripts. *J. Biol. Chem.* **264:** 1854–1859.
56. Vandamme V, Pierce A, Verbert A, Delannoy P. (1993) Transcriptional induction of β galactoside α2,6-sialyltransferase in rat fibroblast by dexamethasone. *Eur. J. Biochem.* **211:** 135–140.
57. Svensson EC, Soreghan B, Paulson JC. (1990) Organization of the β galactoside α2,6-sialyltransferase gene. Evidence for the transcriptional regulation of terminal glycosylation. *J. Biol. Chem.* **265:** 20863–20868.
58. Passantini A, Hart GW. (1988) Cell surface sialylation and tumor metastasis. Metastatic potential of B16 melanoma variants correlates with their relative numbers of specific penultimate oligosaccharide structures. *J. Biol. Chem.* **263:** 7591–7603.
59. Dennis JW, Laferte S. (1985) Recognition of asparagine-linked oligosaccharides on murine tumor cells by natural killer cells. *Cancer Res.* **45:** 6034–6040.

60. Recchi MA, Hebbar M, Hornez L, Harduin-Lepers A, Peyrat JP, Delannoy P. (1998) Multiplex reverse transcription polymerase chain reaction assessment of sialyltransferase expression in human breast cancer. *Cancer Res.* **58:** 4066–4070.
61. Ikehara Y, Kojima N, Kurosawa N, et al. (1999) Cloning and expression of a human gene encoding an N-acetylgalactosamine-alpha2,6-sialyltransferase (ST6GalNAc I): a candidate for synthesis of cancer-associated sialyl-Tn antigens. *Glycobiology* **9:** 1213–1224.
62. Kono M, Tsuda T, Ogata S, et al. (2000) Redefined substrate specificity of ST6GalNAc II: a second candidate sialyl-Tn synthase. *Biochem. Biophys. Res. Commun.* **272:** 94–97.
63. Samyn-Petit B, Krzewinski-Recchi MA, Steelant WF, Delannoy P, Harduin-Lepers A. (2000) Molecular cloning and functional expression of human ST6GalNAc II. Molecular expression in various human cultured cells. *Biochim. Biophys. Acta.* **474:** 201–211.
64. Lee YC, Kaufmann M, Kitazume-Kawaguchi S, et al. (1999) Molecular cloning and functional expression of two members of mouse NeuAcalpha2,3 Galbeta1,3GalNAc GalNAcalpha2,6-sialyltransferase family, ST6GalNAc III and IV. *J. Biol. Chem.* **274:** 11958–11967.
65. Okajima T, Fukumoto S, Ito H, et al. (1999) Molecular cloning of brain-specific GD1alpha synthase (ST6GalNAc V) containing CAG/Glutamine repeats. *J. Biol. Chem.* **274:** 30557–30562.
66. Lee Y-C, Kurosawa N, Hamamoto T, Nakaoka T, Tsuji S. (1993) Molecular cloning and expression of Galβ1,3GalNAc α2,3-sialyltransferase from mouse brain. *Eur. J. Biochem.* **216:** 377–385.
67. Lee Y-C, Kojima N, Wada E, et al. (1994) Cloning and expression of cDNA for a new type of Galβ1,3GalNAc α2,3-sialyltransferase. *J. Biol. Chem.* **269:** 10028–10033.
68. Kojima N, Lee Y-C, Hamamoto T, Kurosawa N, Tsuji S. (1994) Kinetic properties and acceptor substrate preferences of two kinds of Galβ1,3GalNAc α2,3-sialyltransferase from mouse brain. *Biochemistry* **33:** 5772–5776.
69. Brockhausen I, Yang J-M, Burchell J, Whitehouse C, Taylor-Papadimitriou J. (1995) Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells. *Eur. J. Biochem.* **233:** 607–617.
70. Burchell J, Poulsom R, Hanby A, et al. (1999) An alpha2,3 sialyltransferase (ST3Gal I) is elevated in primary breast carcinomas. *Glycobiology* **9:** 1307–1311.
71. Lloyd KO, Burchell J, Kudryashov V, Yin BW, Taylor-Papadimitriou J. (1996) Comparison of *O*-linked carbohydrate chains in MUC-1 mucin from normal breast epithelial cell lines and breast carcinoma cell lines. Demonstration of simpler and fewer glycan chains in tumor cells. *J. Biol. Chem.* **271:** 33325–33334.
72. Dalziel M, Whitehouse C, McFarlane I, et al. (2001) The relative activities of the C2GnT1 and ST3Gal-I glycosyltransferases determine *O*-glycan structure and expression of a tumor-associated epitope on MUC1. *J. Biol. Chem.* **276:** 11007–11015.