Glycopeptide *N*-acetylgalactosaminyltransferase specificities for O-glycosylated sites on MUC5AC mucin motif peptides

Daniel TETAERT*[1], Kelly G. TEN HAGEN†[2], Colette RICHET*, Arnold BOERSMA*, Jean GAGNON‡ and Pierre DEGAND*

*Unité INSERM No. 377, Biologie et Physiopathologie de Cellules Mucipares, Place de Verdun, 59045 Lille Cédex, France, †Center for Oral Biology, Aab Institute for Biomedical Sciences, University of Rochester, 601 Elmwood Avenue, Box 611, Rochester, NY 14642, U.S.A., and ‡Institut de Biologie Structurale JP EBEL, CEA/CNRS/UJF, 41 rue Jules Horowitz, 38027 Grenoble, Cédex 1, France

The recombinant proteins of the two novel UDP-*N*-acetylgalactosamine (GalNAc) glycopeptide:*N*-acetylgalactosaminyltransferases (designated *gp*GaNTase-T7 and *gp*GaNTase-T9) were assayed with O-glycosylated products obtained from the prior action of the ubiquitous transferases (GaNTase-T1 and GaNTase-T2) towards MUC5AC mucin motif peptides (GTTPSPVPTTSTTSAP and peptides with single amino acid substitutions, GTTPS**A**VPTTSTTS**V**P and GTTPSPVPTTS**I**TS**V**P, that are a reflection of mucin molecule polymorphism). *gp*GaNTase-T9 is known to be expressed differentially and more abundantly than *gp*GaNTase-T7 in some tissues; the results of *in vitro* glycosylation also indicates a difference in acceptor substrate specificities between the *gp*GaNTase isoforms. With the use of capillary electrophoresis, MS and Edman degradation, our study suggests that, in the O-glycosylation of mucin-type proteins, approach and recognition signalling by *gp*GaNTase-T7 and *gp*GaNTase-T9 depend largely on the peptide's primary structure (for example the presence of multiple clusters of hydroxy amino acids and the number of GalNAc residues attached to the peptide backbone). O-glycosylation in terms of sites of attachment seems to be less random than previously described and, if sequential reactions are ordered throughout the Golgi stack, the complete O-glycosylation of the mucin molecules seems to be finely tuned to respond to specific damage to, or attack on, epithelia.

Key words: O-glycosylation, UDP-*N*-acetylgalactosamine:polypeptide *N*-acetylgalactosaminyltransferase.

INTRODUCTION

The initial key step in the regulation of mucin-type O-glycosylation is the enzymic transfer of *N*-acetylgalactosamine (GalNAc) from UDP-GalNAc to serine and threonine residues, a reaction catalysed by UDP-GalNAc:*N*-acetylgalactosaminyltransferases (GaNTases). The details of the molecular processes governing the specificity and kinetics of O-glycosylation are still being studied. However, it is now clear that the initiation of O-glycosylation is controlled in a hierarchical manner in the cell and involves a large family of GaNTases [1,2].

At least nine GaNTase isoforms have so far been cloned and functionally expressed from rodent and human cDNA [3–16]; several other genes encoding GaNTases are under investigation. The GaNTase family is extremely old in evolutionary terms, as demonstrated by the existence of multiple members in *Caenorhabditis elegans* in which 11 distinct-sequence GLY homologues of GaNTases share 60–80 % amino acid sequence similarity with the catalytic domain of mammalian GaNTases [17] (GLY refers to glycosylation-related gene products in *C. elegans*).

The repertoire of GaNTase isoforms displays different patterns of tissue-specific expression and varies with cellular differentiation and malignant transformation [18]. It is becoming increasingly clear that the substrate specificity of the large family of GaNTases is the major factor in determining the sites of O-glycan attachments; consequently, the repertoire of GaNTases differentially expressed in cells could control the differential O-glycosylation mechanisms of the mucins [18,19]. The exact substrate specificity of the different members of GaNTases remains largely unexplained and the subject has barely been broached. The emergence of recombinant GaNTases might help in understanding substrate specificity for *in vitro* studies, but few proper experimental structures are currently available for testing in studies of GaNTase activities. The complexity of the GaNTase family suggests that isoforms for specific substrates have to exist and that their glycosylation activity is of biological significance. However, only a few examples of unique GaNTase specificity have been demonstrated so far. These include the O-glycosylation of a peptide substrate derived from the sequence of the V3 loop of the HIV gp120 protein, which has been shown to be glycosylated *in vitro* and *in vivo* exclusively by GaNTase-T3 [6,7]. Other studies *in vitro* have revealed that some GaNTases (GaNTase-T1 and GaNTase-T2) glycosylate a large panel of peptides, whereas others exhibit more restricted (GaNTase-T4 and GaNTase-T5) or unique (GaNTase-T3) substrate preferences [20].

Although no consensus primary amino acid sequence for O-glycosylation has been yet found, there is substantial evidence that the sequence flanking serine, threonine and proline residues significantly affects GalNAc-transferase activity [21]. However, the site specificity of initial O-glycosylation is not merely governed by the peptide sequence around putative target sites. The initial O-glycosylation of a peptide substrate has been shown to influence the subsequent glycosylation events, inducing conformational changes in the peptide and in the accessibility of particular acceptor sites for GaNTases [22]. For example, it has been shown that complete O-glycosylation of the MUC1 tandem repeat peptide requires several GaNTases in a highly ordered process and that human GaNTase-T4 has a highly restricted acceptor substrate specificity that is distinct from that reported

Abbreviations used: GalNAc, *N*-acetylgalactosamine; GaNTase, UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase; G*n*, glycopeptides in which *n* represents the number of O-linked GalNAc residues; *gp*GaNTase, UDP-GalNAc:glycopeptide *N*-acetylgalactosaminyltransferase; MALDI-MS, matrix-assisted laser desorption ionization MS; PTH, phenylthiohydantoin.

[1] To whom correspondence should be addressed (e-mail tetaert@lille.inserm.fr).
[2] Present address: GTTB/NIDCR/NIH, Blg 10, Room 1N113, MSC-1190, Bethesda, MD, U.S.A.

earlier [10]. The occurrence of positive stimulatory effects on initial O-glycosylation that have been shown to be induced by the prior addition of GalNAc could explain the general formation of clustered O-linked glycans on mucins. However, similar positive effects induced by prior glycosylation, as in the MUC1 tandem repeat model, have not yet been described for other mucin motif peptides.

In the present study we examined the role of the peptide moiety and the mode of signal recognition in the O-glycosylation process by the mammalian GaNTase family, with the use of synthetic peptide substrate models derived from sequences that are displayed naturally as polymorphic structures belonging to MUC5AC [23]. In particular, it has been reported that organ systems with discrete accumulation patterns of GaNTase family members include the tissues of, for example, the gastrointestinal tract, lungs, the nervous system and bone [24], in which the *MUC5AC* gene is also highly expressed. The use of the 16-residue mucin motif peptide substrate GTTPSPVPTTSTTSAP (single-letter amino acid codes), deduced from the human *MUC5AC* gene (essentially expressed in tracheo-bronchial and gastric tissues) that contains multiple sites of GalNAc substitution, seemed particularly suitable [25] for also testing the *in vitro* O-glycosylation potential of different mammalian recombinant GalNAc-transferases (GaNTases T1, T2, T3, T4, T5, T7 and T9). The analyses of the resulting products of the O-glycosylation reaction were performed by estimation of the GalNAc incorporation levels (affinity for the substrate), capillary electrophoresis (number of O-linked sites) and Edman degradation after preparative-scale capillary electrophoresis for the purification of some glycopeptide species (the determination of O-linked sites). Finally, by this approach, through the use of MUC5AC motif peptides, we infer that the potential for mucin-specific O-GalNAc formation is governed by at least two novel isoforms characterized as UDP-GalNAc:glycopeptide *N*-acetylgalactosaminyltransferase (*gp*GaNTase), which are distinguishable by their recognition for different O-GalNAc peptide backbones.

MATERIALS AND METHODS

Materials

AG 1-X8 (100–200 mesh, Cl^- form) was purchased from Bio-Rad (Vitry sur Seine, France). Mes, AMP, EDTA, Triton X-100 and unlabelled UDP-GalNAc were from Sigma (St Louis, MO, U.S.A.). UDP-*N*-[1-^{3}H]acetylgalactosamine (8.7 Ci/mmol) was from New England Nuclear Chemicals (Dupont de Nemours, Les Ulis, France). Poly(vinyl alcohol) (molecular mass 15 kDa) was from Fluka (Buchs, Switzerland).

Peptide substrates

The synthetic peptide substrates GTTPSPVPTTSTTSAP, GTTPS*A*VPTTSTTS*V*P and GTTPSPVPTTS*I*TS*V*P were from Neosystem (Strasbourg, France). The purity (more than 95 %) of the peptides was assessed by HPLC analysis and capillary electrophoresis; their masses were verified by MS. Their amino acid sequences were also checked with an Applied Biosystems gas-phase sequencer 477A [12].

Isolation and preparation of recombinant enzyme GaNTase-T9

The isolation of GaNTase-T9 probes and full-length cDNA species was performed as described previously [10]. The conserved amino acid regions EIWGGEN and VWMDEYK were used to design sense and anti-sense PCR primers to amplify products from rat sublingual gland cDNA. These products were cloned, sequenced and used to screen a rat sublingual gland cDNA library. For the generation of GaNTase-T9 secretion constructs, pF4-rT9 was obtained [16].

Functional expression of GaNTases

In brief, 1 μg of pIMKF1 [11], pF1-mT1, pF1-mT2, pF1-mT3, pF1-mT4, pF1-rT5 [11], pF1-rT7 [14] or pF4-rT9 [17] and 8 μl of LIPOFECTAMINE (Life Technologies) were used to transfect a 35 mm well of COS7 cells grown to 90 % confluence in Dulbecco's modified Eagle's medium (Gibco) containing 10 % (v/v) fetal calf serum at 37 °C in air/CO_2 (19:1) as described previously [8]. After 3 days the clarified cell culture medium containing mock supernatant/recombinant enzymes was assayed under the following O-glycosylation reaction conditions: a total volume of 40 μl contained (final concentrations) 1 mM synthetic peptide substrate (10 μl); 125 mM Mes buffer, pH 7.0 (10 μl), containing 0.2 % (v/v) Triton X-100 and 12.5 mM $MnCl_2$; 1.25 mM AMP (5 μl); and the supernatant of recombinant GaNTase (5 μl of each sample at relative concentrations of 1.8–2.8 mg/ml or complemented with 5 μl of 0.25 M sucrose). The reaction was started for 3 or 24 h by the addition of 1 mM UDP-GalNAc diluted by 1.25 nM UDP-[^{3}H]GalNAc [$(0.15–0.2) \times 10^6$ d.p.m.] (5 μl).

Analytical methods

The following procedures for the identification of O-glycosylation reactions (electrospray MS, matrix-assisted laser desorption ionization MS (MALDI-MS) and capillary electrophoresis for analytical or preparative purposes) were performed after sample desalting under conditions described previously [26]. For the separation of the hexadecapeptides, the 2 M formic acid buffer contained 2.5 % (v/v) poly(vinyl alcohol) molecular mass 15 kDa on a P/ACE system model 5000 (Beckman, Fullerton, CA, U.S.A.). Quantitative analyses of capillary electrophoresis profiles were performed by automated integration of the peak areas; calculation of percentages was perfomed by defining the total of parent peptide and newly synthesized glycopeptides as 100 %. To determine the O-linkage sites of the peptide substrates, a preparative-scale procedure was performed with cumulative collection at the precise migration times of the glycosylated fractions (2–20 nmol was isolated) for subjection to Edman degradation [12,27,28].

Identification of O-glycosylation sites

The sequencing procedure to determine the positions of glycosylation was the same as that described in [27]; we used an Applied Biosystems gas-phase sequencer model 477A coupled to an HPLC 120A system for on-line analysis of the phenylthiohydantoin (PTH) derivatives. The assignment of *O*-GalNAc-linked sites was a function of: (1) the PTH-Thr repetitive yield percentage with a decrease at the O-glycosylation position (when compared with the unglycosylated peptide substrate); (2) the appearance of additional PTH products corresponding to the expected PTH-Thr-(*O*-GalNAc) [29].

RESULTS

Incorporation levels by GaNTase isoforms

The MUC5AC mucin motif peptide (GTTPSPVPTTSTTSAP) and their O-glycosylated products obtained after the action of GaNTase-T1 and GaNTase-T2 were subjected to O-glycosylation reaction by the various GaNTase isoforms; incorporation

Table 1 Incorporation rates of [^{3}H]GalNAc by the different GaNTase isoforms with MUC5AC motif peptide (GTTPSPVPTTSTTSAP) and the O-glycosylated preparations

Abbreviation: n.d., not determined. Results are means ± S.E.M.

	Incorporation rate (nmol of GalNAc incorporated per μg of recombinant enzyme)						
Substrate	GaNTase-T1	GaNTase-T2	GaNTase-T3	GaNTase-T4	GaNTase-T5	GaNTase-T7	GaNTase-T9
MUC5AC 'naked' peptide	182 ± 18	284 ± 26	103 ± 6	0.01	47 ± 2	0.01	0.01
O-GalNAc preparation after GaNTase-T1	117 ± 8	144 ± 14	56 ± 4	0.05 ± 0.01	20 ± 2	111 ± 12	14 ± 2
O-GalNAc preparation after GaNTase-T2	171 ± 12	n.d.	118 ± 17	0.01	n.d.	27 ± 2	19 ± 7

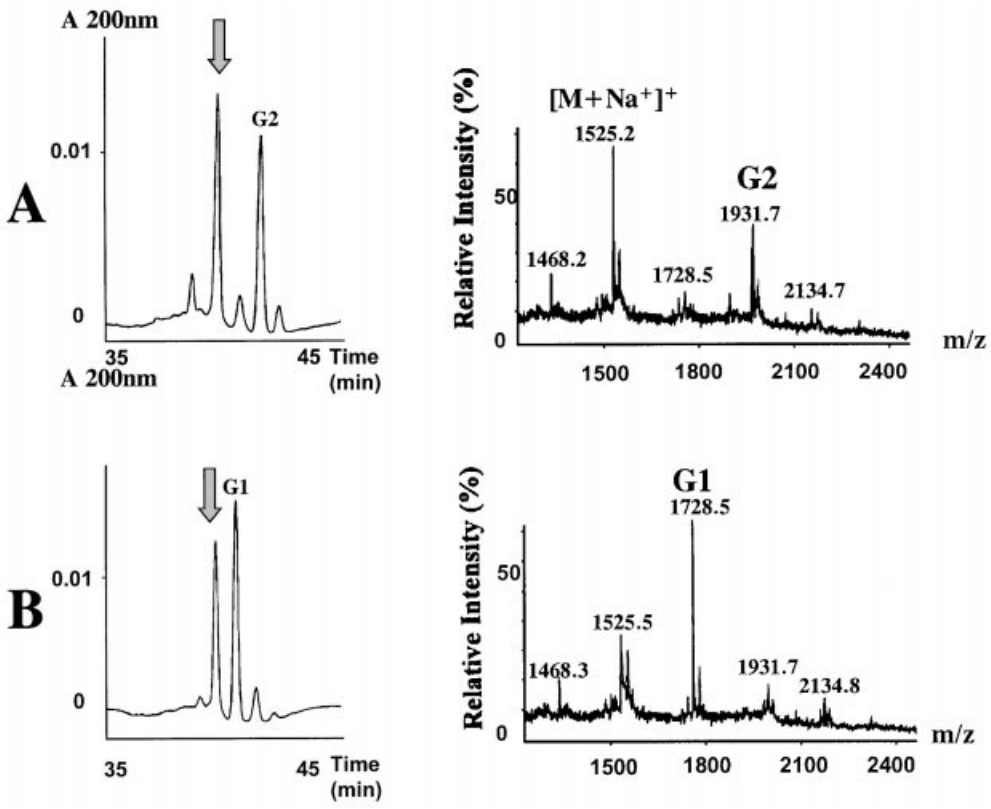


Figure 1 Capillary electrophoresis profiles of the O-glycosylated products obtained by action of the GaNTases

The MUC5AC mucin motif peptide GTTPSPVPTTSTTSAP was incubated for 24 h with GaNTase-T1 (**A**, left panel) and with GaNTase-T2 (**B**, left panel). G1 and G2 are the main glycopeptide products (with one or two O-linked GalNAc residues respectively). The arrow corresponds to the migration time of parent MUC5AC-peptide. The right-hand panels illustrate the corresponding mass spectra of the O-glycosylated products.

levels were evaluated after 24 h of incubation (Table 1). The radioactivities measured indicated some variation between the different GaNTase isoforms: low activities were determined for GaNTase-T3 and GaNTase-T5, whereas the highest level was found for GaNTase-T2 (10-fold background values). A significant value was found for GaNTase-T1 and no activity was detected with GaNTase-T4, GaNTase-T7 or GaNTase-T9. Using the O-glycosylated products as substrates (after action by GaNTase-T1 or GaNTase-T2), the essential change was an activity for the designated *gp*GaNTase-T7 and *gp*GaNTase-T9, whereas GaNTase-T4 has no action towards peptide and/or O-glycopeptide MUC5AC motif peptide as 'naked' or O-GalNAc-substituted forms (Table 1); however, as a positive control this isoform O-glycosylated only the MUC 7 motif (results not shown).

Qualitative analysis of GalNAc substitution on MUC5AC peptide after reaction with GaNTase isoforms

The capillary electrophoresis profiles of the products obtained after 24 h of incubation with the GaNTase-T1 and GaNTase-T2 are given in Figure 1. Identification of the peaks (residual MUC5AC peptide, GTTPSPVPTTSTTSAP, and the O-glycosylated products G*n*, where *n* represents the number of O-linked GalNAc residues) was achieved (1) by comparing the

Table 2 Comparison of the distribution of newly synthesized glycopeptide G*n* after the actions of different GaNTases

Each substrate was incubated first with GaNTase-T1 or GaNTase-T2; the products from this reaction were then incubated with either *gp*GaNTase-T9 or *gp*GaNTase-T7, as indicated. Results are means ± S.E.M. **Bold** numbers represent significant values in the comparisons of percentages.

		Products (% of total)					
Substrate	Enzymes	G1	G2	G3	G4	G5	G6
GTTPSPVPTTSTTSAP	GaNTase-T1	13.7 ± 1.5	**78.1** ± 1.9	8.2 ± 4.2	0	0	0
	GaNTase-T1 then *gp*GaNTase-T9	12.0 ± 2.1	60.6 ± 2.7	**23.3** ± 3.1	4.1 ± 0.3	0	0
	GaNTase-T1 then *gp*GaNTase-T7	11.2 ± 1.1	16.1 ± 2.7	48.8 ± 8.1	**18.1** ± 2.1	5.8 ± 1.1	0.8 ± 0.1
GTTPSPVPTTSTTSAP	GaNTase-T2	**82.3** ± 1.7	14.8 ± 0.8	2.9 ± 0.5	0	0	0
	GaNTase-T2 then *gp*GaNTase-T9	76.0 ± 2.2	15.7 ± 1.1	**6.8** ± 1.1	1.5 ± 0.5	0	0
	GaNTase-T2 then *gp*GaNTase-T7	78.5 ± 2.1	12.8 ± 0.9	3.9 ± 0.3	2.7 ± 0.2	1.4 ± 0.1	0.71 ± 0.02
GTTPS***A***VPTTSTTS***V***P	GaNTase-T1	**51.2 ± 2.2**	29.5 ± 2.1	19.3 ± 2.3	0	0	0
	GaNTase-T1 then *gp*GaNTase-T9	48.6 ± 5.1	24.8 ± 0.8	22.4 ± 1.4	4.2 ± 0.7	0	0
	GaNTase-T1 then *gp*GaNTase-T7	46.2 ± 2.1	4.1 ± 0.8	8.2 ± 1.4	18.3 ± 1.0	17.2 ± 0.2	6.0 ± 0.2
GTTPSPVPTTS***I***TS***V***P	GaNTase-T1	**60.4 ± 3.0**	39.4 ± 0.2	0.2 ± 0.1	0	0	0
	GaNTase-T1 then *gp*GaNTase-T9	36.6 ± 1.8	38.0 ± 2.4	24.5 ± 0.5	0.8 ± 0.1	0	0
	GaNTase-T1 then *gp*GaNTase-T7	38.8 ± 4.1	38.9 ± 2.1	20.9 ± 1.2	0.7 ± 0.4	0	0

profiles of the reaction products obtained in either the presence or the absence of UDP-GalNAc and also those of products resulting from incubations with the mock supernatant (lacking GaNTases) (results not shown), and (2) by identification by MALDI-MS of the masses of the different glycopeptides, finding an additional mass of 203 to the peptide $[(M + Na^+)]^+$ at m/z 1525.2) for each O-linked GalNAc residue. The assignment of capillary electrophoresis fractions, G1 and G2, after the action of GaNTase-T1 and GaNTase-T2 was reinforced by the similarity between relative percentages of the peak area after capillary electrophoresis and the signal intensity in MALDI-MS. In fact, under our 24 h standard conditions (with a molar stoichiometry between nucleotide-sugar and peptide substrate of 1 mM/1 mM) GaNTase-T1 produced mainly G2 (at m/z 1931.5), a di-GalNAc-glycosylated MUC5AC peptide, whereas GaNTase-T2 produced mainly G1 (at m/z 1728.4), a mono-GalNAc-glycosylated MUC5AC peptide (Figure 1). We noted also the formation of an extra peptide fragment at m/z 1468.2 (peptide without its N-terminal glycine residue). The integrations of all peak areas from the profiles (parent peptide and different newly synthesized glycopeptide G*n*) were taken into account to calculate the percentage and distribution of each glycopeptide after 24 h of incubation with GaNTase isoforms: in brief, we noted the presence of a 78 % peak of G2 for GaNTase-T1 (Table 2) and a 82 % peak of G1 for GaNTase-T2 (Table 2).

Evaluation of the co-operative influence between the ubiquitous GaNTase-T1 and GaNTase-T2 isoforms and the *gp*GaNTase-T7 and *gp*GaNTase-T9 isoforms was then achieved with capillary electrophoresis analyses (Figure 2). The capillary electrophoresis profiles showed some discrepancies of action of *gp*GaNTase-T7 and *gp*GaNTase-T9: the glycopeptide products obtained from GaNTase-T1 action (essentially a di-GalNAc peptide, G2) (Figure 2a) was a particular substrate for the subsequent formation of up to G5 glycopeptides by *gp*GaNTase-T7 (Figure 2b), whereas *gp*GaNTase-T9 seemed less active, adding one or two GalNAc residues to generate only G4 glycopeptide in small amounts (Figure 2c). The percentage expression of the newly synthesized glycopeptides showed more clearly the differences between the activities of *gp*GaNTase-T7 and *gp*GaNTase-T9 (Table 2). In contrast, the glycopeptide produced by the action of GaNTase-T2 was mainly a mono-GalNAc peptide, G1 (Figure 2d), a better substrate for *gp*GaNTase-T9 to form larger amounts of G3 and G4 (Figure 2e), whereas *gp*GaNTase-T7 seemed to generate only a small amount of multi-GalNAc-peptide (Figure 2f). Table 2 highlights these discrepancies between the two activities: GaNTase-T2 made essentially G1 but multigalactosaminylation was differently distributed in G2–G5. These results therefore confirm the existence of at least two different activities for *gp*GaNTase-T7 and for *gp*GaNTase-T9: if the *gp*GaNTase-T9 activity required one GalNAc residue for recognition and the further step-by-step addition of GalNAc residues, *gp*GaNTase-T7 acted only when two GalNAc residues were already present on the MUC5AC peptide backbone.

Positions of O-linked GalNAc on the MUC5AC peptide after action of GaNTase isoforms

The identification of GalNAc linked to threonine/serine residue sites was then performed on the MUC5AC motif glycopeptide preparations obtained after reaction with GaNTase-T1 and GaNTase-T2 as well as those resulting from these products after further glycosylation with *gp*GaNTase-T7 and *gp*GaNTase-T9. The glycopeptide fractions were purified after cumulative analytical capillary electrophoresis runs; each individual purified glycopeptide peak (2–20 nmol per isolated fraction) was subjected to Edman amino acid sequencing analysis. The position of a O-linked modification was indicated by the presence of a lower quantity of PTH-Thr than expected, concomitant with the detection of a PTH derivative of GalNAc-Thr. Post-source decay MS–MS was also used for determination of the positions (X. Czeszak, G. Ricart, D. Tetaert, J. C. Michalski and J. Lemoine, unpublished work).

The preferential order of substitution by GalNAc residues into the synthetic peptide mimicking the tandem repeat sequence of MUC5AC incubated with UDP-GalNAc is summarized in Scheme 1. The formation of the di-GalNAc O-linked glycopeptide G2 at positions 13 and 3 (i.e. successively on Thr-13 followed by Thr-3) was obtained mainly with GaNTase-T1 (Scheme 1A), whereas GaNTase-T2 produced essentially the mono-GalNAc O-linked glycopeptide G1 at Thr-9 (Scheme 1B). The activities of the *gp*GaNTases *gp*GaNTase-T7 and *gp*GaNTase-T9 were to substitute the threonine residues in a

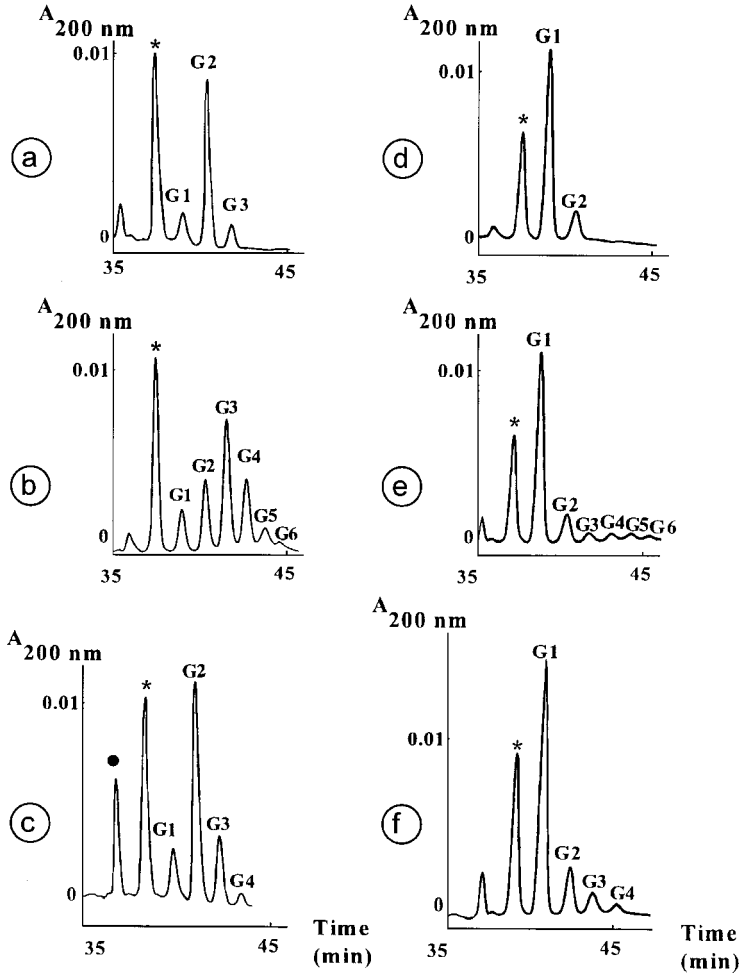


Figure 2 Comparison of capillary electrophoresis profiles between the MUC5AC motif peptide products of O-glycosylation obtained by action of GaNTase-T1 and GaNTase-T2 followed by the action of *gp*GaNTase-T7 and *gp*GaNTase-T9

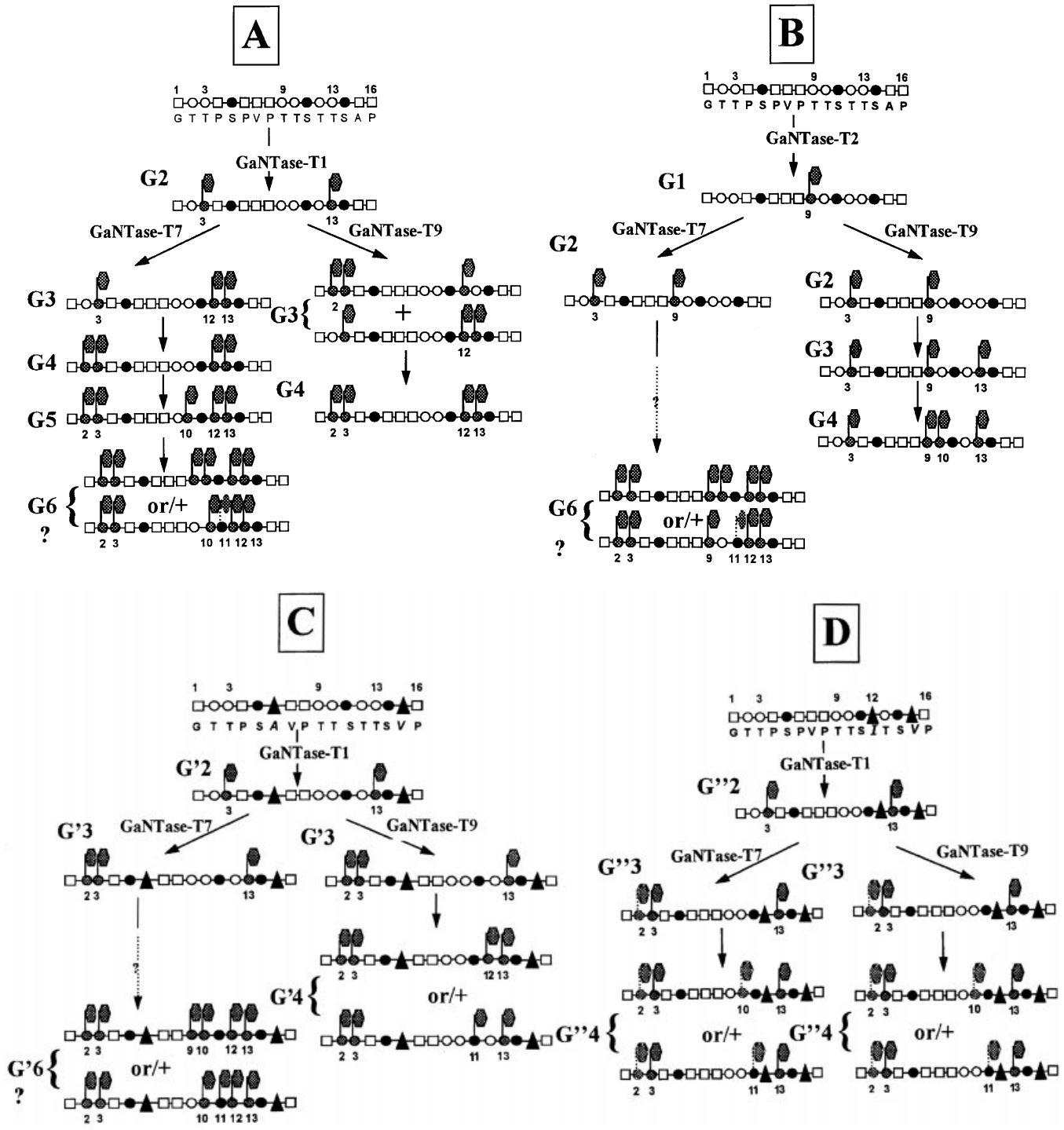
(**a**) Action of GaNTase-T1 and subsequent O-glycosylation of this GalNAc-peptide preparation with *gp*GaNTase-T7 (**b**) and with *gp*GaNTase-T9 (**c**). (**d**) Action of GaNTase-T2 and subsequent O-glycosylation of this GalNAc-peptide preparation with *gp*GaNTase-T7 (**e**) and with *gp*GaNTase-T9 (**f**). The MUC5AC mucin motif peptide GTTPSPVPTTSTTSAP was used as the substrate; G1 to G6 are the resulting O-linked peptides containing one to six GalNAc residues. Symbols: *, parent MUC5AC peptide; ●, fragment of parent peptide without the N-terminal glycine residue.

precise order and also to yield the position isomers for G6 (action of *gp*GaNTase-T7) (Scheme 1A) and for G3 (action of *gp*GaNTase-T9) (Scheme 1B).

O-glycosylation reaction towards modified MUC5AC motif peptides by GaNTase isoforms

To investigate further how recognition by the *gp*GaNTase isoenzymes (*gp*GaNTase-T7 and *gp*GaNTase-T9) of MUC5AC peptide-sequence domains (domain rich in proline residues compared with the domain tandem repeat) might occur, three glycopeptide preparations were made based on our model peptide (peptide 1, GTTPSPVPTTSTTSAP) and those that corresponded to sequences found naturally in mucin-type MUC5AC glycoprotein (peptide 2, GTTPS*A*VPTTSTTS*V*P,and peptide 3, GTTPSPVPTTS*I*TS*V*P). These peptides were subjected to the action of GaNTase-T1 for 24 h to produce glycopeptide preparations (G2, G′2 and G″2 respectively); incorporation rates are presented in Table 3. After this prior O-glycosylation by GaNTase-T1, the incorporation levels by isoforms *gp*GaNTase-T7 and *gp*GaNTase-T9 on the resulting glycopeptide products revealed a higher rate for the G2 preparation containing essentially the glycopeptide GTT(GalNAc)PSPVPTTSTT(GalNAc)SAP; the *gp*GaNTase-T7 rates were halved when modification occurred near Thr-3 of G′2 and decreased to 1/10 for modification in Thr-12 (G″2). Comparison of rates of incorporation by *gp*GaNTase-T9 revealed only a slight difference between preparations of peptides 1 and 2, whereas a lower incorporation level was obtained with peptide 3 (Table 3).

The distribution of different newly synthesized glycopeptides G*n* varied depending on the glycopeptide preparations (Table 2); Scheme 1 summarizes the results of the positions in which threonine residues were assigned as O-GalNAc sites after the action of the two *gp*GaNTases with the use of Edman degradation and/or post-source decay MS–MS. In brief, we noted that the position of a threonine site and its environment were important: for example, Thr-3 in peptide 2 was perturbed by the substitution of an alanine residue in the proline-rich sequence with the formation of a small amount of G2, but *gp*GaNTase-T7 preserved its same qualitative efficiency (the formation of G4, G5 and G6) at a lower level. *gp*GaNTase-T9 activity seemed slightly higher in producing G3 and G4 (Table 2). In addition, the main site Thr-



Scheme 1 Schematic representation of the putative pathways of GalNAc incorporation into different positions of MUC5AC peptides

(**A**) Incubation with *gp*GaNTase-T1 followed by the action of *gp*GaNTase-T7 and *gp*GaNTase-T9 (deduced from results described previously [16]). (**B**) Incubation with *gp*GaNTase-T2 followed by the action of *gp*GaNTase-T7 and *gp*GaNTase-T9. (**C**) Expected positions of the derived MUC5AC peptide 2 (GTTPS***A***VPTTSTTS***V***P) incubated with *gp*GaNTase-T1 followed by the action of *gp*GaNTase-T7 and *gp*GaNTase-T9. (**D**) Expected positions of derived MUC5AC peptide 3 (GTTPSPVPTTS***I***TS***V***P) incubated with *gp*GaNTase-T1 followed by the action of *gp*GaNTase-T7. Symbols: ○, threonine residue; ●, serine residue; □, other amino acid residue; ▲, modified amino acid residue; linked hexagonal and round 'honeycomb stippled' shapes, O-linked GalNAc amino acid residue.

13 in peptide 3 was hindered by the presence of an isoleucine residue in position 12 and the actions of *gp*GaNTase-T7 and *gp*GaNTase-T9 were disturbed thereby. For *gp*GaNTase-T7, a low activity was found with the formation of G3 only; similar amounts of the glycosylated products were recovered after the action of *gp*GaNTase-T9 (Table 2). It seemed that the smaller

Table 3 Incorporation rates of [^{3}H]GalNAc by the different *gp*GaNTase isoforms with the three MUC5AC motif peptides and the O-glycosylated preparations

The peptide substrates used were GTTPSPVPTTSTTSAP (peptide 1), GTTPS***A***VPTTSTTS***V***P (peptide 2) and GTTPSPVPTTS***I***TS***V***P (peptide 3). Results are means ± S.E.M.

	Incorporation rate (nmol of GalNAc incorporated per μg of enzyme)	
Preparation	*gp*GaNTase-T7	*gp*GaNTase-T9
Peptide (1)-glycopeptide	111 ± 6	14 ± 4
Peptide (2)-glycopeptide	53.2 ± 4	12 ± 2
Peptide (3)-glycopeptide	10.2 ± 0.8	8.7 ± 1

amount of G2 formed by GaNTase-T1 was poorly recognized by the *gp*GaNTase-T7 isoform, suggesting an inhibition by the presence of a hydrophobic residue (Ile) replacing a threonine residue at position 12 or by the formation of a G″2 glycopeptide other than the (3,13)-substituted species.

DISCUSSION

Processing by O-glycosylation requires the co-ordinated actions of a large number of enzymes, including sialyltransferases, fucosyltransferases, *N*-acetylglucosaminyltransferases and galactosyltransferases; in the first instance the GaNTases are particularly important. Furthermore, the initial step of O-glycosylation by the action of GaNTases on hydroxy amino acids (threonine and serine) seems less random than expected and the occurrence of several isoforms, acting in a hierarchical manner, should be necessary for the correct completion of O-glycosylated molecules in terms of the numbers and lengths of glycan chains. In previous studies, apoproteins, peptidic fragments and/or synthetic peptides have commonly been used to evaluate the O-glycosylation reactions *in vitro* in an attempt to define the exact substrate acceptor specificities for the different GaNTases (reviewed in [30]). For such investigations, the peptide GTTPSPVPTTSTTSAP, mimicking the human mucin tandem repeat of MUC5AC, has been thought to be of value because of the existence of four identical repeats within the sequence of the clone JER58 [23]; similar repeats in JER58 (GTTPS***A***VPTTSTTS***V***P and GTTPSPVPTTS***I***TS***V***P) with changes that correspond to double-amino-acid polymorphism; and the X^{2}TPXP^6 sequence as a signal favouring both efficient glycosylation [31,32] and also perhaps the accessibility of the hydroxy group of Thr-3 and Thr-13 and/or Thr-9. It is consequently an excellent substrate because a pool of enzymes (such as the microsomal preparation of human gastric mucosa) permits O-linkages to six residues (glycopeptides G1 to G6), [25] whereas an individual isoform has a restricted specificity, for example the formation of G2 (at positions 3 and 13) by GaNTase-T1 [12].

In the present study, the characteristics of different GaNTases were investigated through the use of our MUC5AC motif peptide. Moreover, the *gp*GaNTase activities that require the prior activity of another isoform of GaNTase (such as the ubiquitous GaNTase-T1 or GaNTase-T2) are especially estimated according to the further addition of GalNAc residues on other hydroxy amino acids within the glycoprotein backbone. The study of functional expression *in vitro* allowed us to distinguish between *gp*GaNTase-T7 and *gp*GaNTase-T9, which differ in their recognition of substrate depending on the O-GalNAc peptide environment (mono-GalNAc and/or di-GalNAc substitutions) as well as their polygalactosaminylation level (up to G6 for *gp*GaNTase-T7 and G4 for *gp*GaNTase-T9). *gp*GaNTase-T9 is also notable in being differentially and more abundantly expressed in tissues of embryos during late organogenesis, which increased substantially the distinction between the two *gp*GaNTases [16].

On the basis of sequence relationships, a similarity between *N*-acetylgalactosaminyltransferases and the B-chain of plant toxins has been detected by fold recognition methods (PSI-BLAST or hidden Markov modelling) and thus the presence of a ricin-like sequence was assumed to correspond to the glycan-binding domain [33]. Recently, the general features of the GaNTase structure have been subdivided more precisely into different putative domains: the transmembrane anchor, the stem, the catalytic domains (glycosyltransferase GT1 motif and Gal/GalNAc-T motif) and the ricin-like lectin motif [19]. Furthermore, the glycosyltransferases were thought to be organized as hetero-oligomeric complexes in the membrane of the Golgi cisternae [34]. An elaborate organization could be devised in which co-operativity between the multiple domains of several glycosyltransferases (*gp*GaNTases and/or GaNTases) would increase the sugar/peptide-binding affinity and would participate in a greater specificity of mucin O-glycosylation whatever the peptide sequence. In addition, on the basis of molecular modelling and NMR data, it has been suggested that GaNTases require at least three points of contact to stabilize the spatial orientation of the target hydroxy groups [35]. In the MUC5AC peptide motif, the structural feature of the domain with the X^{2}TPXP^6 sequence therefore seemed of great importance, because the O-glycosylation rates could depend on possible steric hindrances for the *gp*GaNTases in attempting to reach some essential Thr/Ser residues in or near this motif, for example in the PS***A***VP peptide structure. Moreover, the sequence motif X^{2}TPXP^6, which is thought to serve as a signal for mucin-type O-glycosylation [31,32], is present in the MUC5AC peptide, and our results support the hypothesis of an O-glycosylation effectiveness similar to full *in vivo* biological activity towards this kind of peptide structure in mucin.

With point mutations, found as individual polymorphic structural features in some MUC tandem repeat sequences [36], the approach by *gp*GaNTases could also explain the specific response between individuals to any attack in producing mucin-type O-glycoproteins with their own structural and biological properties. Moreover, during carcinogenesis, modification of the *N*-acetylgalactosaminyltransferase repertoire, as well as the core-1 and core-2 glycosyltransferase enzymes [37], would influence all of the glycosyltransferase activities and induce aberrant density of peptide O-glycosylation. Finally, among the *N*-acetylgalactosaminyltransferase family, the *gp*GaNTases seem to be the central participants in the completion of mucin-type O-glycoproteins; further investigations should soon indicate the existence of more numerous *gp*GaNTases. Morever, the failure to identify a unique acceptor sequence for this first step of mucin-type protein O-glycosylation process is probably related to the existence of many GaNTase isoforms, each with different peptide and/or glycopeptide acceptor specificities.

We thank Professor L. A. Tabak for helpful discussion; workers in his laboratory for providing us with the recombinant GaNTases; Dr Guy Ricart USTL Villeneuve d'Ascq for help with MS; and Dominique Demeyer, Nadine Parsy and Nadege Maffioli for technical assistance.

REFERENCES

1 Clausen, H. and Benett, E. P. (1996) A family of UDP-GalNAc: polypeptide N-acetyl galactosaminyltransferases control the initiation of mucin-type O-linked glycosylation. Glycobiology **6**, 635–646

2 Nehrke, K., Hagen, F. K. and Tabak, L. A. (1998) Isoform-specific O-glycosylation by murine UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-T3, *in vivo*. Glycobiology **8**, 367–371

3 Homa, F. L., Hollander, T., Lehman, D. J., Thomsen, D. R. and Elhammer, A. P. (1993) Isolation and expression of a cDNA clone encoding a bovine UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. J. Biol. Chem. **268**, 12609–12616

4 Hagen, F. K., VanWuyckhuyse, B. and Tabak, L. A. (1993) Purification, cloning, and expression of a bovine UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. J. Biol. Chem. **268**, 18960–18965

5 White, T., Bennett, E. P., Takio, K., Sørensen, T., Bonding, N. and Clausen, H. (1995) Purification and cDNA cloning of a human UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. J. Biol. Chem. **270**, 24156–24165

6 Bennett, E. P., Hassan, H. and Clausen, H. (1996) cDNA cloning and expression of a novel human UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyl transferase, GalNAc-T3. J. Biol. Chem. **271**, 17006–17012

7 Zara, J., Hagen, F. K., Ten Hagen, K. G., VanWuyckhuyse, B. C. and Tabak, L. A. (1996) Cloning and expression of mouse UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-T3. Biochem. Biophys. Res. Commun. **228**, 38–44

8 Hagen, F. K., Ten Hagen, K. G., Beres, T. M., Balys, M. M., VanWuyckhuyse, B. C. and Tabak, L. A. (1997) cDNA cloning and expression of a novel UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase. J. Biol. Chem. **272**, 13843–13848

9 Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikematsu, Y., Merkx, G., Van Kessel, Olofsson, S. and Clausen, H. (1999) Cloning and characterization of a close homologue of human UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-T3, designated GalNAc-T6. J. Biol. Chem. **274**, 25362–25370

10 Ten Hagen, K. G., Hagen, F. K., Balys, M. M., Beres, T. M., VanWuyckhuyse, B. C. and Tabak, L. A. (1998) Cloning and expression of a novel, tissue specifically expressed member of the UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyl-transferase family. J. Biol. Chem. **273**, 27749–27754

11 Bennett, E. P., Hassan, H., Mandel, U., Mirgorodskaya, E., Roepstorff, P., Burchell, J., Taylor-Papadimitriou, J., Hollingsworth, M. A., Merkx, G., Van Kessel, A. G. et al. (1998) Cloning of a UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. J. Biol. Chem. **273**, 30472–30481

12 Ten Hagen, K. G., Tetaert, D., Hagen, F. K., Richet, C., Beres, T. M., Gagnon, J., Balys, M. M., Van Wuyckhuyse, B., Degand, P. and Tabak, L. A. (1999) Characterization of a UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase that displays glycopeptide N-acetylgalactosaminyl transferase activity. J. Biol. Chem. **274**, 27867–27874

13 Bennett, E. P., Hassan, H., Hollingsworth, M. A. and Clausen, H. (1999) A novel UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase, GalNAc-T7, with specificity for partial GalNAc-glycosylated acceptor substrates. FEBS Lett. **460**, 226–230

14 White, K. E., Lorenz, B., Evans, W. E., Meitinger, T., Strom, T. M. and Econs, M. J. (2000) Molecular cloning of a novel human UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase, GalNAc-T8, and analysis as a candidate autosomal dominant hypophosphataemic rickets (ADHR) gene. Gene **246**, 347–356

15 Toba, S., Tenno, M., Konishi, M., Mikami, T., Itoh, N. and Kurosaka, A. (2000) Brain-specific expression of a novel human UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase (GalNAc-T9). Biochim. Biophys. Acta **1493**, 264–268

16 Ten Hagen, K. G., Bedi, G. S., Tetaert, D., Kingsley, P., Hagen, F. K., Balys, M. M., Beres, T. M., VanWuyckhuyse, B., Degand, P. and Tabak, L. A. (2001) Cloning and characterization of a ninth member of the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase family, ppGaNTase-T9. J. Biol. Chem. **276**, 17395–17404

17 Hagen, F. K. and Nehrke, K. (1998) cDNA cloning and expression of a family UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase sequence homologs from *Caenorhabditis elegans*. J. Biol. Chem. **273**, 8268–8277

18 Mandel, U., Hassan, H., Therkildsen, M. H., Rygaard, J., Jakobsen, M. H., Juhl, B. R., Dabelsteen, E. and Clausen, H. (1999) Expression of polypeptide GalNAc-transferases in stratified epithelia and squamous cell carcinomas: immunohistological evaluation using monoclonal antibodies to three members of the GalNAc-transferase family. Glycobiology **9**, 43–52

19 Hagen, F. K., Hazes, B., Raffo, R., deSa, D. and Tabak, L. A. (1999) Structure–function analysis of the UDP-*N*-acetyl-D-galactosamine:polypeptide *N*-acetylgalactosaminyl-transferase. Essential residues lie in a predicted active site cleft resembling a lactose repressor fold. J. Biol. Chem. **274**, 6797–6803

20 Bennett, E. P., Weghuis, D. O., Merkx, G., Van Kessel, A. G., Eiberg, H. and Clausen, H. (1998) Genomic organization and chromosomal localization of three members of the human UDP-N-acetyl-α-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase family. Glycobiology **8**, 547–555

21 Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. and Brunak, S. (1998) NetOglyc:prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj. J. **15**, 115–130

22 Kirnarsky, L., Nomoto, M., Ikematsu, Y., Hassan, H., Bennett, E. P., Cerny, R. L., Clausen, H., Hollingsworth, M. A. and Sherman, S. (1998) Structural analysis of peptide substrates for mucin-type O-glycosylation. Biochemistry **37**, 12811–12817

23 Guyonnet-Dupérat, V., Audié, J. P., Debailleul, V., Laine, A., Buisine, M. P., Galiegue-Zouitina, S., Pigny, P., Degand, P., Aubert, J. P. and Porchet, N. (1995) Characterization of the human mucin gene MUC5AC: a consensus cysteine-rich domain for 11p15 mucin gene? Biochem. J. **305**, 211–219

24 Kingsley, P. D., Ten Hagen, K. G., Maltby, K. M., Zara, J. and Tabak, L. A. (2000) Diverse spatial expression patterns of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family member mRNAs during mouse development. Glycobiology **10**, 1317–1323

25 Soudan, B., Hennebicq, S., Tetaert, D., Boersma, A., Richet, C., Demeyer, D., Briand, G. and Degand, P. (1999) Capillary zone electrophoresis and MALDI-mass spectrometry for the monitoring of *in vitro O*-glycosylation of a threonine/serine-rich MUC5AC hexadecapeptide. J. Chromatogr. B **729**, 65–74

26 Hennebicq, S., Soudan, B., Tetaert, D., Briand, G., Richet, C., Demeyer, D. and Degand, P. (1996) Capillary electrophoresis separation of mucin motif MUC5AC glycopeptides from *in vitro* O-glycosylation reactions. Carbohydr. Res. **280**, 121–126

27 Hennebicq, S., Tetaert, D., Soudan, B., Boersma, A., Briand, G., Richet, C., Gagnon, J. and Degand, P. (1998) Influence of the amino acid sequence on the MUC5AC motif peptide O-glycosylation by human gastric UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase(s). Glycoconj. J. **15**, 275–282

28 Boulis, Y., Richet, C., Haupt, K., Hennebicq, S., Soudan, B., Tetaert, D., Degand, P. and Vijayalakshmi, M. A. (1998) Improved capillary electrophoretic separation of glycosylated oligopeptides through addition of poly(vinyl alcohol), and analysis by electrospray mass spectrometry. J. Chromatogr. A **805**, 285–293

29 Zachara, N. E. and Gooley, A. A. (2000) Identification of glycosylation sites in mucin peptides by Edman degradation. In Molecular Biology, Glycoprotein Methods and Protocols: The Mucins (Corfield, A., ed.), vol. 125, pp. 285–293, Humana Press, Totowa, NJ

30 Elhammer, A. P., Kezdy, F. J. and Kurosaka, A. (1999) The acceptor specificity of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. Glycoconj. J. **16**, 171–180

31 Yoshida, A., Suzuki, M., Ikenaga, H. and Takeuchi, M. (1997) Discovery of the shortest sequence motif for high level mucin-type O-glycosylation. J. Biol. Chem. **272**, 16884–16888

32 Asada, M., Orikasa, N., Yoneda, A., Oda, Y., Ota, Y. and Imamura, T. (1999) The AATPAP sequence is a very efficient signal for O-glycosylation in CHO cells. Glycoconj. J. **16**, 321–326

33 Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Holligsworth, M. A., Burchell, J., Taylor-Papadimitriou, J. and Clausen, H. (2000) The lectin domain of UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase T4 directs its glycopeptide specificities. J. Biol. Chem. **275**, 38197–38205

34 Colley, K. J. (1997) Golgi localization of glycosyltransferases: more questions than answers. Glycobiology **7**, 1–13

35 Hazes, B. (1996) The (Q × W)3 domain: a flexible lectin scaffold. Protein Sci. **5**, 1490–1501

36 Vinall, L. E., Hill, A. S., Pigny, P., Pratt, W. S., Toribara, N., Gum, J. R., Kim, J. R., Porchet, N., Aubert, J. P. and Swallow, D. (1998) Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5. Hum. Genet. **102**, 357–366

37 Dalziel, M., Whitehouse, C., McFarlane, I., Brockhausen, I., Gschmeissner, S., Schwientek, T., Clausen, H., Burchell, J. and Taylor-Papadimitriou, J. (2001) The relative activities of the C2GnT1 and ST3 Gall glycosyltransferases determine O-glycan structure and expression of a tumor-associated epitope on MUC1. J. Biol. Chem. **276**, 11007–11015

Received 16 February 2001/3 April 2001; accepted 26 April 2001