

Evidence for glycosylation-dependent activities of polypeptide N-acetylgalactosaminyltransferases rGalNAc-T2 and -T4 on mucin glycopeptides

Franz-Georg Hanisch^{1,2}, Celso A. Reis³, Henrik Clausen³, and Hans Paulsen⁴

²Institute of Biochemistry, Medical Faculty of the University, Joseph-Stelzmann-Str. 52, 50931 Köln, Germany, ³School of Dentistry, Department of Diagnostics, Faculty of Health Sciences, Norre Alle 20, DK 2200, Denmark, and ⁴Institute of Organic Chemistry, University of Hamburg, Martin-Luther-King-Platz 6, 20146 Hamburg, Germany

Received on February 27, 2001; revised on April 11, 2001; accepted on April 12, 2001

We present evidence that site-specific O-glycosylation by recombinant polypeptide N-acetylgalactosaminyltransferases rGalNAc-T2 and -T4 is controlled by the primary sequence context, as well as by the position and structure of previously introduced O-glycans. Synthetic mucin-type (glyco)peptides corresponding to sections of the tandem repeat regions of MUC1, MUC2, and MUC4 were used as substrates for recombinant polypeptide N-acetylgalactosaminyltransferases, rGalNAc-T2 and -T4. By concerted and sequential action the two transferases are able to fully glycosylate MUC1 but only partially MUC2 and MUC4 tandem repeat peptides. GalNAc residues on MUC1 acceptor peptides trigger activity of rGalNAc-T4 directed to Ser in VTSA and Thr in PDTR and of rGalNAc-T2 to Ser/Thr within the GSTA motif of variant MUC1 peptides. However, elongation of GalNAc by β 3-galactosylation inhibits rGalNAc-T4 activity completely and rGalNAc-T2 activity with respect to the acceptor site GSTA. These findings are in accord with the inhibition of rGalNAc-T2 and -T4 by fully GalNAc-substituted MUC1 repeat peptide and support a glycosylation-dependent activity induction or enhancement of both enzymes.

Key words: glycosyltransferases/polypeptide N-acetylgalactosaminyltransferases/O-glycosylation/glycoproteins/mucins

Introduction

A series of parameters can influence and determine the outcome of initial protein O-glycosylation: (1) the availability of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases (ppGalNAc-Ts), which show site-selectivity and are expressed in a cell-specific manner; (2) the amino acid sequence context around putative glycosylation sites that enable or inhibit catalytic action of the transferases; and (3) posttranslational modifications of the peptide substrates,

including initial GalNAc substitution and O-glycan elongation at preferred sites, which can influence the substrate qualities at proximal Ser/Thr residues.

There are currently seven recombinant GalNAc-Ts available, which can serve to simulate *in vitro* the site-specific O-glycosylation of *in vivo* processed glycoproteins (Clausen and Bennett, 1996; Ten Hagen *et al.*, 1998; Bennett *et al.*, 1998, 1999a). Although they have not yet been fully explored, each of these enzymes may have a unique site preference with some degree of overlap between the enzyme-specific target patterns (Wandall *et al.*, 1997; Hanisch *et al.*, 1999). It is expected that suitable combinations of the enzymes and appropriate incubation conditions should allow mimicking of the *in vivo* situation. The relevance of *in vitro* data for the prediction of *in vivo* O-glycosylation sites has conclusively been demonstrated for rGalNAc-T3 and a specific acceptor site on the V3 loop of the HIV gp120 protein (Nehrke *et al.*, 1998).

Calculations using neural network algorithms and sequence information from glycoprotein databases are generally based on the principal assumption that the addition of GalNAc to Ser/Thr is ruled by amino acid sequences around the potential target site. No specific motif for O-glycosylation has been identified as an acceptor site for polypeptidyl GalNAc-Ts and, in particular, for individual isoforms. On the other hand, the vicinal and proximal amino acids (positions -4 to +4), that is, the primary sequence context, play crucial roles in determining the qualities of potential substrate positions (Elhammer *et al.*, 1999).

Recently, it has been shown that the primary peptide sequence context is not the only determining parameter of site-specific O-glycosylation. Instead, there is evidence for a dynamic epigenetic regulation of initial GalNAc addition mediated by competition of different substrate sites and by competition of ppGalNAc-Ts with the core-specific glycosyltransferases (Hanisch *et al.*, 1999). The finding that multiple ppGalNAc-T isoforms in some cases (e.g., high-density O-glycosylation sequences from mucin tandem repeats) can act on the same substrate (albeit often with different kinetics and/or specific acceptor sites) introduces competition and order of GalNAc attachments. In this respect we and others previously showed that GalNAc substitution at certain substrate positions can influence the glycosylation at other vicinal or proximal sites. On the other hand, a substitution with core 1 disaccharides generally depressed proximal and in some cases even distant GalNAc addition, resulting in less dense glycosylation patterns. Furthermore, some isoforms appear to exhibit strict dependence on prior attachments of GalNAc by other isoforms (Bennett *et al.*, 1998, 1999b; Hassan *et al.*, 2000).

On the basis of these considerations the present study attempted to elucidate the influences of epigenetic parameters

¹To whom correspondence should be addressed

on initial O-glycosylation, with a focus on rGalNAc-T2 and -T4 and their action on differently glycosylated acceptor peptides of the MUC1, MUC2, and MUC4 repeat domains. The activity of rGalNAc-T4 with some substrates, including the tandem repeat of MUC1, has previously been shown to be dependent on the presence of GalNAc residues on the acceptor peptide, which may mediate conformational effects on the peptide or serve in the lectin-induced triggering of transferase activity (Bennett *et al.*, 1998; Hassan *et al.*, 2000). On the contrary, rGalNAc-T2 acts very effectively on nonglycosylated substrates, including the MUC1 repeat peptide. All currently known rGalNAc-Ts, however, in particular rGalNAc-T2, are characterized by a putative lectin domain in their C-terminal regions. Hence, the possibility exists that also other isoforms show glycosylation-dependent effects exerted by substrate-linked GalNAc on their activities.

Results

Glycosylation-induced effects on site-specific GalNAc addition to MUC1 tandem repeat peptides

Earlier studies with a panel of glycosylated MUC1 repeat peptides had revealed that the activities of ppGalNAc-Ts from milk toward potential target sites can be positively influenced by previous GalNAc-substitution at vicinal or proximal sites of the substrate (Stadie *et al.*, 1995). The same enzymes were later shown to be negatively affected by Gal-GalNAc-substitution at vicinal and even at distant sites (Hanisch *et al.*, 1999). In contrast to these enzymes, the activity of rGalNAc-T4 is absolutely dependent on previous GalNAc-substitution in the case of MUC1 tandem repeat derived peptide substrates (Bennett *et al.*, 1998; Hassan *et al.*, 2000). This glycosylation-dependent activity of rGalNAc-T4 appears to be triggered by a lectin-like interaction of the enzyme directed by the putative lectin domain found in the C-terminal region of GalNAc-transferases (Hazes, 1996; Hassan *et al.*, 2000).

The question arises whether the triggering effect of glycosylated substrate on rGalNAc-T4 activity is reversed by addition of galactose to the GalNAc-substituted peptides and whether similar effects can be observed with other enzyme isoforms like rGalNAc-T2. Using the corresponding couples with GalNAc- and Gal β 1-3GalNAc-substitution, respectively, it was revealed that rGalNAc-T2 (1) showed a consistent pattern of GalNAc incorporation into three positions, if the substrate was nonglycosylated or GalNAc substituted peptide, and (2) was inhibited with reference to distinct sites, if the substrate was Gal-GalNAc substituted (Figure 1, top). In particular, the Ser in GSTA remained unglycosylated in substrates carrying the disaccharide at Thr or Ser in the distant VTSA motif. Similar inhibition was observed in the Ser-Thr dyad of the GSTA motif, if one of these positions carried a Gal-GalNAc. rGalNAc-T4 did consistently add maximally two GalNAc residues to the monosubstituted GalNAc-peptides of the AHG21 series (Figure 1, bottom). On the other hand, the corresponding Gal β 1-3GalNAc-substituted peptides were not glycosylated by the enzyme, irrespective of the positions and numbers of the disaccharide (Figure 1, bottom). Exceptionally, the disaccharide-substituted peptide AHG21(D10) was an effective substrate and incorporated two GalNAc residues (Figure 1, bottom). This activity, similar to that on nonglyco-

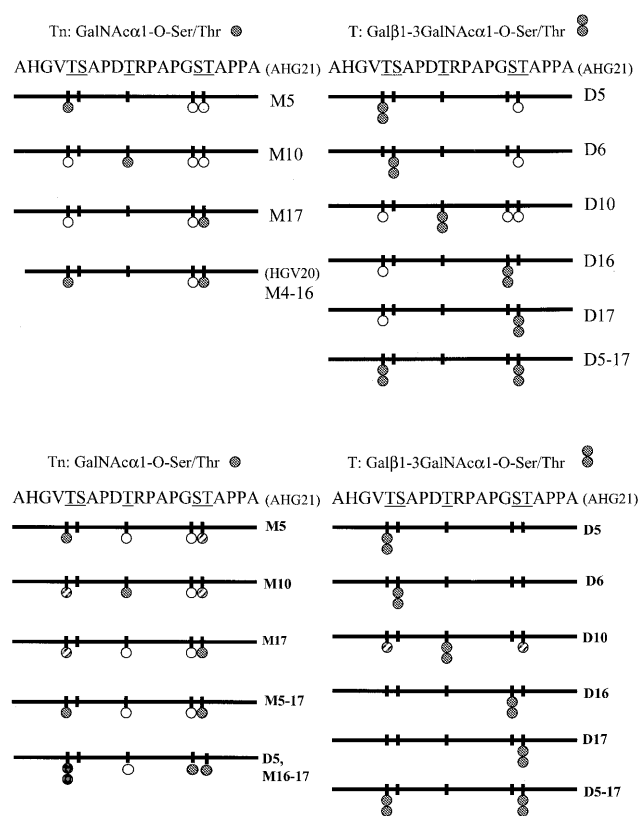


Fig. 1. Action of recombinant polypeptide GalNAc-T2 and -T4 on MUC1 glycopeptides. Top, Action of rGalNAc-T2 on MUC1 glycopeptides. Bottom, Action of rGalNAc-T4 on MUC1 glycopeptides. Glycopeptides of the AHG21 series (0.25 mM) with GalNAc or Gal β 1-3GalNAc substituted at various positions (for structures refer to Table II) were incubated with rGalNAc-T2 or -T4 at 37°C for 18 h and analyzed for the number and sites of incorporated GalNAc residues. Shaded circles indicate the positions of glycans in the substrate, hatched circles those GalNAc residues added by endogenous GalNAc-Ts from CHO cells, and open circles GalNAc residues incorporated by rGalNAc-T2 or -T4.

sylated MUC1 peptide substrates described below, could be attributed to endogenous GalNAc-Ts. Hence, it can be claimed that rGalNAc-T4 is totally inactive on all MUC1 glycopeptide substrates with Gal β 1-3GalNAc substitution. To demonstrate that the disaccharide-induced effect on rGalNAc-T4 can be compensated for by GalNAc in proximal positions, acceptor peptides based on the AHG21 sequence were generated that carried Gal-GalNAc (Thr5) and one GalNAc (Thr17) or two GalNAc residues (Ser16, Thr17). The substrate with a single GalNAc was not further glycosylated compared to the background control, whereas that with two GalNAc residues showed considerable incorporation during overnight incubation (Figure 1, bottom). This observation does not rule out the possibility of steric effects exerted by Gal-GalNAc in a negative and by GalNAc in a positive sense on accessibility of peptide sites. However, a lectin-mediated trigger effect on enzyme activity is in agreement with previous evidence based on lectin domain deletion mutants (Hassan *et al.*, 2000).

The sites of GalNAc incorporation were identified by a combination of alternative techniques; examples of their application on *in vitro* glycosylated peptides are shown in Figure 2. Enzymes

in the rGalNAc-T4 preparation were able to use each of the five positions in the MUC1 tandem repeat peptide AHG21 as a target site including the two Thr residues Thr5 and Thr17, which had previously not been identified as targets of rGalNAc-T4. However, in control experiments with rGalNAc-T4, negative enzyme preparations purified identically to rGalNAc-T4 the Thr5 and Thr17 directed activities were demonstrated to result from copurified endogenous ppGalNAc-Ts. Accordingly, the rGalNAc-T4 related activity could be identified by subtraction of the control patterns (Hassan *et al.*, 2000). The site-specific patterns observed for rGalNAc-T2 were in agreement with previous experiments (Hanisch *et al.*, 1999) and were restricted to the positions Thr5, Ser16, and Thr17 in the AHG21 sequence (Hanisch *et al.*, 1999; Figure 1, top).

Peptide variants of the MUC1 repeat unit have recently been described as major constituents of the VNTR domain in breast cancer cells (amino acid replacements Asp9–Thr10 → Glu–Ser and Pro20 → Ala in AHG21) (Müller *et al.*, 1999). The nonglycosylated variant peptide AHG21-AES was not a substrate for rGalNAc-T4 after subtraction of trace background activity (not shown). Surprisingly, the same peptide

accepted only traces of GalNAc in the presence of rGalNAc-T2, even after 48 h incubation (Figure 3). This reduced activity of rGalNAc-T2 on nonglycosylated variant peptide AHG21-AES contrasts to its high activity on the corresponding reference peptide TAP25 with Asp9–Thr10 and Pro20 (Figure 3). Strong enhancement of rGalNAc-T2 activity was observed on the GalNAc-substituted variant peptide AHG21-AES (M5), which was comparable to that on the reference glycopeptide AHG21 (M5) (Figure 3). The same glycopeptides were also efficient substrates of rGalNAc-T4. Hence, both enzymes showed a dependency of their activities on GalNAc substitution of variant repeat peptides. The variant sequence context APGSTAPA (Ala in +4 of Ser, +3 of Thr) makes the Ser and Thr positions poor acceptor sites for rGalNAc-T2, but this effect is compensated by GalNAc substitution of the substrate peptide.

Glycosylation-induced effects on initial O-glycosylation of MUC2 repeat peptide

In a previous study we could demonstrate that rGalNAc-T4 exhibits low activity toward a partial sequence of the

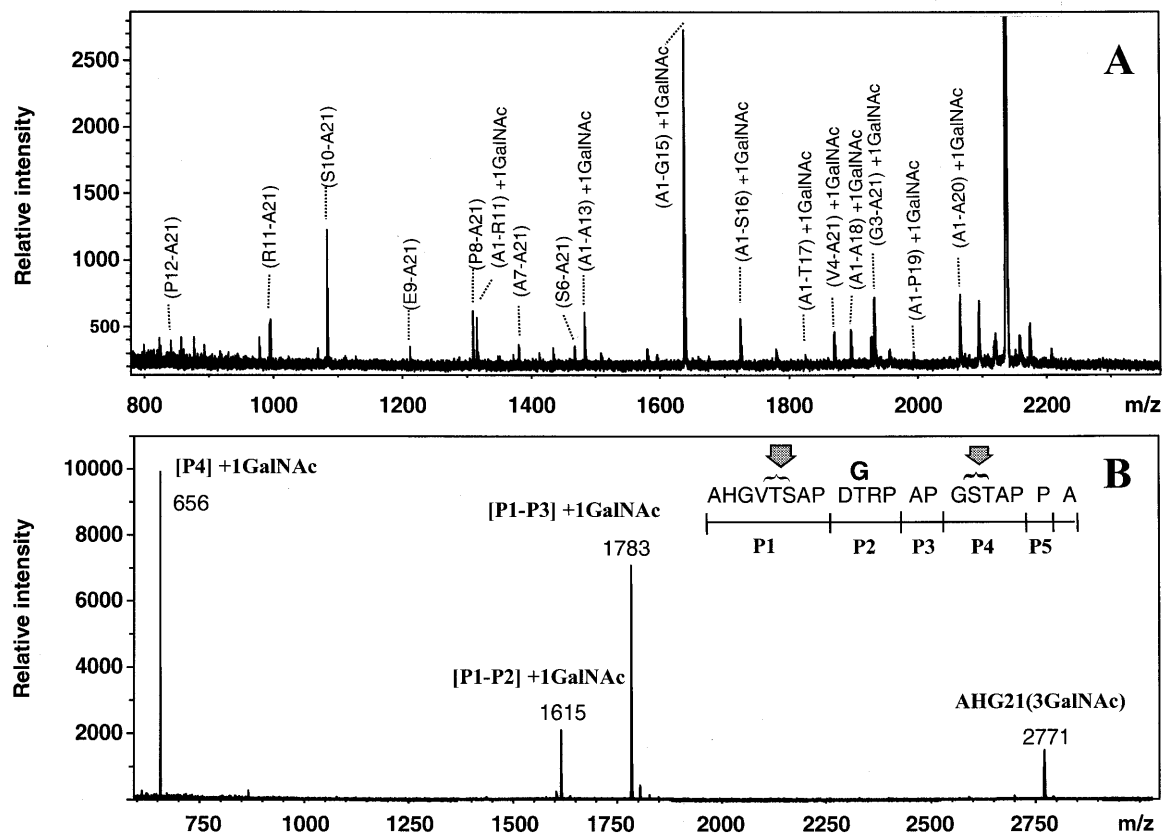


Fig. 2. Localization of glycosylated sites by partial acid hydrolysis or proteolysis of glycopeptide products. Reaction mixtures were chromatographed on HPLC reversed-phase columns, and the collected fractions were subjected to the analysis of glycosylation sites by Edman degradation or alternative complementary techniques. Examples are shown to illustrate effectivity of partial acid hydrolysis and enzymatic approaches on the basis of endopeptidase. MALDI MS of fragments obtained by partial acid hydrolysis of glycopeptide AHG21-AES(M5): AHGVT(GalNAc)SAPESRPAPGSTAPAA in pentafluoropropionic acid. All fragments containing Thr5 show a mass increment increase of 203.2, which corresponds to a GalNAc residue. Other fragments like Ser6–Ala21, Ala7–Ala21, Pro8–Ala21, Glu9–Ala21, Ser10–Ala21, Arg11–Ala21, or Pro12–Ala21 do not show mass increases compared to the calculated masses of nonglycosylated peptides. Hence, position Thr5 is unequivocally identified as glycosylation site. After partial proteolytic cleavage of AHG21(M10 + 2GalNAc) with the Pro-C-specific endopeptidase three fragments can be identified by MALDI MS. Effectivity of cleavage at internal Pro residues was different for the various sites. No fragmentation was observed at Pro8, and hence no P1 peptide could be detected in the mixture. The relative masses of the cleavage products indicate that the two additional GalNAc residues are localized at Thr or Ser of the VTSA and GSTA motifs.

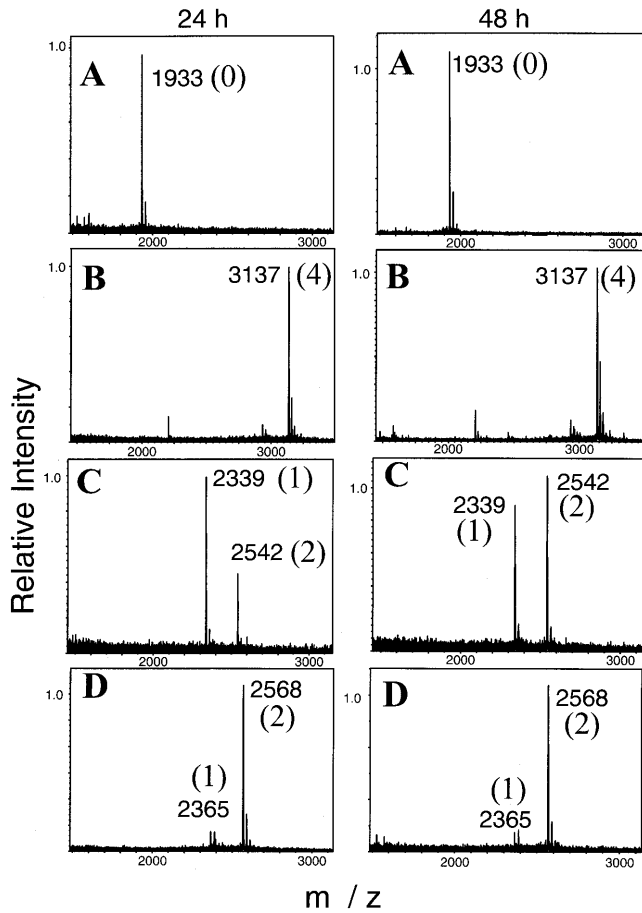
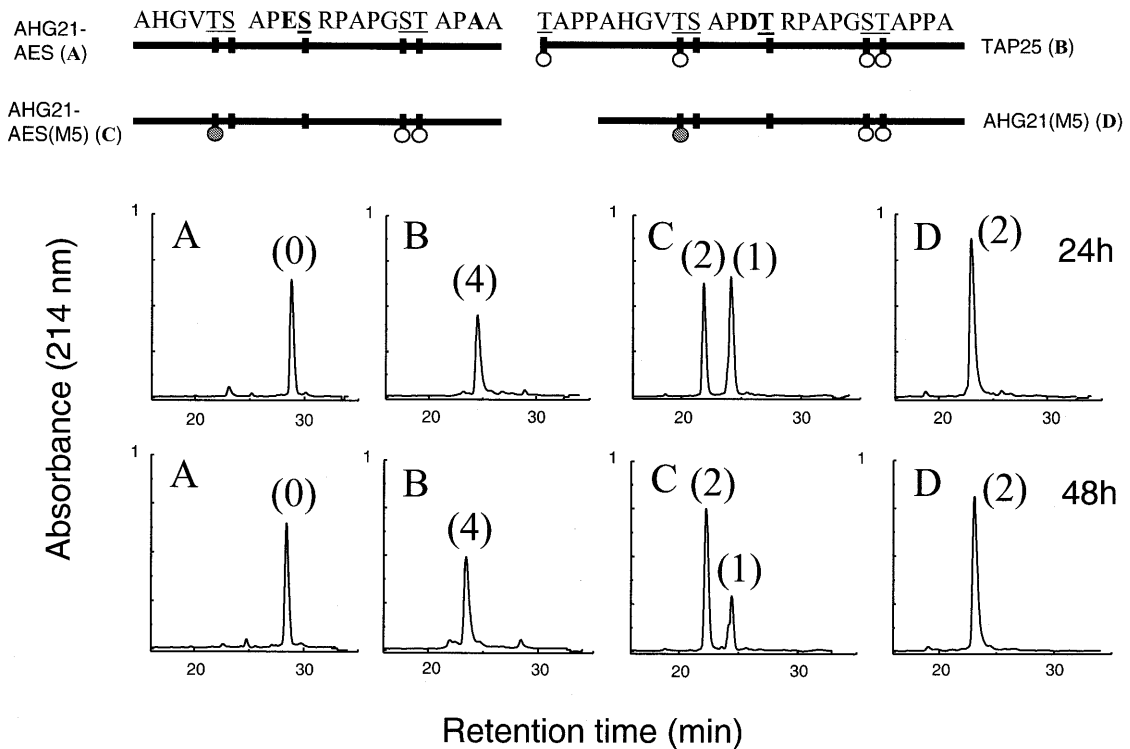


Fig. 3. Action of rGalNAc-T2 on variant MUC1 (glyco)peptides. Peptides and glycopeptides corresponding to variant sequences of the repeat domain were incubated as described under Figure 1 (for structures refer to Table II). Top, MALDI MS analysis of reaction products after 24-h or 48-h incubation times. Average masses define the major signals and the number of incorporated GalNAc residues is indicated in brackets. Substrates with variant sequences were AHG21-AES (A), and AHG21-AES(M5) (C); substrates with the reference sequence were TAP25 (B), and AHG21(M5) (D). Bottom, Analysis of reaction products by RP-HPLC after 24-h and 48-h incubation times. The products are identified by the numbers of GalNAc residues incorporated into the substrates AHG21-AES (A), TAP25 (B), AHG21-AES(M5) (C), and AHG21(M5) (D). Numbers in brackets assigned to each peak refer to mol of GalNAc incorporated per mol of (glyco)peptide. The positions of incorporated GalNAc in the respective peptide sequences are shown ahead. Shaded circles indicate the positions of glycans in the substrate, and open circles GalNAc residues incorporated by rGalNAc-T2.



nonglycosylated MUC2 repeat peptide, which is fivefold increased after substitution of the substrate with one or two GalNAc residues (Bennett *et al.*, 1999b). The peptide substrate PTT15 (PTTTPITTTTTVTPT) used in this study corresponds also to a partial sequence of the 23 amino acid MUC2 repeat but exhibits 2 amino acid substitutions (Ser7 → Thr, Met11 → Thr) compared to the previous substrate and 10 potential glycosylation positions. A maximum of seven GalNAc residues were actually incorporated by rGalNAc-T2 (Figure 4). The sites of glycosylation preferred by rGalNAc-T2 could be partially defined by sequencing in PSD-MALDI MS combined with partial acid hydrolysis in the gas phase, and were confirmed by Edman degradation. The site pattern revealed with PTT15 as substrate was in agreement with results obtained previously for the partial sequence PTTTPLK (Iida *et al.*, 1999). Nonglycosylated PTT15 was not a substrate for rGalNAc-T4 in agreement with its strict dependency on previous GalNAc substitution of MUC1 repeat peptide, contrasting with the previous findings (Bennett *et al.*, 1999b). Using PTT15(GalNAc)₄ as a substrate for rGalNAc-T4 the enzyme incorporated one further residue into the glycopeptide (site not identified).

The position in substrate EPT9 that corresponds to Thr2 in PTT15 remained an effective target site for rGalNAc-T2, if the vicinal Thr3 was substituted with the disaccharide GlcNAcβ1-6GalNAc (core 6). Strikingly, rGalNAc-T4 was unable to add GalNAc to any of the glycosylated substrates, neither to the GlcNAcβ1-6GalNAc nor to GalNAc substituted EPT9 (Figure 4).

Glycosylation-induced effects on initial O-glycosylation of MUC4 repeat peptides

The peptide LPV19 corresponds to one unit and three overlapping amino acid residues of the MUC4 repeat (Figure 5). Although the peptide contains eight potential O-glycosylation sites, the number of GalNAc residues incorporated by rGalNAc-T2 is only five. Using LPV19(GalNAc)₄ as a substrate for rGalNAc-T4 the enzyme incorporated one further GalNAc residue into the glycopeptide (site not identified). rGalNAc-T4 activity did not exceed background controls on

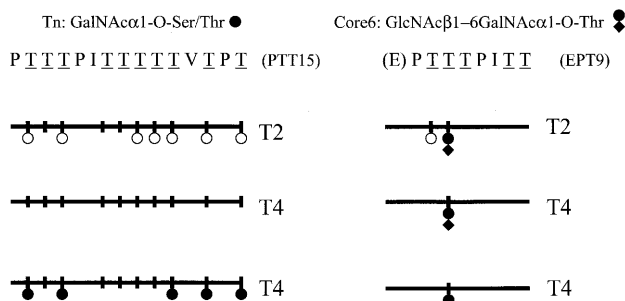


Fig. 4. Action of rGalNAc-Ts on MUC2 (glyco)peptides. Synthetic peptide PTT15 corresponding to a partial sequence of the 23 amino acid tandem repeat peptide of MUC2 and the glycopeptides EPT9(d4) and EPT9(M4) (for structures refer to Table II) were incubated with the transferases rGalNAc-T2 and -T4 and analyzed by alternative methods for the sites of GalNAc incorporation. Black circles indicate glycosylated positions of the substrate, open circles those sites where GalNAc residues were added.

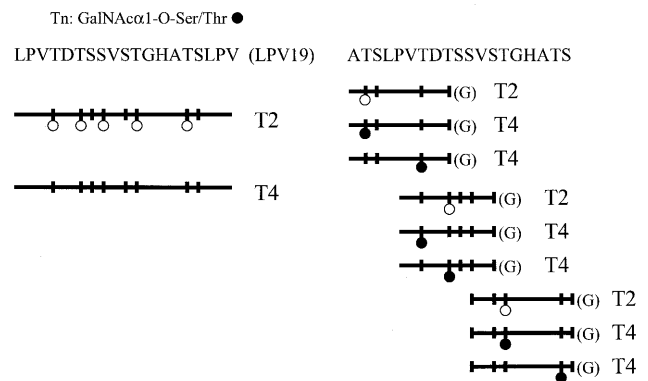


Fig. 5. Action of rGalNAc-Ts on MUC4 (glyco)peptides. The synthetic peptide LPV19 corresponding to the MUC4 repeat peptide and three overlapping residues, and partial sequences of the repeat peptide carrying GalNAc at various positions (for structures refer to Table II) were incubated with the indicated rGalNAc-Ts. Black circles indicate glycosylated positions of the substrate, open circles those sites where GalNAc residues were added.

nonglycosylated LPV19 and on all nonglycosylated peptides corresponding to partial sequences of the MUC4 repeat (SVS10, ATS10, PVT10). Unexpectedly, also the GalNAc-substituted decapeptides were not substrates of rGalNAc-T4 (Figure 5).

The sites of GalNAc addition to LPV19 were identified by Edman sequencing (rGalNAc-T2: Thr4, Thr6, Ser8, Thr11, Thr15). Results obtained with the decapeptide substrates corresponding to overlapping partial sequences of MUC4 repeat peptide did agree with these site-specific patterns. rGalNAc-T2, was, however, unable to use any of the GalNAc-substituted decapeptides as substrate (Figure 5). This finding indicates that a minimum peptide length may be required, which is larger for glycosylated substrates.

Inhibition of rGalNAc-T2 and rGalNAc-T4 action

According to previous evidence the glycosylation-dependent action of rGalNAc-T4 appears to be mediated by a lectin-like GalNAc interaction of the enzyme with the substrate glycopeptide (Hassan *et al.*, 2000). Such interaction has been substantiated by inhibitory effects of GalNAc, but not of other monosaccharides, on rGalNAc-T4 activity (Hassan *et al.*, 2000). We intended to further evaluate the nature of these effects and to demonstrate that also other isoforms of the ppGalNAc-Ts show similar inhibition of their activities. To this end we used free GalNAc and GalNAc-peptide (fully substituted at each potential site) as inhibitors of the enzyme(s). Incorporation rates in the presence and absence of inhibitor were measured for rGalNAc-T2 and -T4 and compared to those for rGalNAc-T1 as a reference (Figure 6 and Table I). Incubation of substrates, nonglycosylated TAP25 (rGalNAc-T1, and -T2 in Figure 6, bottom) and GalNAc substituted AHG21(M10) (rGalNAc-T1 and -T4 in Figure 6, top), over a period of 18 h resulted in reproducible proportions of glycosylated sites versus total potential acceptor sites on the substrate (set as 100%). In the presence of glycopeptide inhibitor HGV20(M4-5-9-15-16) the rGalNAc-T4 catalyzed addition of GalNAc was reduced to 17% of the control (Table I). On a molar basis the glycopeptide inhibitor was more effective by at least two orders of magnitude

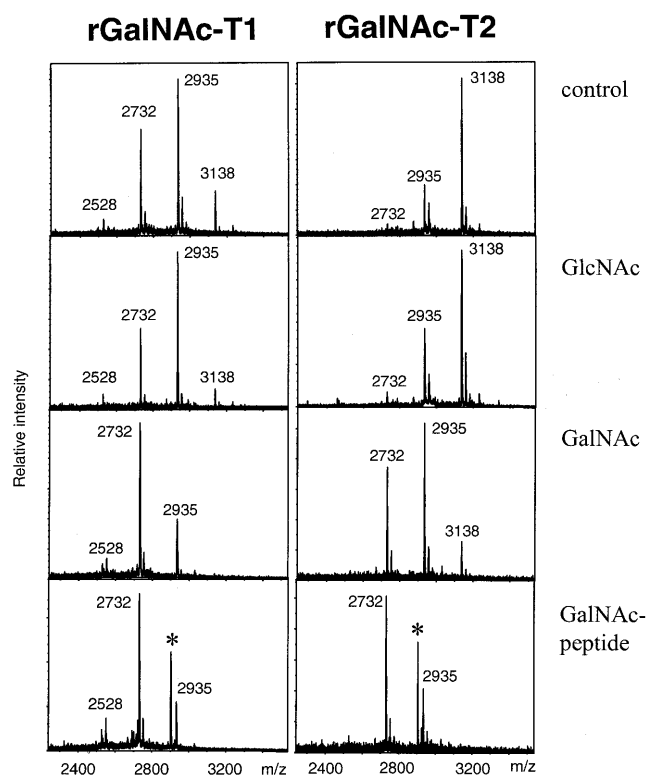
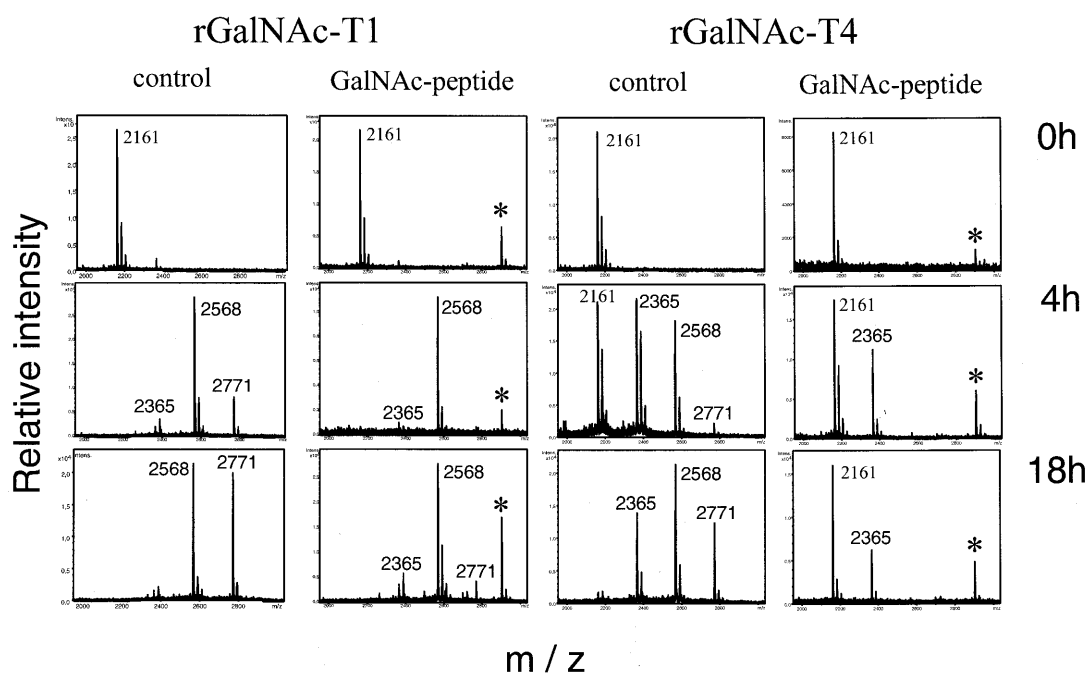


Fig. 6. Inhibition of rGalNAc-Ts by sugars and glycopeptide. Top, time-course MALDI MS analysis of GalNAc transfer by rGalNAc-T1 or rGalNAc-T4 in the absence or presence of GalNAc-peptide. The enzymes were incubated with substrate AHG21(M10), average mass of MH 2161, for the indicated time intervals in the absence (control) or presence of 0.35 mM inhibitor HGV20(M4-5-9-15-16) (GalNAc-peptide). The ion masses at *m/z* 2365, 2568, and 2771 indicate the addition of one, two, or three GalNAc residues, respectively. Asterisk indicates the signal corresponding to the GalNAc-peptide inhibitor with an average mass of MH 2903. Bottom, comparative MALDI MS analysis of GalNAc transfer by rGalNAc-T1 and rGalNAc-T2 to TAP25 in the presence and absence of inhibitors. The enzymes were incubated for 18 h in the absence (control) or presence of sugars (GalNAc, GlcNAc) or glycopeptide HGV20(M4-5-9-15-16) (GalNAc-peptide) and the reaction mixtures were analysed by MALDI MS for the number of incorporated GalNAc residues. The ion masses at *m/z* 2528, 2731, 2935, and 3138 indicate nonglycosylated TAP25 or its reaction products with one, two, or three GalNAc residues, respectively. Asterisk indicates the signal corresponding to the GalNAc-peptide inhibitor with an average mass of MH 2903.

compared to the monosaccharide GalNAc (Table I). No effect was measurable with GlcNAc or with the nonglycosylated MUC1 tandem repeat peptide TAP25.

A similar set of experiments was performed with the enzymes rGalNAc-T1 and -T2 (Figure 6, bottom). Activities of

both enzymes showed no decrease in the presence of GlcNAc, but were considerably inhibited by GalNAc and glycopeptide HGV20(M4-5-9-15-16). Comparing the glycopeptide induced inhibition of the three enzymes it is obvious, however, that the effect on rGalNAc-T4 activity is much stronger (Figure 6, top).

Table I. Inhibition of polypeptide GalNAc-transferase activities in rGalNAc-T preparations by glycopeptide and sugars

Inhibitor	% substitution of theoretical acceptor sites (percentage of glycosylation relative to controls)		
	rGalNAc-Ts (substrates)		
	rGalNAc-T1 (TAP25)	rGalNAc-T2 (TAP25)	rGalNAc-T4 (AHG21-M10)
Control	45.6 (100)	61.5 (100)	45.2 (100)
GlcNAc (50 mM)	44.5 (97.5)	60.0 (97.5)	41.0 (90.7)
GalNAc (50 mM)	36.8 (80.7)	46.5 (75.6)	23.6 (47.8)
TAP25 (0.4 mM)	n.d.	n.d.	38.3 (84.7)
GalNAc-peptide ^a (0.34 mM)	35.5 (77.8)	37.7 (61.3)	7.7 (17.0)

The assays were performed in duplicate and the reaction mixtures were analysed by MALDI MS after 18 h incubation. Signal intensities for each glycopeptide were averaged and the degree of substitution (%) at theoretical acceptor sites in each substrate was calculated. The percentage of glycosylation relative to controls in the absence of inhibitors is given in parentheses.

^aThe GalNAc-peptide used as inhibitor of rGalNAc-Ts is the fully glycosylated MUC1 repeat peptide HGV20 (M4-5-9-15-16).

Discussion

The present study has revealed further evidence for the modulating effects exerted by substrate glycosylation on following GalNAc addition to proximal Ser/Thr positions. Using recombinant GalNAc-transferases, rGalNAc-T2 and -T4, this evidence was obtained by *in vitro* O-glycosylation of peptide and glycopeptide substrates corresponding to sections of the MUC1, MUC2, and MUC4 tandem repeat domains. Both enzymes have previously been characterized with regard to their site specificities in the MUC1 repeat peptide sequence context (Wandall *et al.*, 1997; Hanisch *et al.*, 1999; Hassan *et al.*, 2000). The findings of this study suggest that beyond the sequence context the two enzymes show site-specific enhancement of their activities on GalNAc-substituted acceptor peptides. The type of sugar linked to the acceptor peptide was revealed to play a crucial role in glycosylation-induced enzyme activation, because a substitution of the substrate with Gal β 1-3GalNAc at varying positions generally resulted in completely or partially suppressed GalNAc addition to adjacent or distant positions. In case of rGalNAc-T2 this site-specific inhibition was restricted to Ser16 within the GSTA motif. rGalNAc-T4 showed a strikingly different activity on the same substrates by (1) lacking any activity on nonglycosylated repeat peptides of MUC1 and (2) transferring GalNAc to target sites, which are not accessible to rGalNAc-T2, if at least one GalNAc residue was available on the substrate peptide. This is in agreement with previous observations and was explained by a lectin-mediated trigger effect on the enzyme activity (Hassan *et al.*, 2000). This assumption is supported by the findings of this study, which revealed complete suppression of rGalNAc-T4 activity on substrates carrying Gal-GalNAc instead of GalNAc only. Blocking of the lectin ligand by galactosylation is one explanation of rGalNAc-T4 inactivity. However, the assumption of steric or conformational effects exerted by the core 1 disaccharide on enzyme access to the peptide substrate cannot be ruled out. Inhibition experiments with sugars and glycopeptides demonstrated that rGalNAc-T4 activity on its preferred substrates can be specifically blocked by soluble and peptide-linked GalNAc but not by GlcNAc. However, blocking of the catalytic domain, which should have binding affinity to peptide-linked

GalNAc in the enzyme-product complex, cannot be ruled out. Strikingly, also rGalNAc-T1 and T2 showed inhibition by GalNAc and peptide-linked GalNAc. In Figure 6 (top) it is obvious that the addition of the third GalNAc residue by rGalNAc-T1 (to Ser16 in the AHG21 repeat peptide of MUC1) is strongly inhibited by the GalNAc peptide, whereas addition of the second residue is not. It cannot be excluded, accordingly, that other isoforms of rGalNAc-Ts exhibit some glycosylation dependency with respect to certain sites of a peptide substrate. Both enzymes, rGalNAc-T1 and -T2, are not strictly dependent on GalNAc substitution of their acceptor peptides, but at least for rGalNAc-T2 it has been shown in this study that enzyme activity is considerably enhanced on glycosylated variant repeats. Also in case of rGalNAc-T4 the glycosylation of substrates is not an absolute prerequisite for its activity, because there is at least one example, where the enzyme exhibits non-glycosylation-dependent activity, e.g. the PSGL-1 peptide substrate (Bennett *et al.*, 1998). In conclusion, it can be speculated that the glycosylation dependent action of rGalNAc-T2 and rGalNAc-T4 on specific sites in the MUC1 glycopeptide substrate is ruled by the sequence context and is positively or negatively regulated by GalNAc or Gal-GalNAc substitution, respectively.

This study has revealed further evidence for the previously postulated epigenetic regulation of glycosylation density by competition between initial GalNAc introduction and core-specific enzymes (Hanisch *et al.*, 1999). The dependency of rGalNAc-T4 on previously introduced GalNAc residues leads to a situation where the fill-up reactions of partially glycosylated MUC1 peptides can be inhibited by β 3-galactosylation of GalNAc-peptides catalyzed by the core 1-specific C1GT1. Assuming colocalization of both enzymes their relative activities in the cis-Golgi compartments may finally determine to which degree the MUC1 repeat peptide is actually glycosylated at each of the five putative sites.

In the instance of MUC1 the different glycoforms found in human milk and on carcinoma cells were demonstrated to show an inverse relationship between glycosylation density and the lengths and complexity of the glycans (Hanisch *et al.*, 1989; Müller *et al.*, 1997, 1999; Hanisch and Müller, 2000). Though 50% of the putative sites in MUC1 are substituted with polyactosamine-type chains during lactation (Müller *et al.*,

1997), nearly full glycosylation of the repeat peptides with core-type chains is reached in the breast cancer cell line T47D (Müller *et al.*, 1999). Similar results were obtained with recombinantly expressed MUC1 glycosylation probes after transfection into a variety of established breast cancer cell lines (Müller and Hanisch, unpublished data).

Not only expression of a repertoire of glycosyltransferases but also their localization in the Golgi, their relative activities in each subcompartment, and their spatial arrangements in the Golgi membrane will influence the site-specific O-glycosylation state of a mucin. The impact of the last on antigenicity and immunogenicity of MUC1, which represents a primary target in the worldwide development of cancer vaccines, has been recently reviewed (Hanisch and Müller, 2000).

Materials and methods

Peptides and glycopeptides

The synthetic peptides and glycopeptides used in this study were designated throughout by using the N-terminal tripeptide motif followed by the number of amino acid residues (Table II). Glycosylation is indicated in brackets by M (GalNAc substitution), D (Gal β 1-3GalNAc substitution), or d (GlcNAc β 1-6GalNAc substitution) followed by the substituted position within the peptide sequence. Sequence variants of the MUC1 repeat peptide were labeled by the replaced amino acids written in italics.

MUC1 TAP25 peptide was kindly supplied by Dr. Joyce Taylor-Papadimitriou (ICRF, London). The peptide variants exhibiting amino acid replacements, AHG21(AES), APE21(EST), and GST20(AES), were synthesized in-house by a local service facility. All glycopeptides based on the AHG21 peptide (Table II), the HGV20-based glycopeptide, and the glycosylated peptide variants AHG21(AES) were synthesized and analyzed according to previously published protocols (Klich *et al.*, 1997; Mathieux *et al.*, 1997). For structural details refer to Table II. AHG21(D5-M17) and AHG21(D5-M16-17) were derived by *in vitro* glycosylation of synthetic AHG21(D5) using rGalNAc-T2 and purified from the reaction mixtures by reversed-phase high-performance liquid chromatography (RP-HPLC).

MUC2 The glycopeptide EPT9 is structurally based on a section of the MUC2 tandem repeat peptide carrying an additional Glu residue at the N-terminus and a GlcNAc β 1-6GalNAc (core 6) substitution at Thr4 EPT9(d4). Its synthesis has been described in a previous publication (Mathieux *et al.*, 1997). The peptide is modified N-terminally by acetylation and C-terminally by amidylation. The corresponding GalNAc substituted EPT9(M4) peptide was generated by enzymatic cleavage of the GlcNAc residue using β -hexosaminidase (jack bean, Glyco, Novato). The nonglycosylated peptide PTT15 corresponding to a section of the MUC2 repeat was synthesized by a local service facility.

MUC4 The peptide LPV19 covering one tandem repeat unit of MUC4 was synthesized in-house by a local service facility. The glycopeptides corresponding to sections of the MUC4 tandem repeat peptide, SVS10, ATS10, and PVT10, were synthesized as previously described (Mathieux *et al.*, 1997). These glycopeptides terminate with an additional Gly residue

Table II. Synthetic mucin (glyco)peptide substrates used in this study

Designation	MUC1 TR (glyco)peptides
TAP25	TAPPAHGVTSPADTRPAPGSTAPPA
AHG21(M5)	AHGVTSPADTRPAPGSTAPPA
AHG21(M10)	AHGVTSPADTRPAPGSTAPPA
AHG21(M17)	AHGVTSPADTRPAPGSTAPPA
HGV20(M4-16)	HGVTSPADTRPAPGSTAPPA
AHG21(D5)	AHGVT <u>S</u> APDTRPAPGSTAPPA
AHG21(D6)	AHGVT <u>S</u> APDTRPAPGSTAPPA
AHG21(D10)	AHGVTSPAD <u>T</u> RAPGSTAPPA
AHG21(D16)	AHGVTSPADTRPAPG <u>S</u> TAPPA
AHG21(D17)	AHGVTSPADTRPAPGS <u>T</u> APPA
AHG21(D5-17)	AHGVT <u>S</u> APDTRPAPGS <u>T</u> APPA
	MUC1 TR variant (glyco)peptides
AHG21-AES	AHGVTSAPE <u>S</u> RPAPGSTAPAA
AHG21-AES(M5)	AHGVTSAPE <u>S</u> RPAPGSTAPAA
AHG21-AES(M10)	AHGVTSAPE <u>S</u> RPAPGSTAPAA
AHG21-AES(M17)	AHGVTSAPE <u>S</u> RPAPGSTAPPA
AHG21-ES(M5)	AHGVTSAPE <u>S</u> RPAPGSTAPPA
AHG21-ES(M10)	AHGVTSAPE <u>S</u> RPAPGSTAPPA
AHG21-ES(M17)	AHGVTSAPE <u>S</u> RPAPGSTAPPA
AHG21-ESS(D10)	AHGVTSAPE <u>S</u> RPAPGSSAPPA
AHG21-ES(D17)	AHGVTSAPE <u>S</u> RPAPGS <u>T</u> APPA
AHG21-S(D17)	AHGVTSPADTRPAPGS <u>S</u> APPA
GST20-AES	GSTAPAAHGVTSAPE <u>S</u> RPAP
APE21-EST	APESRPAPGSTAPTAHGVTSA
	MUC2 TR peptides
PTT15(GalNAc ₄)	PTTTPTTTT [*] TVTPT [*]
PTT15	PTTTPTTTT [*] TVTPT
EPT9(d4)	EPT <u>T</u> TPITT
EPT9(M4)	EPTTTTPTT
	MUC4 TR peptides
LPV 19	LPVTDTSVSTGHATSLPV
SVS10	SVSTGHATS(G)
SVS10(M4)	SVSTGHATS(G)
SVS19(M8)	SVSTGHATS(G)
ATS10	ATSLPVTD(T)(G)
ATS10(M2)	ATSLPVTD(T)(G)
ATS10(M7)	ATSLPVTD(T)(G)
PVT10	PVTDTSVVS(G)
PVT10(M3)	PVTDTSVVS(G)
PVT10(M5)	PVTDTSVVS(G)

The first three letters refer to the starting sequence of the peptides, followed by the number of amino acid residues. Amino acid replacements are given in italics. Glycosylation site and structures are designated with M (GalNAc) or D (Gal-GalNAc) or d (GlcNAc-GalNAc). Glycosylated Thr/Ser is shown in boldface (GalNAc) or underlined boldface (Gal-GalNAc).

^{*}Labeled peptide substrate, which was generated by *in vitro* glycosylation without site localization of incorporated GalNAc residues.

and are modified by acetylation (N-terminal) and amidylation (C-terminal).

In vitro glycosylation

The enzyme sources used were semipurified as previously described by successive sequential ion-exchange chromatographies on Amberlite (IRA95, Sigma) or DEAE Sephacel (Pharmacia), S-Sepharose Fast-Flow (Pharmacia), and Mini-STM (PC3.2/3, Pharmacia) using the Smart system (Pharmacia) (Bennett *et al.*, 1998). Secreted rGalNAc-T4 was obtained from stably transfected Chinese hamster ovary (CHO) line (CHO/GalNAc-T4/21A) (Hassan *et al.*, 2000).

Peptide or glycopeptide substrates (100–500 μ M) were solubilized in 25 mM cacodylate buffer, pH 7.4, containing 10 mM MnCl₂ and 0.25% TritonX-100 and mixed with UDP-GalNAc (200 μ M) and rGalNAc-Ts in a total volume of 20 μ l. In case of rGalNAc-T1 and -T2 (obtained from pAc-GP67-GalNAc-T1-sol and pAc-GP67-GalNAc-T2-sol baculovirus-infected High Five cells) enzyme activities were used, which yielded complete incorporation of GalNAc within 18h into the sites Thr9, Ser20, and Thr21 of TAP25. For all other substrates the reaction mixture was incubated at 37°C for 18 h to reach a state where even poor substrate positions were at least partially glycosylated (Hassan *et al.*, 2000). Insect cells as well as CHO cells produce and secrete endogenous GalNAc-transferases. These endogenous activities are merely directed to Thr5 and Thr17 in the MUC1 repeat peptide starting with the AHG motif. Enzymatic background controls included a mutant rGalNAc-T4 (459D/H) with a defect lectin domain (Hassan *et al.*, 2000), a truncated catalytically inactive enzyme (T2C-sol), and a preparation of β 4-galactosyltransferase β 4-Gal-T2 (Almeida *et al.*, 1999). All enzymes were purified from High Five media as reported previously (Bennett *et al.*, 1998).

Analysis of reaction products

In initial experiments the reaction times and enzyme amounts were optimized for each rGalNAc-T preparation to achieve a maximum of incorporated GalNAc residues into the various peptide and glycopeptide substrates. Generally amounts of enzyme were used, which allowed glycosylation of all acceptor sites accessible for a particular isoform of GalNAc-Ts. In most cases reaction times of 18 h were sufficient to achieve exhaustive glycosylation in terms of incorporation of the maximal number of GalNAc residues. A prolongation of the reaction times up to 48 h did not increase this number. Only these “endpoint” products were used for site identification experiments after HPLC purification.

Aliquots of the reaction mixtures were diluted 1:10 in 0.1% aqueous trifluoroacetic acid (TFA) and mixed with an equal volume of α -cyano-4-hydroxy-cinnamic acid matrix (saturated solution in 0.1% TFA/acetonitrile, 1:2, v/v) on the target. The samples were analyzed by matrix-assisted laser desorption and ionization mass spectrometry (MALDI MS) on a Bruker Reflex III using nitrogen laser ionization ($\lambda = 337$ nm), and positive ion detection in the reflectron mode as previously described (Müller *et al.*, 1997). Pseudomolecular ions MH⁺ were registered and the number of GalNAc residues added to the substrate peptides was calculated on the basis of incremental mass increase (1 HexNAc residue corresponding to 203.2 mass units). The remaining reaction mixtures were chromatographed by RP-HPLC on a semipreparative scale and

the eluting products were quantitated on the basis of their UV spectroscopic profiles as described previously (Hanisch *et al.*, 1999; Müller *et al.*, 1997). Aliquots were also run on a capillary electrophoresis apparatus (MR5000, Beckman Coulter, Munich) using uncoated 25 cm fused silica capillaries (75 μ m) and 50 mM phosphate, pH 2.5, as electrophoresis buffer with normal polarity. Detection was at 200 nm and the thermostat for the cooling fluid was set at 25°C.

Localization of O-glycosylation sites

To identify the actual acceptor sites utilized by a particular enzyme, we chose reaction times which allowed at least the partial conversion into a final product with the maximum number of GalNAc residues. Prolonging the reaction times exceeding 18 h did not increase the number of substituted sites, and hence all structural studies were performed on those species exhibiting a final state of O-glycosylation.

O-glycosylated Ser/Thr residues were identified by a variety of complementary methods. HPLC-purified glycopeptides were sequenced by Edman degradation on an ABI 473A (PE Biosystems, Foster City, CA) with microcartridge under standardized conditions (Zachara and Gooley, 1998). Alternatively, partial sequences were obtained by analysis of postsourc-decay fragments in MALDI MS (Müller *et al.*, 1997). In cases where MS fragmentation yielded insufficient sequence information, the glycosylation sites of HPLC-pure glycopeptides (10 μ g) were localized by partial acid hydrolysis in a gas-phase reaction using pentafluoropropionic acid (20% v/v in water containing 0.5% dithiothreitol), which was incubated with the dry sample in a Mininert (Pierce) under vacuum at 90°C for 1 h (Mirgorodskaya *et al.*, 1999). The fragments were taken up in 0.1% aqueous TFA and analyzed by MALDI MS as above. Partial information on glycosylation sites was also obtained by enzymatic fragmentation of the glycopeptides with the endopeptidases from *Flavobacterium meningosepticum* (Pro-C specific, 100 mU in 0.1 M phosphate buffer, pH 7.2, 2 h at 30°C) or from *Papaya* (Gly-C specific, 100 mU, 0.1 M phosphate buffer, 1 mM EDTA, pH 6.8, 18 h at 37°C). Both enzymes were purchased from ICN Biomedicals (Eschwege, Germany).

Acknowledgments

This study was supported by the grants Ha 2092/4-4 from the Deutsche Forschungsgemeinschaft and by the Danish Research Council and Cancer Society.

Abbreviations

CHO, Chinese hamster ovary; MALDI, matrix-assisted laser desorption and ionization; MS, mass spectrometry; ppGalNAc-T, polypeptidyl N-acetylgalactosaminyltransferases; RP-HPLC, reversed-phase high-performance liquid chromatography; TFA, trifluoroacetic acid.

References

- Almeida, R., Levery, S.B., Mandel, U., Kresse, H., Schwientek, T., Bennett, E.P., Clausen, H. (1999) Cloning and expression of a proteoglycan UDP-galactose: β -xylose β 1-galactosyltransferase I. A seventh member of the human β 4-galactosyltransferase gene family. *J. Biol. Chem.*, **274**, 26165–26171.

- Bennett, E.P., Hassan, H., Mandel, U., Hollingsworth, M.A., Akisawa, N., Ikematsu, Y., Merckx, G., van Kessel, A.G., Olofsson, S., and Clausen, H. (1999a) Cloning and characterization of a close homologue of human UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase T3, designated GalNAc-T6. Evidence for genetic but not functional redundancy. *J. Biol. Chem.*, **274**, 25362–25370.
- Bennett, E.P., Hassan, H., Hollingsworth, M.A., and Clausen, H. (1999b) A novel human UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase, GalNAc-T7, with specificity for partial GalNAc-glycosylated acceptor substrates. *FEBS Lett.*, **460**, 226–230.
- Bennett, E.P., Hassan, H., Mandel, U., Mirgorodskaya, E., Roepstorff, P., Burchell, J., Taylor-Papadimitriou, J., Hollingsworth, M.A., Merckx, G., Van Kessel, A.G., and others. (1998) Cloning of a human UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. *J. Biol. Chem.*, **273**, 30472–30481.
- Clausen, H. and Bennett, E.P. (1996) A family of UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases control the initiation of mucin-type O-linked glycosylation. *Glycobiology*, **6**, 635–646.
- Elhammer, A.P., Kezdy, F.J., and Kurosaka, A. (1999) The acceptor specificity of UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases. *Glycoconj. J.*, **16**, 171–180.
- Hanisch, F.-G. and Müller, S. (2000) MUC1: the polymorphic appearance of a human mucin. *Glycobiology*, **10**, 439–449.
- Hanisch, F.-G., Müller, S., Hassan, H., Clausen, H., Zachara, N., Gooley, A.A., Paulsen, H., Alving, K., and Peter-Katalinic, J. (1999) Dynamic epigenetic regulation of initial O-glycosylation by UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases. *J. Biol. Chem.*, **274**, 994–9954.
- Hanisch, F.-G., Uhlenbruck, G., Peter-Katalinic, J., Egge, H., Dabrowski, J., and Dabrowski, U. (1989) Structures of neutral O-linked polygalactosaminoglycans on human skim milk mucins. *J. Biol. Chem.*, **264**, 872–883.
- Hassan, H., Reis, C., Bennett, E.P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M.A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2000) The lectin domain of UDP-GalNAc: N-acetylgalactosaminyltransferase T4 directs its glycopeptide specificities. *J. Biol. Chem.*, **275**, 38197–38205.
- Hazes, B. (1996) The (QxW)₃ domain: a flexible lectin scaffold. *Protein Sci.*, **5**, 1490–1501.
- Iida, S., Takeuchi, H., Hassan, H., Clausen, H., and Irimura, T. (1999) Incorporation of N-acetylgalactosamine into consecutive threonine residues in MUC2 tandem repeat recombinant human N-acetylgalactosaminyltransferase-T1, -T2, and -T3. *FEBS Lett.*, **449**, 230–234.
- Klich, G., Paulsen, H., Meyer, B., Meldal, M., and Bock, K. (1997) Synthesis and characterization of highly glycosylated glycopeptides with Tn-antigenic structures corresponding to human glycoporphin AN. *Carbohydr. Res.*, **299**, 33–48.
- Mathieux, N., Paulsen, H., Meldal, M., and Bock, K. (1997) Synthesis of glycopeptide sequences of repeating units of the mucins MUC2 and MUC3 containing oligosaccharide side-chains with core 1, core 2, core 3, core 4 and core 6 structure. *J. Chem. Perkin Trans.*, **1**, 2359–2368.
- Mirgorodskaya, E., Hassan, H., Wandall, H.H., Clausen, H., and Roepstorff, P. (1999) Partial vapor-phase hydrolysis of peptide bonds: A method for mass spectrometric determination of O-glycosylated sites in glycopeptides. *Anal. Biochem.*, **269**, 54–65.
- Müller, S., Alving, K., Peter-Katalinic, J., Zachara, N., Gooley, A.A., and Hanisch, F.-G. (1999) High density O-glycosylation on tandem repeat peptide from secretory MUC1 of T47D breast cancer cells. *J. Biol. Chem.*, **274**, 18165–18172.
- Müller, S., Goletz, S., Packer, N., Gooley, A., Lawson, A.M., and Hanisch, F.-G. (1997) Localization of O-glycosylation sites on glycopeptide fragments from lactation-associated MUC1. *J. Biol. Chem.*, **272**, 24780–24793.
- Nehrke, K., Hagen, F.K., and Tabak, L.A. (1998) Isoform-specific O-glycosylation by murine UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase-T3, *in vivo*. *Glycobiology*, **8**, 367–371.
- Stadie, T., Chai, W., Lawson, A.M., Byfield, P., and Hanisch, F.-G. (1995) Studies on the order and site specificity of GalNAc transfer to MUC1 tandem repeats by UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase from milk or mammary carcinoma cells. *Eur. J. Biochem.*, **229**, 140–147.
- Ten Hagen, K.G., Hagen, F.K., Balys, M.M., Beres, T.M., Van Wuyckhuysse, B., and Tabak, L.A. (1998) Cloning and expression of a novel, tissue-specifically expressed member of the UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase family. *J. Biol. Chem.*, **273**, 27749–27754.
- Wandall, H.H., Hassan, M., Mirgorodskaya, E., Kristensen, A.K., Roepstorff, P., Bennett, E.P., Nielsen, P.A., Hollingsworth, M.A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) Substrate specificities of three members of the human UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. *J. Biol. Chem.*, **272**, 23503–23514.
- Zachara, N.E. and Gooley, A.A. (1998) Identification of glycosylation sites in mucin peptides. In T. Corfield, ed., *Mucin methods and protocols. Methods in molecular biology*. Humana Press, Totowa.