Cloning and Characterization of a Close Homologue of Human UDP-N-acetyl-α-D-galactosamine:Polypeptide
N-Acetylgalactosaminyltransferase-T3, Designated GalNAc-T6

EVIDENCE FOR GENETIC BUT NOT FUNCTIONAL REDUNDANCY

(Received for publication, January 25, 1999, and in revised form, May 26, 1999)

Eric Paul Bennett‡, Helle Hassan‡, Ulla Mandel†, Michael A. Hollingsworth‡, Naoaki Akisawa§, Yoshito Ikematsu§, Gerard Merkx‡, Ad Geurts van Kessel¶, Sigvard Olofsson, and Henrik Clausen**

From the ‡Faculty of Health Sciences, School of Dentistry, DK-2200 Copenhagen, Denmark, §Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska 68198, ¶University Hospital Nijmegen, Department of Human Genetics, 6500HB Nijmegen, The Netherlands, and **University of Gothenburg, Department of Virology, S-413 46 Gothenburg, Sweden

The UDP-GalNAc-polypeptide N-acetylglactosaminyli
transferase, designated GalNAc-T3, exhibits unique
functions. Specific acceptor substrates are used by Gal
NAc-T3 and not by other GalNAc-transferases. The ex
pression pattern of GalNAc-T3 is restricted, and loss of
expression is a characteristic feature of poorly differen
tiated pancreatic tumors. In the present study, a sixth
human UDP-GalNAc-polypeptide N-acetylglactosaminyli
transferase, designated GalNAc-T6, with high similar
ity to GalNAc-T3, was characterized. GalNAc-T6 exhibited
high sequence similarity to GalNAc-T3 throughout the
coding region, in contrast to the limited similarity that
exists between homologous glycosyltransferase genes,
which is usually restricted to the putative catalytic do
main. The genomic organizations of GALNT3 and
GALNT6 are identical with the coding regions placed in
10 exons, but the genes are localized differently at 2q31
and 12q13, respectively. Acceptor substrate specificities
of GalNAc-T3 and -T6 were similar and different from
other GalNAc-transferases. Northern analysis revealed
distinct expression patterns, which were confirmed by
immunocytochemistry using monoclonal antibodies. In con
trast to GalNAc-T3, GalNAc-T6 was expressed in WI38
fibroblast cells, indicating that GalNAc-T6 represents a
candidate for synthesis of oncofetal fibronectin. The re
sults demonstrate the existence of genetic redundancy
of a polypeptide GalNAc-transferase that does not pro
vide full functional redundancy.

The initiation of mucin-type O-linked protein glycosylation is
controlled by a family of UDP-GalNAc:polypeptide N-acetylga
lactosaminyltransferases (GalNAc-transferases)† (EC 2.4.1.41)
(1). To date five distinct members of the animal GalNAc-trans
ferase family have been reported (2–7). The GalNAc-trans
ferase family appears to be highly conserved as nine distinct
genomes have been identified in Caenorhabditis elegans (8). Four
human GalNAc-transferases have been characterized, and
several characteristics of the human family of GalNAc-trans
ferases are now apparent. (i) There is overall sequence similari
ties of approximately 40–45%, with regions of high similarity
(80%) in GalNAc-transferase motifs but little similarity among
the N-terminal regions encoding the cytoplasmic tail, the pu
tative signal anchor sequence, and the stem regions (1); (ii) the
chromosomal localizations and genomic organizations are dif
ferent (1, 9); (iii) the substrate specificities as determined by
in vitro assays are different, but there is overlap among some
substrates, especially those derived from mucin tandem
chains (5, 6, 10); and (iv) the patterns of expression in human
cells and organs are different (2, 3, 5, 6). Furthermore, the
catalytic action of the different GalNAc-transferases can be
cooperative, since glycosylation of certain acceptor sites in the
MUC1 tandem repeat by one GalNAc-transferase is required be
fored other sites can be glycosylated by another GalNAc-
transferase (6). These data suggest that each GalNAc-trans
ferase has distinct biological functions that are mainly deter
mined by the kinetic properties and expression patterns of the
enzymes. Many properties of the enzyme function are still not
fully understood: the importance of sub-Golgi localization (11),
the importance of large variation in the length of stem regions
(1, 7), and the significance of a putative C-terminal lectin-like
domain of approximately 100 amino acids, which does not ap
pear to be essential for catalytic activity (12, 13).

GalNAc-T3 exhibits acceptor substrate specificities not seen
with other enzymes, including glycosylation of a single in vivo
defined O-glycosylation site in fibronectin, which forms the

† The abbreviations used are: GalNAc-transferase, UDP-N-acetyl-
α-D-galactosamine:polypeptide N-acetylglactosaminyltransferase;
GalNAc-T1, -T2, -T3, and -T4 represents human GalNAc-transferases
cloned and expressed by Meurer et al. (4), White et al. (5), Bennett et al.
(5), and Bennett et al. (6) (GenBank® accession numbers are U41514,
X85018, X85019, X92689, and Y08564); GalNAc-T5 represents a rat
gene cloned and expressed by Hagen et al. et al. (7) (GenBank® accession
number AF043944); EST, expressed sequence tag; HIV, human immu
nodeficiency virus; bp, base pairs; kbp, kilobase pairs; PCR, polymerase
chain reaction; FITC, fluorescein isothiocyanate; MALDI-TOF, matrix
assisted laser desorption/ionization mass spectrometry time of flight;
mAb, monoclonal antibody.
oncelfal fibronectin isoform (14) and a single site in the V3 loop of HIV gp120 (5). The unique specificity of GalNAc-T3 for the oncelfal fibronectin peptide was reproduced with plasma fibronectin, clearly indicating the importance of the primary sequence context as a major factor in determining O-glycosylation (10). Furthermore, the specificity of GalNAc-T3 for the HIV sequence has been confirmed in vivo. A reporter construct containing the acceptor sequence was O-glycosylated only if GalNAc-T3 was co-expressed in the host cell (15). This demonstrated that specificity in vitro reflects in vivo specificity. Recent studies with a panel of monoclonal antibodies to human GalNAc-transferases demonstrated that GalNAc-T3 is not expressed in connective tissue cells in normal or tumor tissues or in a fibroblast cell line synthesizing oncelfal fibronectin (16). It is therefore unlikely that GalNAc-T3 represents the native fibronectin GalNAc-transferase activity found in tumor tissues and fibroblast cell lines, as originally described by Matsuura et al. (17). This suggests the existence of an additional GalNAc-transferase with similar properties as GalNAc-T3 but with a different expression pattern.

Here, we report the cloning and expression of such a novel human GalNAc-transferase, which appears to represent a high similarity duplicate gene of GalNAc-T3. The novel GalNAc-transferase, designated GalNAc-T6, displayed an acceptor substrate specificity similar to GalNAc-T3, although GalNAc-T6 showed better kinetic properties with the fibronectin substrate. The two GalNAc-transferases exhibit different expression patterns as analyzed by Northern analysis and immunocytology. The cloned cDNA was expressed in a fibroblast cell line synthesizing oncelfal fibronectin. The existence of high similarity pairs of genes within the GalNAc-transferase gene family is significant to the biological function of this large gene family and of practical significance for studies of transgenic knock-out models.

**EXPERIMENTAL PROCEDURES**

**Identification and Cloning of cDNA for GalNAc-T6—**The dbEST data base at The National Center for Biotechnology Information (NCBI), was searched for sequences similar to the coding region of the human ZAP-Ac-T3 gene (5) using the tBLASTx and BLASTn algorithms. The 5′ region of GalNAc-T3, with no known similarity to other human GalNAc-transferases (bp 1–600), was used to identify a rat EST (GenBank® accession number H32001) predicted to encode a protein sequence with 63% similarity to residues 99–181 of GalNAc-T3. The putative human counterpart of this rat sequence was isolated by PCR with a sense primer, EBHC500 (5′-AGCGGATCCGACAAGACAGTGGTGGTGAGC-3′) and EBHC504 (5′-AGCGGATCCACTCCTGCCTTCC-3′), derived from the rat EST sequence, and an antisense primer, EBHC506N (5′-AGCGGATCCTGGGGATGATCTGGGTTCC-3′), derived from the conserved GalNAc-transferase motif (Fig. 1, panel A). Four cDNA libraries from MKN45 (3), Colo205 (Stratagene), salivary glands (5), and spleen were screened by PCR with 0.5 μm primers and library lysates containing 1 × 107 plaque-forming units. A spleen a ZAP (Stratagene) random-primed cDNA library was prepared from 2.0 μg of human spleen mRNA (CLONTECH) using a time saver cDNA kit (Amerham Pharmacia Biotech) and constructed as recommended by the manufacturer. PCR was performed by using 35 cycles of 95 °C for 20 s, 55 °C for 5 s, 72 °C for 2 min under standard conditions.

Two cDNA libraries derived from salivary glands and spleen yielded a single PCR product, whereas two libraries derived from the cell lines Colo205 and MKN45 gave no product. The PCR products from the salivary gland and spleen libraries were cloned and sequenced, and the sequences were found to be identical. The identified novel human sequence (clone #1) covering the central part of a putative GalNAc-transferase gene was further studied by cloning 5′ and 3′ sequences using rapid cDNA library screening (5) (Fig. 1, panel B). Brie fly, the spleen library was aliquoted into 20 sublibraries, and these were screened by PCR with 5′ primer primer pair A2.12423 and the reverse primer derived from the sequence and the A2AP vector primers T3 and T7, and the products were confirmed by hybridization with either EBHC505 or EBHC505 (5′-AGCGGATCCTGGGGATGATCTGGGTTCC-3′), EBHC505 (5′-AGCGGATCCTGGGGATGATCTGGGTTCC-3′), EBHC505 (5′-AGCGGATCCTGGGGATGATCTGGGTTCC-3′), EBHC505 (5′-AGCGGATCCTGGGGATGATCTGGGTTCC-3′) was used for the 3′ PCR, and EBHC505 (5′-AGCGGATCCTGGGGATGATCTGGGTTCC-3′) was used for the 5′ PCR (Fig. 1, panel B). PCR with EBHC505/TTT yielded a single product of 2 kb for the 3′ sequence, and PCR with EBHC504/T3 gave a single product of 250 bp. Both products were blunt end-cloned and sequenced. The 3′ sequence contained an in-frame stop codon. The 5′ sequence had a potential open reading frame but was shorter than the coding region of GalNAc-T3 and had no translation initiation codon and fibroblast cell lines, as originally described by Matsuura et al. (17). This suggests the existence of an additional GalNAc-transferase with similar properties as GalNAc-T3 but with a different expression pattern.

In Situ Hybridization to Metaphase Chromosomes—Fluorescence in situ hybridization (FISH) was performed on normal human lymphocyte metaphase chromosomes using procedures described previously (6, 18). Briefly, P1 DNA (P1-12423) was labeled with biotin-14-dATP, precipitated with human Cot 1 DNA, dissolved in hybridization solution (2× SSC (1× SSC = 0.15 μ NaCl and 0.015 μ sodium citrate)), 10% dextran sulfate, 1% Tween 20, and 0.05% formamide, pH 7.0, and heat-denatured. Slides were incubated with the probe for 4 h, followed by immunocytochemical detection using avidin fluoroscin isothiocyanate (FITC) and successive steps with rabbit-anti-FITC and mouse-rabbit FITC-conjugated antibodies. Slides were evaluated in a Zeiss epifluorescence microscope, and hybridization signals and chromosomes counter-stained with 4,6-diamidino-2-phenylindole (DAPI) were analyzed using the BDS image software package (Oncor). Recombinant Expression of GalNAc-T6 in Insect Cells—An expression construct of a secreted form of the putative GalNAc-transferase gene, pAcGP67-GalNAc-T6-sol, was prepared by reverse transcription-PCR using primers EBHC514 (5′-AGCGGATCCTGGGGATGATCTGGGGTTCC-3′) and EBHC511 (5′-AGCGGATCCTGGGGATGATCTGGGGTTCC-3′) with BoomH overhangs (Fig. 1, panel A), and the product was cloned into the vector pAcGP67 (Pharmingen) and sequenced. Salivary gland mRNA (CLONTECH) was used as template for reverse transcription-PCR. Control constructs included pAcGP67-GalNAc-T1-sol, pAcGP67-GalNAc-T2-sol, pAcGP67-GalNAc-T3-sol, pAcGP67-GalNAc-T4-sol, and pAcGP67-O-sol, which were prepared as described previously (3, 5, 6). Co-transfection of SF9 cells with pAcGP67 containing the baculovirus expression vector was performed according to the manufacturer’s specifications. Briefly, 0.4 μg of construct was mixed with 0.1 μg of Baculool™ DNA and co-transfected in SF9 cells in 24-well plates. Ninety-six h post-transfection recombinant virus was amplified in 6-well plates at dilutions of 1:10 and 1:50. The titer of amplified virus was estimated by titration in 24-well plates with monitoring of GalNAc-transferase activities (5).

**Polypeptide GalNAc-transferase Assay—**Standard assays were performed in 50 μl of total reaction mixtures containing 25 mM Tris (pH 7.4), 10 mM MnCl2, 0.25% Triton X-100, 50 μM UDP[14C]GalNAc (2,000 cpm/nmol) (Amerham Pharmacia Biotech), 200–500 μM acceptor peptide (see Tables I and II for structures), and 5–10 μl of culture supernatants. In some experiments, as indicated, purified recombinant GalNAc-transferase preparations were used. GalNAc-T6 was purified as described previously (10) using sequential ion-exchange chromatography on Amberlite (IRA95, Sigma) and DEAE-Septahel, S-Sepharose fast flow, and Mono-S (PC1/6.5, Smart-System) (Amerham Pharmacia Biotech) columns. Final purification to apparent homogeneity was performed on S12 gel filtration (Superose 12 pc3/20, Smart-system, Amerham Pharmacia Biotech). Purity and protein concentration of the products were determined by SDS-PAGE and densitometry. Purified recombinant GalNAc-T6 was used to incubate with 2 μl of 2× gels in the gel filtration buffer containing 0.5 μg/ml heparin and 1 μg/ml heparin and 1 μg/ml heparin.
Fig. 1. **Structure of human GalNAc-T6 and cloning strategy.** Panel A, multiple sequence alignment (ClustalW, DNASTAR) of human GalNAc-T3 and -T6. Positions of introns are indicated by open numbered arrows, the putative transmembrane regions are indicated by underlining, and potential N-glycosylation sites are indicated by asterisks. The positions of primers used for the secreted expression construct (EBHC514 and EBHC511) are indicated by arrows. Panel B, schematic representation of the cloning and sequencing strategy. See "Experimental Procedures" for a description.
Cloning and Expression of GalNAc-T6

GalNAc-T3, and 0.05 unit/mg for GalNAc-T4, measured using peptides derived from MUC2, MUC1, and MUC7 tandem repeats (6, 10).

Peptides were synthesized by ourselves, or by Carbiotech (Copenhagaen), or Neosystems (Strasbourg), and quality was ascertained by amino acid analysis and mass spectrometry. For analysis of the donor substrate specificity, assays were performed with 100 μM UDP[14C]Gal or UDP[14C]GlCNac (4,000 cpm/nmol). Products were routinely quantified by scintillation counting after Dowex-1 formic acid chromatography. At least once for all combinations of enzyme sources and peptides, the products were evaluated by C-18 reverse phase chromatography (PC3-293 or mRcP C2/C18 SC2/10 Amersham Pharmacia Biotech, Smart System) with scintillation counting of peptide peak fractions. Peptides and products produced by in vitro glycosylation were confirmed by mass spectrometry, and reaction kinetics were monitored by capillary electrophoresis. Reaction mixtures for preparative glycosylation included 2 mM cold UDP-GalNAc and 25 μg of acceptor peptide in a total volume of 100 μl. Reactions were incubated in the sample carousel of an Applied Biosystems model HT270 (Perkin-Elmer) at 30 °C, and injections were performed at 60-min intervals. Capillary zone electrophoresis was performed on coated fused silica capillaries, 72 cm × 50 μm, with 49 cm between sample injection and optical cell. Electrophoresis were performed at 30 °C using 50 mM phosphate buffer (pH 2.5). Voltage across the capillary was 20 kV in positive mode with the anode at the injection side, and the runs were monitored at 310 nm. At the beginning of each cycle the capillary was flushed with 0.1 M NaOH for 2 min followed by flushing with 50 mM phosphate buffer (pH 2.5) for 4 min. After 8 h of reaction the glycopeptides were purified by C-18 high performance liquid chromatography and analyzed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-TOF). Spectra were acquired on either Voyager-DE mass spectrometer (Perseptive Biosystem Inc.) equipped with delay extraction. The matrix used was 2,5-dihydroxybenzoic acid (25 mg/ml) dissolved in 2:1 mixture of 0.1% trifluoroacetic acid in water and acetonitrile.

Samples dissolved in 0.1% trifluoroacetic acid to a concentration of approximately 80 fmol–2 pmol/μl were prepared for analysis by placing 1 μl of sample solution on a probe tip followed by 1 μl of matrix.

Northern Hybridization—Multiple tissue northern (MTN) blots were obtained from CLONTECH. Cell line blots were prepared with total RNA isolated from human colon and pancreatic adenocarcinoma cell lines SUIT2 and sublines (S2-007, S2-013, S2-020, S2-028, S2-CP9, and S2-VP10), PANC-1, BxPC3, ASPC1, COLO 357, SUIT2 and sublines (S2-007, S2-013, S2-020, S2-028, S2-CP9, and S2-007), MiaPaca, HCG25, HT-29(S2-VP10), PANC-1, HPAF, Capan-2, BxPC3, ASPC1, COLO 357, with 0.1× SSC, 0.1% SDS in a mini-hybridization oven (Hybaid). The microarray was equipped with FITC interference filters and a 75-W xenon lamp (FITC).

RESULTS

Cloning of GalNAc-T6—GalNAc-T6 was identified from a rat EST (H32001), and the human sequence was obtained by a combination of PCR-based cDNA cloning and by genomic cloning. The composite sequence contained an open reading frame of 1869 bp (GenBank® accession number Y08565), which is similar to GalNAc-T3 (1902 bp). An additional 33 bp in the coding region of GalNAc-T3 compared with T6 is attributable to the position of the translation initiation codons in GalNAc-T3, having a longer N-terminal cytoplasmic sequence (Fig. 1, panel A). The entire coding sequence was confirmed by sequencing of P1 clones covering the entire coding sequence. Sequencing on P1 clone, PL-12423, revealed that the coding region of GALNT6 was contained in 10 exons (Fig. 1, panel A). Sequences flanking the introns are shown in Fig. 2. A compar-
ison of intron/exon boundaries between GALNT6 and GALNT3 showed that these were positioned identically.

The deduced sequence of GalNAc-T6 predicts a type II transmembrane protein with a hydrophobic signal anchor sequence in residues 9–27 (Fig. 1, panel A). Both GalNAc-T3 and -T6 have 2 potential N-glycosylation sites; one is conserved in the C-terminal region, whereas the other is located in the putative stem region in a nonconserved position.

Results of fluorescence in situ hybridization revealed that GALNT6 is located on chromosome 12q13 (Fig. 3). No specific hybridization signals were observed at other chromosomal sites. A total of 20 cells in metaphase were analyzed. Several human ESTs with sequences identical to the coding region and available 3'-untranslated-region sequence of GalNAc-T6 were identified.

Expression of GalNAc-T6—Expression of the secreted construct of GalNAc-T6 in Sf9 cells resulted in GalNAc-transferase activity in the culture medium of infected cells that was significantly greater than background values obtained with uninfected controls or cells infected with the histo blood group O* gene (Table I). In general, GalNAc-T6 showed activity and specificity similar to GalNAc-T3, including greater activity with Muc1a as compared with Muc1b (derived from the tandem repeat of MUC1) and unique activities with the HIV V3 loop and the fibronectin peptide. GalNAc-T3 and -T6 did not transfer to the hCG-β peptide derived from human chorionadotropin β-chain. One striking difference was the peptide derived from CD59 (Table II), which was a reasonable substrate for GalNAc-T6 but a poor substrate for GalNAc-T3. The CD59 sequence was identified as a putative O-glycosylation site in a recent study by Rudd et al. (20). CD59 was not a substrate for GalNAc-T2 and -T4, and GalNAc-T1 showed very low activity. GalNAc-T6 had a $K_m$ of approximately 2 mM with CD59, but the $K_m$ of GalNAc-T3 with this substrate could not be measured due to quantities of peptides required. A sequence from the tandem repeat of zonadhesin (21) was a substrate for GalNAc-T1, -T2, -T3, and -T6 (Table II). Analysis of the incorporation of GalNAc residues into the zonadhesin peptide by capillary zone electrophoresis and MALDI-TOF revealed that up to 6 mol of GalNAc could be incorporated. Minor differences in the rate of incorporation was found between the enzymes (not shown).

This study identified another novel unique substrate for GalNAc-T3 and -T6 in the prion protein (22). A disulfide-bonded loop sequence consisting of 36 amino acids is found in the prion protein, and this loop contains two N-glycosylation sites that are utilized both in the normal and disease form of the protein (179CVNTIKQHVTVTGKENETETDVKKMMERVEQMC214) (23, 24). The 15 residues between these two N-linked glycosylation sites contain a cluster of likely O-glycosylation sites. No other obvious potential O-glycosylation sites are found in the prion protein. Two peptides derived from the prion loop, prion-a and prion-b (Table II), were tested as substrates, and only GalNAc-T3 and -T6 utilized the prion-a peptide, whereas no enzymes transferred to the prion-b peptide. GalNAc-T1 did show very low activity with the prion-a peptide. Terminal glycosylation with GalNAc-T3 and -T6 of the prion-a peptide revealed that up to four sites out of five potential sites could be glycosylated (not shown), but the fourth site could not be quantitatively glycosylated. GalNAc-T3 and -T6 showed similar low activities with substrates based on MUC7 and rat submaxillary gland mucin (EA2 peptide), which are derived from mucin tandem repeats and represent efficient substrates for several other GalNAc-transferase including GalNAc-T1 and -T2 (Table I) and rat GalNAc-T5 (7).

A comparative analysis revealed differences in the kinetic parameters of purified recombinant GalNAc-T3 and -T6. GalNAc-T3 showed better catalytic activity with the HIV V3-loop peptide, and GalNAc-T6 showed better activity with the fibronectin peptide (Tables I and II). The former was due to a higher $V_{max}$ for GalNAc-T3, whereas the latter was related to both a lower $K_m$ and higher $V_{max}$ for GalNAc-T6.

**Northern Blot Analysis of Human Organs and Tumor Cell Lines**—Northern blots with mRNA from 23 human adult organs showed that GalNAc-T6 hybridized to a single mRNA of approximately 5 kbp in placenta and trachea, with weak signals in brain and pancreas (Fig. 4). The transcript size in brain appeared to be different, with two faint bands of approximately 6.5 and 3 kb. This pattern is different than that found previously for GalNAc-T3 (5), although both GalNAc-T3 and -T6 were expressed in placenta.

Further analysis of mRNA expression among a panel of pancreatic and breast adenocarcinoma cell lines showed a differential expression pattern for the two genes. GalNAc-T3 was strongly expressed in nine cell lines, weakly expressed in four cell lines, and not detected in five cell lines (Fig. 5). GalNAc-T3

![Fig. 3. Fluorescence in situ hybridization of GALNT6 to metaphase chromosomes. GALNT6-labeled 12q13 is shown.](image-url)

| Peptide          | Sequence                  | GalNAc-T1 | GalNAc-T2 | GalNAc-T3 | GalNAc-T4 | GalNAc-T5 |
|------------------|---------------------------|-----------|-----------|-----------|-----------|-----------|
| Muc1a            | AHGVTSAIPDR              | 2.71 ± 0.28 | 1.47 ± 0.19 | 7.14 ± 0.12 | 0 ± 0 | 4.69 ± 0.16 |
| Muc1b            | RPAGGSTAPPA              | 0 ± 0 | 6.36 ± 0.02 | 0.17 ± 0.02 | 0 ± 0 | 0.15 ± 0.02 |
| Muc2             | PTTPSTSTTMTVTPPTPTC       | 3.84 ± 0.06 | 5.62 ± 0.41 | 4.94 ± 0.27 | 0.79 ± 0.09 | 3.56 ± 0.43 |
| Muc7             | Ac-CPTPSATPPAPPSSAPPETTA | 9.30 ± 0.14 | 14.34 ± 0.15 | 1.42 ± 0.15 | 1.21 ± 0.12 | 0.69 ± 0.16 |
| OSM fragment     | LSESTDIQPQGPGCPGA         | 2.21 ± 0.38 | 0 ± 0 | 0.20 ± 0.06 | 0 ± 0 | 0.50 ± 0.16 |
| EA2              | PTTDSTTPAPPETK           | 11.4 ± 0.25 | 21 ± 0.34 | 4.12 ± 0.28 | 0.92 ± 0.03 | 1.59 ± 0.34 |
| HIV V3 gp120     | Ac-CIRIQQGPRAVTFIGKGNIMR | 0 ± 0 | 0 ± 0 | 2.42 ± 0.01 | 0 ± 0 | 0.88 ± 0.01 |
| Fibronectin      | Ac-PFVTTHPGYD            | 0 ± 0 | 0 ± 0 | 0.82 ± 0.06 | 0 ± 0 | 1.59 ± 0.01 |

*a The assay was performed using culture supernatants of Sf9 cells expressing soluble recombinant human GalNAc-transferases as described under "Experimental Procedures."

*b One unit is defined as the amount of enzyme that transfers 1 μmol of GalNAc in 1 min using the standard reaction mixture.

*c Background values obtained with irrelevant expression constructs were subtracted. Assays with Muc2 and EA2 substrates yielded background levels.
### Table II

| Peptide | Sequence | $K_m$, milliunits/ml | $V_{max}$, milliunits/ml | mRNA/mg protein |
|---------|----------|----------------------|--------------------------|----------------|
| HIVHIBgp120 | Ac-CIRIQRGPGRAFVTIGKIGNMR | NA | NA | NA |
| Fibronectin | Ac-PFVTHPGYD | NA | NA | NA |
| Prion-a | KQHTVTTTTKGEN | 0.4 | ND | ND |
| Prion-b | KGENFTETDIKI | 0.1 | ND | ND |
| Zonadhesin | PTERTTTPTKRTTTPTIR | 0.14 | 0.05 | 0.05 |

#### DISCUSSION

This study demonstrates the existence of a subfamily of genes with high similarity within the large polypeptide GalNAc-transferase gene family. GalNAc-T3 and -T6 have high similarity in DNA and amino acid sequence throughout the coding region, identical organization for nine conserved intron/exon boundaries, and identical organization for nine conserved intron/exon boundaries. GalNAc-T6 has been shown to be expressed in several of these cell lines has previously been confirmed by Western blotting (19). GalNAc-T6 transcript (5 kb) was detected in WI38 fibroblasts (Fig. 6), and only one of the carcinoma cell lines, MiaPaca (Fig. 5), analyzed. MiaPaca did not express GalNAc-T3.

**Generation of Monoclonal Antibodies to Human GalNAc-T6**—A hybridoma line, UH7 (2F3), secreting IgG1 specifically reacting with GalNAc-T6 was selected for use. The specificity of the antibody was confirmed by immunocytochemistry with Sf9 cells expressing human GalNAc-T1, -T2, -T3, -T4, and -T6, and differential immunoprecipitation of GalNAc-transferase activities was derived from the medium of infected Sf9 cells. This strategy for selection and characterization has been described in detail in Mandel et al. (16). UH7 reacted with Sf9 cells expressing GalNAc-T6 and not with Sf9 cells expressing other GalNAc-transferases or irrelevant proteins (not shown). UH7 immunoprecipitated recombinant GalNAc-T6 from spent medium of infected Sf9 cells, whereas GalNAc-T1, -T2, -T3, and -T4 activities did not immunoprecipitate. UH7 removed all GalNAc-T6 activity from the medium, but only low activity (approximately 10%) was recovered in the immunoprecipitate (not shown). UH7 did not react by Western blot analysis. Another antibody, UH8 (2E11) was reactive in Western blots but did not react with the native enzymes by immunocytochemistry or immunoprecipitation. The same finding was found for antibodies to GalNAc-T2 and -T3, where antibodies reactive with the native, active enzymes fail to react by Western blot, whereas other antibodies reactive by Western blot failed to react with the native proteins (16).

**Immunolabeling of Cell Lines and Sperm Cells**—The immunocytochemical staining of GalNAc-T3 and -T6 was distinct in human tumor cell lines (Fig. 7). Previously, we have shown that GalNAc-T3 is not expressed in fibroblast cell lines or connective tissue in normal and tumor specimens (16). This is significant because cultured fibroblasts and fetal and tumor tissues express a GalNAc-transferase activity capable of utilizing the fibronectin peptide substrate form oncofetal fibronectin (17). GalNAc-T6 was expressed in the fibroblast cell line WI38, which produces oncofetal fibronectin (14, 25), as evaluated by immunostaining with mAb UH7 and Northern analysis (Figs. 6 and 7). mAb UH7 did not label the pancreatic carcinoma lines, SUIT2 and ASPC1, in agreement with the Northern analysis and the finding that sublines derived from SUIT2 show differences in expression of GalNAc-T3 (Fig. 7). Anti-GalNAc-T3 mAb UH5 labeled SUIT2 cells heterogeneously (Fig. 7, panel B), which is in agreement with the Northern analysis and the finding that sublines derived from SUIT2 show differences in expression of GalNAc-T3 (Fig. 5) (19). GalNAc-T6 was also detected by immunocytochemistry in A431 cells (not shown).

Previously, we showed that GalNAc-T3 was strongly expressed in ejaculated spermatozoa, whereas neither GalNAc-T1, -T2, or -T4 were expressed (16). As shown in Fig. 7 GalNAc-T6 was not expressed in spermatozoa. Thus, GalNAc-T3 remains the only GalNAc-transferase found in spermatozoa. One likely substrate for GalNAc-T3 in spermatozoa is the zonadhesin cell membrane adhesion molecule, which plays a role in sperm-egg binding (21). Zonadhesin contains a mucin-like tandem repeat region with a degenerate eight-amino acid repeat sequence, and zonadhesin is heavily O-glycosylated. GalNAc-T3 utilized the zonadhesin peptide substrate but exhibited excess substrate inhibition (Table II).

**Cloning and Expression of GalNAc-T6**—25367

---

$^a$ Acceptor substrate concentration used in standard reaction mixture. $^b$ NA, not applicable. $^c$ ND, not determined. $^d$ NM, not measured due to substrate inhibition at concentrations greater than 0.5 mM.
exon boundaries in the coding regions, and similar kinetic properties that are distinct from other GalNAc-transferases. These parameters strongly indicate that GalNAc-T3 and -T6 are derived from a late gene duplication event and raises the possibility that they provide genetic and functional redundancy. However, there were minor differences in the activities of GalNAc-T3 and -T6, and more importantly, the expression patterns of the two transferases were shown to be entirely different by Northern analysis and immunocytology. Thus, although they may represent duplicated genes, the products of the genes must have different functions and apparently do not provide functional back-up in all cells.

The pattern of sequence similarity found between GalNAc-T3 and -T6 differs from that generally found between members of homologous glycosyltransferases. Other members of the GalNAc-transferase family share sequence similarity primarily in the putative catalytic domain, consisting of a major part of the mid- and C-terminal sequence (6, 7). The recently cloned rat GalNAc-T5 had a stem region of more than 500 residues, which underscores the fact that there are large differences outside the catalytic domain (7). One previous exception to this pattern of similarity was found with three of the α-1,4-fucosyltransferases (FUT3, FUT5, and FUT6) (26, 27), which share an exceptionally high degree of sequence similarity. Interestingly, these three fucosyltransferases also share a similar simple genomic organization (one coding exon), and the genes localize to a single locus on chromosome 19 (28). A single putative common ancestral gene for the α-1,4-fucosyltransferase family was identified in cow, suggesting that a duplication event of this gene occurred very late in evolution (29). GALNT3 and GALNT6 share genomic organizations but do not co-localize to a single locus and are found on different chromosomes. Orthologous rodent genes for both GALNT3 (30) (GenBank® accession number U70535) and GALNT6 (GenBank® accession number AJ133523) exist, but it is possible that lower organisms only have one copy of these high similarity copy genes. Thus, Hagen and Nehrke (8) found no evidence of such closely related pairs of GalNAc-transferase genes in the Gal-
in another cell type that lacks a single GalNAc-transferase, specific positions may not be O-glycosylated. Minor differences in the kinetic properties of GalNAc-T3 and -T6 were identified. It is important to note that for practical and economical reasons only representative panels of potential substrates can be studied, which may not reflect the full range of specificities exhibited by the enzymes. GalNAc-T6 showed almost 10-fold better activity with a CD59-derived substrate, which was recently identified as an in vivo O-glycosylation site (20). Detailed kinetic parameters of this substrate with GalNAc-T3 and -T6 could not be studied, since the $K_m$ of GalNAc-T6 was at least 2 m$^2$. GalNAc-T3 and -T6 activities also differed for HIV and fibronectin peptide substrates. GalNAc-T6 showed the best kinetic properties with the fibronectin peptide. The finding that GalNAc-T6 showed relative better kinetic properties with the fibronectin peptide may be significant, since this enzyme and not GalNAc-T3 is expressed in the WI38 fibroblast cell line, which produces O-glycosylated oncofetal fibronectin (Figs. 6 and 7). O-Glycosylation of the fibronectin IIICS domain was previously suggested to be related to an onco-developationally regulated polypeptide GalNAc-transferase activity (17). Previously, we suggested that GalNAc-T3 was a candidate for this activity (10). However, recent immunocytochemical studies with anti-GalNAc-T3 mAbs revealed that GalNAc-T3 is not expressed in fibroblast cell lines, connective tissue cells in normal organs, or in connective tissue associated with carcinomas (16). Since, GalNAc-T6 shows better kinetic properties for synthesis of oncofetal fibronectin and is expressed in WI38 fibroblasts, it is likely that this enzyme represents the GalNAc-transferase activity responsible for induction of oncofetal fibronectin. The finding that GalNAc-T6 shows a more restricted expression pattern than GalNAc-T3 and other GalNAc-Ts corroborates this hypothesis. In preliminary immunohistological studies of squamous cell carcinomas of the mouth, which produce oncofetal fibronectin in the connective tissue compartment of carcinomas (25, 31, 32), we have not yet been able to demonstrate expression of GalNAc-T6 in the connective tissue compartment. GalNAc-T6 was not expressed in normal stratified squamous epithelium of the oral cavity, but variable expression was detected in the epithelial compartment of squamous carcinomas. Further studies are needed to evaluate this, but the initial implication is that expression of GalNAc-T6 is cancer-associated in stratified squamous epithelium.

The high similarity genes, GalNAc-T3 and -T6, are differentially expressed in human organs and cell lines. Previously, Northern analysis of human organs showed that GalNAc-T3 was highly expressed in pancreas and testis, with weaker expression in several other organs including kidney, prostate, and intestine (5). A similar expression pattern of GalNAc-T3 was found in the mouse, and additional organs with high expression levels, were identified (30). GalNAc-T6 has a more restricted expression pattern and was expressed in placenta and trachea and weakly in brain and pancreas (Fig. 4). The Northern analysis of cell lines shown in Fig. 5 further illustrate the restricted expression of GalNAc-T6 and the differential expression of GalNAc-T3 and -T6. Human GalNAc-T1, -T2, and -T4 were found to be more ubiquitously expressed (2, 5, 6). The murine GalNAc-transferases designated GalNAc-T4 and -T5 have more restricted expression patterns, with high expression being found only in stomach, small intestine, colon, and sublingual glands (7, 33). Recently, we have initiated the production of monoclonal antibodies to human GalNAc-transferases to circumvent inherent technical and practical problems and limitations associated with Northern analysis and in situ hybridization.

**Fig. 7.** GalNAc-transferase expression in human cell lines. FITC immunocytology with mAbs UH5 (GalNAc-T3) is shown in panels A—D and with mAb UH7 (GalNAc-T6) in panels E and H. Panels A and E, WI38 fibroblast cells; panels B and F, SUIT2 pancreatic carcinoma cells; panels C and G, ASPC1 pancreatic carcinoma cells; and panels D and H, human ejaculated sperm cells (magnification, ×450).

*GalNAc-transferase family in *C. elegans.*

The kinetic properties of GalNAc-T6 were similar to GalNAc-T3 and quite different from those of other GalNAc-transferases (Table I). It is clear that a number of peptide sequences derived from mucin tandem repeats with high density of O-glycosylation sites serve as substrates for several GalNAc-transferases; however, it is equally clear that each GalNAc-transferase has unique acceptor specificities. Unique substrate peptides have been identified for all the human GalNAc-transferases characterized to date (6, 10). The unique specificities of GalNAc-T3 and GalNAc-T6 have been extensively studied, and several substrates have been identified. The first unique substrate defined for a GalNAc-transferase was the HIV$_{189}$gp120-derived peptide (5). Recently, Nehrke et al. (15) showed that this unique specificity was not solely an in *vitro* phenomenon but is reproduced in an in *vivo* model. Interestingly, immunocytochemical and Northern analysis revealed that GalNAc-T6 and not -T3 is expressed in CD4-positive H9 cells, which are often used for *in vitro* propagation of HIV. However, the expression pattern in freshly isolated CD4-positive lymphocytes is not known. These data suggest that the repertoire of GalNAc-transferases in a cell plays a major role in determining the O-glycosylation pattern. Thus, a glycoprotein may be O-glycosylated at certain positions in one cell type but when expressed

---

2. U. Mandel, K. Schionning, and H. Clausen, unpublished observation.

3. U. Mandel and H. Clausen, unpublished observation.
function of both of these genes using the knock-out strategy. A subfamily is therefore an important basis for studies of the GalNAc-transferase family in normal cells and tissues as well as in disease states. The novel strategy for the generation and selection of hybridomas applied here appears to have overcome past difficulties in making antibodies to glycosyltransferases.

The identification of potential O-glycosylation sites in the prion protein is intriguing. Presently, there is no evidence that the prion protein carries O-glycans (22); however, this may relate to lack of contact with GalNAc-T3 and/or -T6. Northern analysis did show expression of GalNAc-T6 (and not GalNAc-T3) in brain using whole brain mRNA (Fig. 4), but we have no evidence that GalNAc-T6 is co-expressed with the prion protein. Full appreciation of the expression pattern of these enzymes in brain tissues requires detailed immunohistological analysis with monoclonal antibodies. If co-expression occurs, it is highly likely that O-glycosylation would occur in the prion sequence. This is supported by the in vivo data provided for the HIV V3-loop sequence by Nehrke et al. (15) but also by the fact that the acceptor sequence in the prion protein is contained on a sequence with flanking N-glycosylation (22), which would be predicted to be exposed to the appropriate Golgi compartment.

Despite evidence indicating distinct specificities for individual GalNAc-transferases, the first knock-out experiment of a GalNAc-transferase revealed no apparent phenotypic changes in homozygously deficient mice (34). However, it is now clear that the targeted gene was not the intended GalNAc-T1. Instead a close homologue with high sequence similarity and a function of this gene, the function of this gene is dispensable, or the function of this gene is activated (36). The information gathered here for the GalNAc-T3/T6 family has not yet been cloned and expressed; nonetheless, the lack of an observed phenotype when the gene was targeted may be explained by several possibilities: a lack of function of this gene, the function of this gene is dispensable, or the function of the gene is compensated for by another gene. The impaired gene has higher sequence similarity to GalNAc-T1 than other GalNAc-transferases. Furthermore, the genomic organizations of the two genes appear to be identical (35). This may indicate that these two genes represent another subfamily in the GalNAc-transferase family, but it is also important to note that one pseudogene derived from GalNAc-T1 has been identified (36). The information gathered here for the GalNAc-T3/T6 subfamily is therefore an important basis for studies of the function of both of these genes using the knock-out strategy. A parallel type of mannosyl O-glycosylation in yeast has been studied in more detail; although this type of protein glycosylation is initiated by a family of seven homologous mannosyl transferases, disruption studies showed that at least two genes must be targeted to yield changes in growth and viability (37). In conclusion, the present results demonstrate the existence of subfamilies of polypeptide GalNAc-transferase genes.

4 E. P. Bennett and H. Clausen, unpublished observations.

REFERENCES
1. Clausen, H., and Bennett, E. P. (1996) Glycobiology 6, 635–646
2. Homa, F. L., Hollander, T., Eiberg, H., de Haan, D. J., Thomsen, D. R., and Elhammer, A. P. (1993) J. Biol. Chem. 268, 12609–12616
3. White, T., Bennett, E. P., Takio, K., Sorensen, T., Bonding, N., and Clausen, H. (1995) J. Biol. Chem. 270, 24156–24165
4. Meurer, J. A., Naylor, J. M., Baker, C. A., Thomsen, D. R., Homa, F. L., and Elhammer, A. P. (1995) J. Biochem. (Tokyo) 118, 568–574
5. Bennett, E. P., Hassan, H., and Clausen, H. (1996) J. Biol. Chem. 271, 17006–17012
6. Bennett, E. P., Hassan, H., Mandel, U., Mirgorodskaya, E., Roepstorff, P., Burchell, J., Taylor-Papadimitriou, J., Hollingsworth, M. A., Merkx, G., Geerds van Kessel, A., Geurts van Kessel, E., Steffensen, R., and Clausen, H. (1998) J. Biol. Chem. 273, 30472–30481
7. Ten Hagen, K. G., Hagen, F. K., Balya, M. M., Beres, T. M., Van Wuykhusye, B., and Tabak, L. A. (1998) J. Biol. Chem. 273, 27749–27754
8. Homa, F. L., and Nehrke, K. (1998) J. Biol. Chem. 273, 8268–8277
9. Bennett, E. P., Weghuis, D. O., Merkx, G., Geurts van Kessel, A., Elber, H., and Clausen, H. (1998) Glycobiology 8, 547–555
10. Wurzburg, S. B., Hassan, H., Mandel, U., Hassan, E. K., Roepstorff, P., Bennett, E. P., Nielsen, P. A., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) J. Biol. Chem. 272, 23563–23574
11. Bennett, E. P., Hassan, E. K., Stark, A., Olivo, J-C., Oosterhuis, C., Berger, E. G. Clausen, H., and Nilsson, T. (1998) J. Cell Sci. 111, 45–60
12. Imberty, A., Piller, V., Piller, F., and Breton, C. (1997) Protein Eng. 10, 1353–1356
13. Hazes, B. (1996) Protein Sci. 5, 1490–1501
14. Matsuura, H., Takio, K., Titani, K., Greene, T., Levy, S. B., Salyan, M. E., and Hakomori, S. (1991) J. Biol. Chem. 266, 3314–3322
15. Nehrke, K., Hagen, T. K. G., Hagen, F. K., and Tabak, L. A. (1997) Glycobiology 7, 1053–1060
16. Mandel, U., Hassan, H., Therkildsen, M. H., Bygaard, J., Jacobson, M., Juhl, B. R., Dalesten, E., and Clausen, H. (1999) Glycobiology 9, 43–52
17. Matsuura, H., Greene, T., and Hakomori, S. (1989) J. Biol. Chem. 264, 10472–10476
18. Suijkerbuijk, R. F., van de Veen, A. Y., van Echten, J., Buys, C. H., de Jong, B., Oosterhuis, J. W., Warburton, D. A., Cassiman, J. J., Schonk, D., and Geurts van Kessel, A. (1991) Am. J. Hum. Genet. 48, 269–273
19. Satherlin, M. R., Nishimori, I. L., Caudie, J., Bennett, E. P., Hassan, H., Mandel, U., Mack, D., Iwamura, T., Clausen, H., and Hollingsworth, M. A. (1997) Cancer Res. 57, 4744–4748
20. Rudd, P. M., Morgan, B. P., Wormald, M. R., Harvey, D. J., van den Berg, C. W., Davis, S. J., Ferguson, M. A. J., and Dwek, R. A. (1997) J. Biol. Chem. 272, 7229–7244
21. Hardy, D. M., and Garbers, D. L. (1995) J. Biol. Chem. 270, 26025–26028
22. Hay, B., Barry, R. A., Liebesburg, I., Prusiner, S. B., and Lingappa, V. R. (1987) Mol. Cell. Biol. 7, 914–920
23. Haraguchi, T., Fisher, S., Olfsson, S., Endo, T., Groth, D., Tarenino, A., Teplow, D., Burchell, J., Lyle, L., Costache, M., Palcz, M. M., Mollincone, R., Oriol, R., and Julien, R. (1997) RJL J. Biol. Chem. 272, 8764–8773
24. Zara, J., Hagen, F. K., Ten Hagen, K. G., Van Wuykhusye, B. C., and Tabak, H. (1996) Biochem. Biophys. Res. Commun. 236, 38–44
25. Loidon-Rosa, B., Vehil, P., Matsuura, S., Clausen, H., Cuadrado, C., and Burtin, P. (1990) Cancer Res. 50, 1608–1612
26. Mandel, U., Gaggero, B., Reibel, J., Therkildsen, M. H., Dalesten, E., and Clausen, H. (1994) APMIS 102, 655–672
27. Hagen, F. K., Ten Hagen, K. G., Beres, T. M., Balya, M. M., Van Wuykhusye, B. C., and Tabak, H. (1997) J. Biol. Chem. 272, 13843–13848
28. Hennet, T., Hagen, F. K., Tabak, L. A., and Math, J. D. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 12070–12074
29. Marth, J. D. (1996) Glycobiology 6, 701–705
30. Meurer, J. A., Drong, R. P., Hema, F. L., Slightom, J. L., and Elhammer, A. P. (1996) Glycobiology 6, 231–241
31. Gentzsch, M., and Tanner, W. (1996) EMBO J. 15, 5752–5759
32. Kozak, M. (1992) Annu. Rev. Cell Biol. 8, 197–225