Substrate Specificities of Three Members of the Human
UDP-N-Acetyl-α-d-galactosamine:Polypeptide
N-Acetylgalactosaminyltransferase Family, GalNAc-T1, -T2, and -T3*

(Received for publication, February 22, 1997, and in revised form, June 13, 1997)

Hans H. Wandall, Helle Hassan, Ekaterina Mirgorodskaya, Anne K. Kristensen, Peter Roprestorf, Eric P. Bennett, Peter A. Nielsen, Michael A. Hollingsworth, Joy Burchell, Joyce Taylor-Papadimitriou, and Henrik Clausen

From the §School of Dentistry, University of Copenhagen, Nørre Allé 20, 2200 Copenhagen N, Denmark, the ¶Department of Molecular Biology, University of Odense, Campusvej 47, Odense, Denmark, the ¶¶Eppeley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska 68198, and the **Imperial Cancer Research Fund, 44 Lincoln’s Inn Fields, London WC2A 3PX, United Kingdom

Mucin-type O-glycosylation is initiated by UDP-N-acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferases (GalNAc-transferases). The role each GalNAc-transferase plays in O-glycosylation is unclear. In this report we characterized the specificity and kinetic properties of three purified recombinant GalNAc-transferases. GalNAc-T1, -T2, and -T3 were expressed as soluble proteins in insect cells and purified to near homogeneity. The enzymes have distinct but partly overlapping specificities with short peptide acceptor substrates. Peptides specifically utilized by GalNAc-T2 or -T3, or preferentially by GalNAc-T1 were identified. GalNAc-T1 and -T3 showed strict donor substrate specificities for UDP-GalNAc, whereas GalNAc-T2 also utilized UDP-Gal with one peptide acceptor substrate. Glycosylation of peptides based on MUC1 tandem repeat showed that three of five potential sites in the tandem repeat were glycosylated by all three enzymes when one or five repeat peptides were analyzed. However, analysis of enzyme kinetics by capillary electrophoresis and mass spectrometry demonstrated that the three enzymes react at different rates with individual sites in the MUC1 repeat. The results demonstrate that individual GalNAc-transferases have distinct activities and the initiation of O-glycosylation in a cell is regulated by a repertoire of GalNAc-transferases.

To date three human UDP-N-Acetylgalactosamine:polypeptide N-acetylgalactosaminyltransferases (1–3) (GalNAc-transferases) have been identified and characterized (1–4). Although the three GalNAc-transferases show similarities in primary structure with regard to predicted domain structures, sequence motifs, and conserved cysteine residues, the overall amino acid sequence similarity of only 45% suggests that the members of the GalNAc-transferase family have undergone significant changes during evolution. The genes encoding these enzymes are located on different chromosomes and have distinct structures, although some intron positions are conserved, suggesting an evolutionary relationship. The genes are differentially expressed in organs as revealed by Northern analysis (1–3); in particular GalNAc-T3 exhibited a restricted expression pattern. One question addressed here is whether these three GalNAc-transferases are isoenzymes with redundant or unique functions.

Hennet et al. (5) recently addressed this question by analyzing mice rendered deficient in a close homologue of GalNAc-T1 by gene targeting. No obvious phenotypic differences were observed and preliminary characterization of the residual GalNAc-transferase activity with a few substrates did not reveal differences in enzyme activities. There was a reduction in GalNAc-transferase activity in ES cells in which the gene was inactivated. It is difficult to assess the full significance of these findings because the enzyme deleted in these studies is not well characterized with respect to substrate specificity and tissue expression pattern. Disruption of Dol-P-Man:polypeptide mannosyltransferases which initiate O-glycosylation in yeast showed that loss of one enzyme did not affect cell growth and O-glycosylation in a severe manner. In contrast, disruption of two or more genes affected growth or was lethal (6).

The parameters that determine sites of O-glycan attachment to glycoproteins are poorly understood (7–10). Unlike N-linked glycosylation and most other types of protein glycosylation, a consensus peptide sequence motif for acceptor sites has not emerged for either GalNAc- or Man-type O-glycosylation (6, 8–11). Despite extensive studies of the acceptor substrate specificities of different GalNAc-transferase preparations (8, 12–14), analysis of sequences around confirmed sites O-glycosylated in vivo failed to reveal a simple model for prediction of glycosylation (9, 15, 16). Attempts to infer sequence specificity from analysis of the substrate specificities of GalNAc-transferase activities obtained from extracts is likely to be misleading.

* This work was supported by The Danish Cancer Society, the Ingeborg Rulkjær Foundation, the Velux Foundation, the Danish Medical Research Council, the Danish Natural Science Research Council, the Lundbeck Foundation, the Novo Nordisk Foundation, National Institutes of Health Grant 1 RO1 CA66234, and funds from the EU-Biotech and Human Research Fund, 44 Lincoln’s Inn Fields, London WC2A 3PX, United Kingdom.

Contributed equally to the results of this work.

To whom correspondence should be addressed: Henrik Clausen, School of Dentistry, Nørre Allé 20, DK-2200 Copenhagen N, Denmark. Tel.: 45-35326835; Fax: 45-35326505; E-mail: henrik.clausen@odont.ku.dk.

The abbreviations used are: GalNAc-transferase, UDP-N-acetyl-α-d-galactosamine:polypeptide N-acetylgalactosaminyltransferase; MALDI-TOF, matrix-assisted laser desorption mass spectrometry; GalNAc-T1, -T2, and -T3, GalNAc-transferases cloned and expressed by Homa et al. (1), White et al. (2), and Bennett et al. (3), respectively; HPLC, high performance liquid chromatography; CE, capillary electrophoresis; PAGE, polycrylamide gel electrophoresis; Man transferase, Dol-P-Man:polypeptide mannosyltransferases; HIV, human immunodeficiency virus; Bis-Trris, 2-(bis(2-hydroxyethyl)amino)-2-(hydroxymethyl)-propane-1,3-diol.

© 1997 by The American Society for Biochemistry and Molecular Biology, Inc.
due to variable expression of a number of different GalNAc-trans-ferases, which may show distinct specificities for acceptor substrates. Some GalNAc-transf erases may compete for acceptor substrate sites, even if they do not glycosylate the acceptor substrate site (17). Mathematical models designed to predict sequence preferences of O-glycan sites are flawed because of the limited number of sites identified to date and the selected class of glycoproteins these represent (9, 16), and the fact that the analyzed sites were obtained from a number of different organisms and cell types that probably express different repertoires of GalNAc-transf erases.

In the present study we investigated the in vitro specificity and kinetic properties of purified recombinant GalNAc-transf erases, GalNAc-T1, -T2, and -T3. The results demonstrate unique but partly overlapping acceptor substrate specificities among the three enzymes. Specific sites on peptides were glycosylated, however, there were differences in kinetic properties at these sites. The same sites were glycosylated in a 20-mer or a 105-mer peptide based on the MUC1 tandem repeat. Selective specificity of GalNAc-T3 for a 6-mer sequence in fibronectin was maintained with the intact fibronectin molecule. The results indicate that the acceptor substrate specificities of the GalNAc-transf erases is largely dependent on the primary sequence of the acceptor substrate.

EXPERIMENTAL PROCEDURES

Expression and Purification of Recombinant GalNAc-transf erases

Expression constructs of soluble human GalNAc-T1, -T2, and -T3 were prepared in the vector pAcGP67 as described previously (2, 3). The constructs for GalNAc-T1 and -T2 were designed to correspond to the previously identified N-terminal sequence of the purified soluble enzyme. (1, 2). The N terminus of the expressed recombinant forms included the following residues (underlined) derived from the vector construct: T1, NH2-DLGSGRL T2, NH2-DPTGTLLEPPKK; and T3, NH2-DLGSTMEMR. Sf9 cells were grown at 27 °C in TMN-FH medium containing 10% fetal calf serum (Pharmingen). Plasmids pAcGP67-GalNAc-T1-sol, pAcGP67-GalNAc-T2-sol, and pAcGP67-GalNAc-T3-sol were cotransfected with Baculo-Gold™ DNA (Pharmingen) as described previously (3). Recombinant Baculovirus was obtained after two successive amplifications in Sf9 cells grown in serum-containing medium, at 0.6 rpm. Cells which could be loosened by centrifugal agitation were harvested and resuspended in 100 ml of SF-900 II medium in the original shaker bottle and infected with 1:1,000 to 1:5,000 of a stock of the second amplification of virus (3).

After 1 h 350 ml of SF-900 II medium containing 2 mgl of GalNAc from UDP-GalNAc in 1 ml using the standard reaction mixture as described under "Experimental Procedures" with 25 μg of Muc2 peptide as acceptor substrate for GalNAc-T1, 25 μg of Muc1b as acceptor substrate for GalNAc-T2, and 25 μg of Muc1a as acceptor substrate for GalNAc-T3.

### Purification of recombinant GalNAc-transf erases from Serum-free Medium

Purification of recombinant enzymes from serum-free medium was performed as follows (Table I). Approximately 400 ml of medium containing 2–5 units of enzyme were harvested and processed individually. Medium was dialyzed against 25 mM Bis-Tris, pH 6.0, 10 mM NaCl, 2 mM MnCl2, and 2 mM EDTA, centrifuged at 10,000 g, and passed through a 120-ml DEAE (Sigma) column equilibrated in dialysis buffer without EDTA. The excluded fractions were applied to a 30-ml S-Sepharose Fast-flow (Pharmacia) column equilibrated in the same buffer and GalNAc-transf erases were eluted with a gradient of NaCl from 10 to 500 mM. Fractions containing enzyme were pooled and simultaneously dialyzed and concentrated using a Spectrum Dialysis Concentrator with 10,000 cut off (Spectrum). The concentrated GalNAc-T1 and -T2 were diluted 5-fold in Bis-Tris buffer with 10 mM NaCl, applied to a Mono-S column (HR 5/5, Pharmacia), and eluted with a NaCl gradient from 10 to 500 mM. GalNAc-T3 was not subjected to the second cation exchange chromatography as this step inactivated the enzyme. Mono-S fractions of GalNAc-T1 and -T2 as well as concentrated S-Sepharose fractions of GalNAc-T3 were further purified by S12 gel filtration chromatography (PC3/230, Smart System, Pharmacia) and SDS-PAGE using bovine serum albumin as a standard.

### Polypeptide GalNAc-transf erase 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) (Amersham), 5 mM 2-mercaptoethanol (only in assays to determine Km and V max), 0.01–0.5 milliunits of GalNAc-transf erase, and 25 μg of acceptor peptide (see Table II for structures). Peptides were synthesized by ourselves or by Carbiotech (Copenhagen) and Neosystems (Strasbourg), and quality was ascertained by amino acid analysis and mass spectrometry. Products were routinely determined by scintillation counting after Dowex-1 formic acid cycle chromatography. At least once for all combinations of enzyme sources and peptides, the products were evaluated by C-18 reverse phase chromatography (PC3/230 or μRPC C2/C18 SC2.1/10 Pharmacia, Smart System) with scintillation counting of peptide peak fractions. Finally, peptides and products produced by in vitro glycosylation were in most cases also confirmed by mass spectrometry.

GalNAc-transf erase assays used for determination of Km of acceptor substrates were modified to include 200 μM UDP-[14C]GalNAc (2,000 cpm/nmol) with peptides having concentrations from 0.005 to 2 mM. Assays to determine Km for UDP-GalNAc were performed with saturating concentrations of acceptor substrates (GalNAc-T1, 500 μM Muc2; GalNAc-T2, 100 μM Muc2; and GalNAc-T3, 500 μM Muc2). Analysis of specificity for other sugar nucleotides was performed in standard reaction mixtures with 500 μM UDP-GalNAc, UDP-Gal, or UDP-GlcNAc using 250 μM Muc2 and 0.25 milliunits of purified enzymes (spe-
cific activity estimated with Muc2 peptide). Assays were performed in duplicate or quadruplicate. Assays to determine the metal ion requirement were performed in standard reaction mixtures without MnCl₂ using 0.25 milliunits of GalNAc-transferase (specific activity estimated with Muc5C peptide) purified by gel filtration (run in buffer phosphate-buffered saline with 1 M NaCl) in the absence of MnCl₂. Analysis of GalNAc-transferase activity without addition of Mn²⁺ revealed no detectable activity. The activity was assessed with Muc2 peptide in the presence of 5, 10, or 20 mM CaCl₂ or MgCl₂ with MnCl₂ as control.

Preparative glycosylation of peptides was performed with 10–50 nmol of peptide, 0.5–2.5 mmol of UDP-[¹⁴C]GalNAc (10-fold excess of potential Ser/Thr acceptor sites), and 0.25–5 milliunits of GalNAc-transferase (specific activity determined using the relevant acceptor peptide to be glycosylated) in a final volume of 200 μl. Reactions were allowed to incubate for 24–48 h at 37 °C, and at 18–24 h additional enzyme and UDP-GalNAc (50% of originally added) were added. A peptide was considered terminally glycosylated by a GalNAc-transferase when addition of enzyme and UDP-GalNAc did not result in further incorporation over 4 h as estimated by [¹⁴C]GalNAc incorporation, CE, HPLC, or MALDI-TOF. In vitro glycosylation of plasma fibronectin (Sigma) was performed using the standard reaction mixture with 0.5 milliunits of GalNAc-T3 or a mixture of 0.25 milliunits of GalNAc-T1 and 0.25 milliunits of GalNAc-T2 for 6 h at 37 °C. Controls included

FIG. 1. Panels A–C, SDS-PAGE analysis of different steps in the purification scheme of recombinant GalNAc-transferases from serum-free medium. Panel A, GalNAc-T1; panel B, GalNAc-T2; and panel C, GalNAc-T3. The gel was stained with Coomassie Blue. Standards were prestained molecular weight markers as indicated in the margin. Designations of fractions correspond to purification steps in Table 1. Panels D–F, MALDI-TOF analysis of purified GalNAc-transferases (step 5). Panel A, GalNAc-T1; panel B, GalNAc-T2; and panel C, GalNAc-T3. The single and double charged ions for GalNAc-T1 and -T3 have an average Mₙ of 63,516.3 and 65,690.7, respectively. GalNAc-T2 has a lower average Mₙ of 60,361.3. The Mₙ predicted from the amino acid sequences of secreted recombinant GalNAc-T1, -T2, and -T3 are 59,913.7, 60,028.3, and 67,044.1, respectively. The mass analysis corresponds with the molecular weights estimated by SDS-PAGE in panels A–C. The higher mass of GalNAc-T1 than predicted are most likely due to N-glycosylation, and this is also suggested by the heterogeneous peaks obtained. GalNAc-T2 has only one potential N-linked glycosylation site, and this was not predicted to be utilized. In agreement with this the obtained mass of GalNAc-T2 corresponds well with the predicted mass. The observed mass of GalNAc-T3 was lower than predicted, and this could be attributed to cleavage in the stem region. In contrast to T1 and T2 the expression construct of T3 included the entire stem region as the cleavage site for this enzyme is unknown.
heat-inactivated (5 min 95 °C) enzyme. SDS-PAGE Western blotting using monoclonal antibodies FDZ to fibronectin and FDC-6 and 5C10, which specifically detect the oncofetal fibronectin epitope, as well as enzyme-linked immunosorbent assay using the same antibodies.

**Monitoring of in Vitro O-Glycosylation by Capillary Electrophoresis**

A reaction mixture for preparative glycosylation was used with cold 1–2 μM UDP-GalNAc and 0.5–1 μM acceptor peptides in a total volume of 50–100 μl. The assay was incubated in the sample carousel at 30 °C and injections performed at 30–60-min intervals. Capillary zone electrophoresis was performed on a Applied Biosystems model HT270 (Perkin-Elmer). Coated fused silica capillaries, 72 cm in length between sample injection and optical cell were used. Electrophoresis was performed at 30 °C using 50 mM phosphate buffer (pH 7.5). Voltage across the capillary was 20 KV in the positive mode with the anode at the injection side, and the runs were monitored at 210 nm. The assay was incubated in the sample carousel at 30 °C

### Table II

| Peptide    | Peptide sequence | GalNAc-T1     | GalNAc-T2     | GalNAc-T3     |
|------------|------------------|---------------|---------------|---------------|
|            |                  | K<sub>m</sub> | V<sub>max</sub> | K<sub>m</sub> | V<sub>max</sub> | K<sub>m</sub> | V<sub>max</sub> |
| Muc1a      | AHHGTVSAPDTR     | 0.66          | 102           | 2.13          | 105.7         | 0.09          | 63.3          |
| Muc1b      | RPAPGSTAPPDA     | 1.42          | 139.7         | ND<sup>a</sup> | 1.05          | 75.9          | ND            |
| Tap24      | PDDPAPGSGTAPPAC  | ND            | ND            | 0.82          | 114.8         | ND            | ND            |
| Muc2c      | FTTFSPIMTMTPTPTSC | 0.12          | 102.0         | 0.01          | 19.0          | 0.10          | 87.2          |
| Muc2a      | AC-SAPPTGSTAT2   | 0.34          | 43.2          | 0.84          | 54.2          | 1.53          | 42.5          |
| hCG-β      | NFQossaSSSSSSPPLPSPSRPG | NA<sup>a</sup> | NA            | 1.20          | 47.9          | NA            | NA            |
| OSM fragment | LSESTQ LPGGF GGCA | 0.30          | 88.2          | ND            | ND            | 1.61          | 53.7          |
| Erythropoietin | PFDAAASAPLR | ND            | ND            | ND            | ND            | ND            | ND            |
| Erythropoietin | PFDATAAAPLR | ND            | ND            | ND            | ND            | ND            | ND            |
| HIVgp120   | AC-CHRQPGFAVTGKGNMR | NA          | NA            | NA            | NA            | 0.41          | 25.6          |
| VTHPGY     | AC-PFVTHPGY      | NA            | NA            | NA            | NA            | ND            | ND            |

<sup>a</sup> ND, not determined, indicates that although incorporation is observed K<sub>m</sub> was higher than 2 μM and therefore not analyzed due to required quantities of peptides.

<sup>b</sup> NA, not applicable, indicates that no incorporation is observed with this substrate even after prolonged incubations (24 hrs).

<sup>c</sup> Due to substrate inhibition at concentration of 1 μM.

### Structure Determination

**Matrix-assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF)**—All mass spectra, except for the Muc1 105-mer peptide, were acquired on a Voyager-Elite MALDI time of flight mass spectrometer (Perseptive Biosystem Inc., Framingham, MA), equipped with deconvolution and matrix calibration software. Data were acquired on a Voyager-Elite MALDI time of flight mass spectrometer (Bruker-Franzen Analytik, Bremen, Germany). Asp-N endopeptidase digestion—Reactions containing Muc1 105-mer (2 μmol) in 100 μl 0.1 M sodium phosphate (pH 8.0) with 0.2 μg of Asp-N endopeptidase were incubated for 18 h at 37 °C. The digest was injected directly into a reverse phase HPLC column and cleaved peptides eluted by a gradient of 0–90% acetonitrile in 0.1% trifluoroacetic acid. The kinetic parameters of the three GalNAc-transferases in this study are potentially affected by the following factors: (i) the enzymes are derived from human cDNA sequences; (ii) the enzymes are soluble constructs with GalNAc-T1 and -T2 having a N-terminal sequence similar to the forms originally purified and GalNAc-T3 having a longer N-terminal sequence designed to exclude the hydrophobic retention signal and 12

<sup>3</sup> T. Nilsson, personal communication.
to purified human enzymes in membrane bound and/or secreted forms would determine whether or not these factors influence the results; however, this is currently not possible. The following observations suggest that the data obtained is not significantly affected by the design of the expression constructs and the expression system. The kinetic parameters of purified, recombinant GalNAc-T2 were found to be similar to those of the originally purified enzyme (2, 17) with respect to $K_m$ for the acceptor substrates including Muc2. The acceptor substrate specificity of GalNAc-T3 is unique and consistent with an activity detected in organ extracts that was not seen with GalNAc-T1 or -T2 (3, 17). Human GalNAc-T1 has not been purified, but data with purified bovine colostrum GalNAc-T1 and recombinant GalNAc-T1 (bovine placenta cDNA) expressed in the Baculo-system or in COS-7 cells are available (1, 4, 20). The acceptor substrate specificity of purified GalNAc-T1 matched that of bovine GalNAc-T1 expressed in COS-7 cells with respect to Muc1a, Muc1b, and Muc2 peptides. In contrast, however, the $K_m$ for UDP-GalNAc estimated with purified, recombinant GalNAc-T1 (human cDNA) in the present study was higher than that measured with the purified bovine enzyme (14, 21) and recombinant bovine enzyme expressed in COS-7 cells (22).

**Acceptor Substrate Specificities—**

An analysis of acceptor substrate specificity is summarized in Table II. Peptides derived from the tandem repeat region of secreted mucins, MUC2 and MUC5AC, which have a high density of Ser/Thr, were utilized efficiently by all three enzymes. The MUC2 sequence exhibited the lowest $K_m$ with all three enzymes. Previously, affinity chromatography of GalNAc-transferase preparation from placenta with the Muc2 peptide as ligand was found to selectively bind GalNAc-T2, and the non-bound GalNAc-transferase activity was measured to have higher $K_m$ for the Muc2 peptide (17). The relatively low $K_m$ for recombinant GalNAc-T2 (Table II) is consistent with this finding. The estimated $K_m$ of the non-bound fraction was 0.254 mM which is close to the $K_m$ values obtained for both GalNAc-T1 and -T3 (Table II).

All three recombinant enzymes utilized MUC1-derived peptide substrates; however, the kinetics with each of the peptides were very different. GalNAc-T2 showed a lower $K_m$ for MUC1 peptides that included the GSTAP sites, whereas GalNAc-T1 and -T3 showed lower $K_m$ for the peptides with GVTSA sites. This differential activity was also found in previous analysis of the specificity of crude organ extracts (17). Enzyme activity was greater for MUC1 peptides with GVTSA in extracts from rat testis and salivary glands, and human placenta, whereas enzyme activity in extracts from rat kidney and human liver was greater with peptides that included the GSTAP sites. It is hypothesized that these differences in specificity among organ extracts reflect the differential expression of GalNAc-transferases; however, interpretation of these results is restricted by our limited knowledge of the expression pattern of the enzymes and the possible expression of additional un-
Substrate Specificities of Three GalNAc-transferases

A

Muc 1 105-mer - Intact Protein

NONGLYCOSYLATED

GLYCOSYLATED - GalNAc-T1

GLYCOSYLATED - GalNAc-T2

Muc 1 105-mer - after Asp-N Digestion

NONGLYCOSYLATED

GLYCOSYLATED - GalNAc-T1

GLYCOSYLATED - GalNAc-T2

B

-DTRPAPGSTAPPAHGVTSAIP-

Cycle #2 #8 #9 #17 #18

GalNAc - T1

Cycle #2 #8 #9 #17 #18

GalNAc - T2

FIG. 3
Substrate Specificities of Three GalNAc-transferases

Characterized GalNAc-transferases. Acceptor substrate peptides selectively utilized by a single GalNAc-transferase were identified. Previously, an acceptor peptide, HIV-1 V3, derived from the V3-loop of HIV gp120, was found to be glycosylated exclusively by GalNAc-T3 (3), and this was verified using the purified recombinant enzymes (Table II). Another peptide derived from fibronectin (Table II) was also found to be glycosylated only by GalNAc-T3; its \( K_m \) could not be estimated, apparently because of substrate inhibition at concentrations over 1 mM (not shown). Essentially, no incorporation of GalNAc into the fibronectin peptide was detected when GalNAc-T1 or -T2 were used, with less than 5% incorporation after 24-h reactions. The peptide substrate human chorionic gonadotropin \( \beta \) chain contains serine sites previously shown to be utilized by Muc2-affinity purified placenta GalNAc-T2 (17). Only recombinant GalNAc-T2 utilized this substrate, and the \( K_m \) was higher than for most other peptides (1.20 mM). The peptide derived from a fragment (LSESTQQLP-) of ovine submaxillary mucin (23), which exhibits sequence similarity to the N terminus of glycophorin A, was preferentially utilized by GalNAc-T1 with a \( K_m \) of 0.30 mM. GalNAc-T3 showed low activity with this peptide.

A serine glycosylation site on erythropoietin was previously shown to be an efficient \( \text{in vitro} \) serine acceptor sequence with a \( K_m \) of 4.4 mM for purified porcine GalNAc-transferase (24). This porcine enzyme was later suggested to represent a GalNAc-T1 homologue (25). Hagen et al. (4) found that the same peptide with a Thr substitution of the Ser site was a 58-fold better substrate for bovine GalNAc-T1 than the Ser containing peptide. The same Ser-containing erythropoietin peptide was evaluated with the three human recombinant GalNAc-transferases (Table II). There was very poor \( \text{in vitro} \) O-glycosylation of this peptide by the three enzymes, and \( K_m \) values could not be determined. The Thr-substituted peptide was a better substrate for all three enzymes. To evaluate the kinetics of glycosylation of the erythropoietin peptides, the reaction was monitored by CE (Fig. 2) which showed that all three enzymes exhibited similar reaction patterns with both peptides. The data presented here further support the hypothesis that one GalNAc-transferase may utilize both Ser and Thr sites (2, 4, 24). However, as yet the best Ser containing peptide identified is the human chorionic gonadotropin \( \beta \) chain peptide specifically utilized by GalNAc-T2 and the \( K_m \) of this reaction is more than 100-fold higher than the best Thr containing peptide substrate. In agreement with the results presented here the \( \text{in vivo} \) glycosylation of the single Ser glycosylation site in erythropoietin is less efficient than the Thr-substituted site (26). It remains possible that undiscovered enzymes will show better reaction kinetics with the Ser containing peptide.

**Donor Substrate Specificities**—The \( K_m \) for UDP-GalNAc with GalNAc-T1 and -T2 were 10 and 29 \( \mu \)M, respectively, and GalNAc-T1 was found to have a higher \( K_m \) of 62 \( \mu \)M. The \( K_m \) for purified bovine colostrum GalNAc-T1 and recombinant T1 expr...
Evidence for glycosylation of serine included disappearance of the Ser-PTH peak and appearance of a pseudo-peak between Asp-PTH and Asn-PTH. Evidence of threonine glycosylation was a reduction in the Thr-PTH peak and emergence of pseudopeaks between Asn-PTH and Ser-PTH. Previously, we observed that threonine and ST in GSTAP (Fig. 3, panel B). Evidence for glycosylation of serine included disappearance of the Ser-PTH peak and appearance of a pseudo-peak between Asp-PTH and Asn-PTH.
GalNAc glycosylation resulted in complete disappearance of the Thr-PTH peak, and emergence of pseudopeaks around Gin-PTH (17), but the data presented here were obtained with a different automated sequencer, which has a different sensitivity for detection of Ser-PTH and Thr-PTH derivatives. The peptide fragment with only 2 mol of GalNAc incorporated by GalNAc-T1 was also sequenced and found to be glycosylated in Thr at GVTSQA and Thr at GSTAP (not shown).

Analysis of a Single Repeat of MUC1 (TAP24)—The surprising finding that the two GalNAc-T1 and -T2 produced the same final product with the 105-mer peptide in long-term assays despite large differences in the reaction kinetics with the short MUC1-based peptide substrates prompted us to analyze the kinetics of the glycosylation of Muc1 peptides in more detail. Glycosylation of the TAP24 MUC1 peptide, which contains 6 potential acceptor sites, was analyzed by CE at different time points during a 24-h reaction (Fig. 4). Reactions with GalNAc-T1 and -T3 showed three peaks, which corresponded to 1, 2, or 3 mol of GalNAc incorporated at Thr in GVTSA and ST in GSTAP. Reaction with GalNAc-T2 produced the same three peaks representing the same sites of incorporation. In addition, a fourth peak was detected, which corresponded to incorporation of a fourth mole of GalNAc in the N-terminal Thr residue. There was no evidence of incorporation at Ser in GVTSA and Thr in -DTR- in any of the peptides analyzed.

The kinetics of the reactions were different with respect to timing of the order of addition of the first, second, and third moles of GalNAc. The relative product developments of 1, 2, or 3 mol of GalNAc incorporated at Thr in GVTSA and ST in GSTAP over time with GalNAc-T2 being the most unstable (not shown). GalNAc-T1 lost 60% of activity in 24 h. The three enzymes showed comparable levels of inactivation over time with GalNAc-T2 being the most unstable (not shown).

All three enzymes incorporated a third GalNAc only when a substantial fraction of the peptide substrate was converted to a diglycosylated form. Differences between the enzymes ability to incorporate the third GalNAc were also found in that GalNAc-T2 incorporated the third GalNAc much quicker than GalNAc-T1 and -T3. Furthermore, GalNAc-T2 incorporated a fourth GalNAc at the initial threonine in TAP24. To ensure that the observed differences in activity between GalNAc-T1, -T2, and -T3 were not due to differences in stability, the loss of activity for each enzyme was assessed by incubation at 30 °C in the reaction buffer without substrates over a time period of 24 h. The three enzymes showed comparable levels of inactivation over time with GalNAc-T2 being the most unstable (not shown). GalNAc-T1 and -T3 lost 40% of activity and GalNAc-T2 lost 60% of activity in 24 h.

Analysis of 11-mer Peptides Covering Either GVTSQA or GSTAP of MUC1 Repeat—Further details in the sequence of glycosylation was tested using a mixture of two short 11-mer peptides covering either GVTSQA or GSTAP. CE analysis was possible because these two peptides and their glycoforms were well separated by CE. As shown in Fig. 6 (panels A-C) glycosylation of Muc1a (GVTSQA) by all three transferases resulted in one glycoform that was glycosylated at Thr in GVTSQA as evaluated by MALDI-TOF and amino acid sequencing (not shown). Glycosylation of Muc1b (GSTAP) by all three transferases also yielded one glycoform that was glycosylated at Thr in GSTAP (not shown). The peptide design apparently did not
allow incorporation of GalNAc into the serine site because the N-terminal sequence was too short as the 15-mer Muc1b peptide incorporated ST at GSTAP.5 Nevertheless, CE monitoring of the glycosylation reaction with the mixed peptides clearly demonstrated that GalNAc-T1 and -T3 glycosylate the Thr in GVTSAA site with better efficiency than the GSTAP sites, whereas GalNAc-T2 glycosylates Thr in GSTAP more efficiently than Thr at GVTSAA. These data are in accordance with the observed kinetic values for the peptides as well as with the set of longer peptides (Table II).

**Substrate Specificities: The Fibronectin Model**

Matsuura et al. (29) originally identified an O-glycosylation site in the sequence VTHPGY, which was derived from fibronectin, and showed that extracts from fetal and tumor tissues expressed GalNAc-transferase activity capable of O-glycosylating different peptide designs with this sequence under in vitro conditions (28). As discussed above, GalNAc-T3 exclusively utilized this substrate sequence (Table II), and therefore is a candidate for the enzyme activity identified previously. To determine whether or not GalNAc-T3 glycosylates fibronectin, we evaluated plasma fibronectin as a substrate for the enzyme. Plasma fibronectin lacks O-glycosylation at the VTHPGY sequence and two monoclonal antibodies are available to monitor glycosylation at this specific sequence (29, 30). As shown in Fig. 7 O-glycosylation of fibronectin by GalNAc-T3, but not GalNAc-T1 and -T2, created the FDC-6 and 5C10 epitope on fibronectin as evaluated by enzyme-linked immunoadsorbent assay and SDS-PAGE Western analysis, supporting the hypothesis that fibronectin is a physiological substrate for GalNAc-T3.

**DISCUSSION**

Mucin-type O-glycosylation is initiated by a family of polypeptide GalNAc-transferases. The data in this report support the hypothesis that each individual enzyme has a unique function that includes the ability to glycosylate different acceptor substrates as evaluated by in vitro analysis. However, it is also apparent that there is overlap in acceptor substrate specificities, suggesting that the enzymes have some redundant functions. Within the inherent limitations of in vitro analysis, our results suggest that primary sequence of the acceptor site is one determining factor for position and rate of O-glycosylation.

It is clear that the GalNAc-transferases evaluated here utilize a wide range of acceptor sequences. GalNAc-T3 is unique in its ability to utilize substrates with sequences flanking the glycosylation site that contain charged side chains and do not contain proline, which are often associated with O-glycan sites (8, 9, 15). Recently, Nehrke et al. (31) developed an in vivo model for assessment of the influence of flanking sequences for O-glycosylation. Prior studies of the acceptor specificity of bovine GalNAc-T1 suggested that charged residues in the flanking region of the acceptor sequence PHMAQVTVGPGL severely affected the activity (14); however, similar substitutions introduced in a chimeric reporter construct expressed in COS-7 cells revealed little if any effect. One interpretation of this finding is that analysis of acceptor substrate specificities of GalNAc-transferases by in vitro methods does not correctly reflect the in vivo condition. Our results suggest another possibility. In vivo evaluation of the specificity of GalNAc-T1 with the total repertoire of GalNAc-transferases in COS-7 cells will reflect the activities of all endogenous enzymes. The observation of overlapping acceptor specificity among the GalNAc-transferases suggest that several enzymes could be involved in the glycosylation of a single substrate. It is also possible that single site substitutions with charged residues may have different influences on the acceptor specificity of different GalNAc-transferases and thus not reveal effects in this system. Preliminary studies of the HIV-V3 sequence in the same reporter construct show that glycosylation in wild-type COS-7 cells is very low, whereas co-transfection with a Gal-

---

5 N. Kaufman and M. A. Hollingsworth, unpublished data.
NAc-T3 construct resulted in full in vivo glycosylation. Studies of O-mannosylation in yeast has similarly shown that loss of a single Man transferase can be accompanied by a selective loss of in vitro as well as in vivo glycosylation of a specific site without affecting the general activity (32). Thus, further studies are needed to fully understand in vivo specificity of these enzymes.

Mucin-like domains are one major site of O-glycosylation. Analysis of the factors that regulate initiation of O-glycosylation of mucin is important for understanding some aspects of disease-associated aberrant O-glycosylation. The cell membrane-associated mucin termed MUC1 was originally identified as a cancer-associated mucin found in breast and pancreatic cancers (33, 34). Subsequent studies showed that cancer-associated forms of MUC1 arise by differential O-glycosylation of the tandemly repeated region in cancer cells (35, 36). Recent studies have indicated that MUC1 in tumor cells has fewer O-glycan chains and the O-glycans have reduced β1-6GlcNAc branching (37, 38). The data presented here demonstrate that three of five serine and threonine residues in the repeat are acceptors for in vitro glycosylation that is catalyzed by GalNAc-T1, -T2, and -T3. Previously, Stadie et al. (39) reported glycosylation of the same three sites by crude enzyme preparations from skimmed milk, colon and breast carcinoma cell lines, although Ser in GSTAP was only partially glycosylated. Nishimori et al. (40) found no difference of in vitro O-glycosylation of MUC1 based peptides with GalNAc-transferase preparations from normal and cancer tissues. Stadie’s (39) study found that O-glycosylation was initiated at Thr in the GVTSA motif followed by Thr in the GSTAP motif in extracts from all sources tested. This pattern was similar to that determined by us for GalNAc-T1 and -T3, and is in agreement with our previous studies of crude extracts from salivary glands (17). GalNAc-T2 showed an opposite pattern of reactivity with highest efficiency for the GSTAP motif. This activity may not be present in sufficient quantities to be detected in the sources tested by Stadie et al. (39). It is presently unclear what effect differences in kinetic parameters of the individual GalNAc-transferases may have for mucin O-glycosylation in vivo. It is possible that alterations in the expression of particular GalNAc-transferases in cancer cells may result in the production of glycoforms that differ in number of attached O-glycans. Stadie et al. (39) found that MUC1 tandem repeat peptides with 2–3 mol of GalNAc showed a reduced reactivity with the antibody SM3 which defines one cancer-associated form of MUC1 (35, 41).

The use of capillary electrophoresis for analysis of in vitro O-glycosylation of peptides derived from mucins with multiple acceptor sites proved to be a useful technique because it allowed direct assessment of enzyme kinetics on individual acceptor sites. CE was recently used by Hennebicq-Reig et al. (42) to analyze glycosylation of a MUC5C tandem repeat peptide. We adapted this technique to allow direct sampling and electrophoresis of on-going reaction mixtures with individual GalNAc-transferases. Structural analysis of glycoforms corresponding to the peaks separated by CE revealed that each peak represented a unique glycoform with GalNAc residues attached.

6 K. Nehrke, F. Hagen, and L. A. Tabak, personal communication.
at a specific site(s). CE is a powerful tool for monitoring glycosyltransferase reactions and the sensitivity may be increased dramatically by using fluorescence acceptor or donor substrates (43).

Among the first evidence for the existence of multiple GalNAc-transferases was the finding that fetal and cancer tissues selectively expressed GalNAc-transferase activity capable of utilizing a peptide sequence, VTHPGY, derived from fibronectin (28). We showed that only recombinant GalNAc-T3 utilized a peptide sequence, VTHPGY, derived from fibronectin (28). In contrast, the oncofetally regulated activity previously identified (28) utilized either Mn2+ or Mg2+. In contrast, the oncofetally regulated activity previously identified (28) utilized either Mn2+ or Ca2+. This does not support the hypothesis that GalNAc-T3 is the GalNAc-transferase involved in glycosylation of the fibronectin site and suggests the existence of additional enzymes with this specificity. Additional putative GalNAc-transferase genes have been identified and cloned and several of these have recently been expressed and found to have GalNAc-transferase activity.7

One potentially important finding from this work is that GalNAc-T2 exhibited high donor substrate specificity for UDP-Gal that is specifically associated with one acceptor substrate. Initiation of O-glycosylation with galactose has not been reported for eucaryotic cells, but it is found in prokaryotes (44). It is not known if GalNAc-T2 performs Gal O-glycosylation in vivo, since such structures have not been reported in association with Mac2 or any other mucins to the best of our knowledge; however, it is not clear that this has been looked for exhaustively. As expected, anti-Tn antibodies and VLA lectin did not react with the Gal-glycosylated glycopeptide, whereas the glycopeptide was highly reactive with Jacalin (data not shown). Future studies should address whether this donor substrate specificity is found for membrane-bound GalNAc-T2 and if it leads to Gal O-glycosylation in vivo. Other glycosyltransferases have been shown to be able to use two different sugar nucleotides as donor substrates. The histo-blood group A enzyme, which normally utilizes UDP-GalNAc, show weak specificity for UDP-Gal, although with higher Km value (45). Similarly the histo-blood group B enzyme, which normally utilizes UDP-Gal, can utilize UDP-GalNAc in high concentrations (46). A chimeric AB enzyme that utilizes efficiently both sugar nucleotides was produced by introducing a single amino acid substitution into the protein (47). Furthermore, recently it was shown that α-lactalbumin alters not only the acceptor substrate specificity but also the donor substrate specificity of the β1,4Galtransferase so that the enzyme utilizes both UDP-Gal and UDP-GalNAc (48).

In summary the studies reported here show that the kinetic properties of three GalNAc-transferases including acceptor and donor substrate specificities differ significantly. The results suggest that the initiation of O-glycosylation is a selective process that is controlled by the repertoire of GalNAc-transferases expressed. These findings have wide implications for the understanding of O-glycosylation in disease and also for the recombinant expression technology.

REFERENCES

1. Homa, F. L., Hollander, T., Lehman, D. J., Thomsen, D. R., and Elhammer, Å. P. (1993) J. Biol. Chem. 268, 12699–12616

7 E. P. Bennett, H. Hassan, H. Wandall, and H. Clausen, manuscript in preparation.
Substrate Specificities of Three Members of the Human UDP-N-Acetyl-α-d-galactosamine:Polypeptide N-Acetylgalactosaminyltransferase Family, GalNAc-T1, -T2, and -T3

Hans H. Wandall, Helle Hassan, Ekaterina Mirgorodskaya, Anne K. Kristensen, Peter Roepstorff, Eric P. Bennett, Peter A. Nielsen, Michael A. Hollingsworth, Joy Burchell, Joyce Taylor-Papadimitriou and Henrik Clausen

J. Biol. Chem. 1997, 272:23503-23514.
doi: 10.1074/jbc.272.38.23503

Access the most updated version of this article at http://www.jbc.org/content/272/38/23503

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 45 references, 27 of which can be accessed free at http://www.jbc.org/content/272/38/23503.full.html#ref-list-1