The Lectin Domain of UDP-N-acetyl-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase-T4 Directs Its Glycopeptide Specificities*

Received for publication, June 30, 2000, and in revised form, September 5, 2000
Published, JBC Papers in Press, September 12, 2000, DOI 10.1074/jbc.M005783200

Helle Hassan§§, Celso A. Reis§§§, Eric Paul Bennett §§§, Ekaterina Mirgorodskaya¶, Peter Roepstorff, Michael A. Hollingsworth**, Joy Burchell†¶, Joyce Taylor-Papadimitriou‖, and Henrik Clausen§§§

From the §Faculty of Health Sciences, School of Dentistry, Copenhagen DK-2200, Denmark, the ¶Institute of Molecular Pathology and Immunology of University of Porto, Porto 4200, Portugal, the †Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense University, Odense DK-5230, Denmark, the **Eppley Institute for Research in Cancer and Allied Diseases, University of Nebraska Medical Center, Omaha, Nebraska 68198, and the ‖Imperial Cancer Research Fund, London WC2A 3Ps, United Kingdom

The initiation step of mucin-type O-glycosylation is controlled by a large family of homologous UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases (GalNAc-transferases). Differences in kinetic properties, substrate specificities, and expression patterns of these isoenzymes provide for differential regulation of O-glycan attachment sites and density. Recently, it has emerged that some GalNAc-transferase isoforms in vitro selectively function with partially GalNAc O-glycosylated acceptor peptides rather than with the corresponding unglycosylated peptides. O-Glycan attachment to selected sites, most notably two sites in the MUC1 tandem repeat, is entirely dependent on the glycosylation-dependent function of GalNAc-T4. Here we present data that a putative lectin domain found in the C terminus of GalNAc-T4 functions as a GalNAc lectin and confers its glycopeptide specificity. A single amino acid substitution in the lectin domain of a secreted form of GalNAc-T4 selectively blocked GalNAc-glycopeptide activity, while the general activity to peptides exerted by this enzyme was unaffected. Furthermore, the GalNAc-glycopeptide activity of wild-type secreted GalNAc-T4 was selectively inhibited by free GalNAc, while the activity with peptides was unaffected.

The first step in mucin-type O-glycosylation is catalyzed by one or more members of a large family of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases (GalNAc-transferases). Differences in kinetic properties, substrate specificities, and expression patterns of these isoenzymes provide for differential regulation of O-glycan attachment sites and density. Recently, it has emerged that some GalNAc-transferase isoforms in vitro selectively function with partially GalNAc O-glycosylated acceptor peptides rather than with the corresponding unglycosylated peptides. O-Glycan attachment to selected sites, most notably two sites in the MUC1 tandem repeat, is entirely dependent on the glycosylation-dependent function of GalNAc-T4. Here we present data that a putative lectin domain found in the C terminus of GalNAc-T4 functions as a GalNAc lectin and confers its glycopeptide specificity. A single amino acid substitution in the lectin domain of a secreted form of GalNAc-T4 selectively blocked GalNAc-glycopeptide activity, while the general activity to peptides exerted by this enzyme was unaffected. Furthermore, the GalNAc-glycopeptide activity of wild-type secreted GalNAc-T4 was selectively inhibited by free GalNAc, while the activity with peptides was unaffected.

The first step in mucin-type O-glycosylation is catalyzed by one or more members of a large family of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases (GalNAc-transferases)§ (EC 2.4.1.41), which transfer GalNAc to serine and threonine acceptor sites (for reviews, see Ref. 1–3). To date seven members of the mammalian GalNAc-transferase family have been identified and characterized (4–12), and several additional putative members of this gene family have been predicted from analysis of genome data bases (13). The GalNAc-transferase isoforms have different kinetic properties and show differential expression patterns temporally and spatially, suggesting that they have distinct biological functions (2). Sequence analysis of GalNAc-transferases has led to the hypothesis that these enzymes contain two distinct subunits: a central catalytic unit, and a C-terminal unit with sequence similarity to the plant lectin ricin (14–17). Previous experiments involving site-specific mutagenesis of selected conserved residues confirmed that mutations in the catalytic domain eliminated catalytic activity. In contrast, mutations in the lectin domain had no or only little effects on catalytic activity of at least one GalNAc-transferase isoform, GalNAc-T1 (14). However, recent evidence demonstrates that some GalNAc-transferases in vitro exhibit unique activities with partially GalNAc-glycosylated glycopeptides. The catalytic activities of two GalNAc-transferase isoforms, GalNAc-T4 and -T7, selectively act on glycopeptides corresponding to mucin tandem repeat domains where only some of the clustered potential glycosylation sites have been GalNAc glycosylated by other GalNAc-transferases (7, 11, 12). Importantly, GalNAc-T4 and -T7 recognize different GalNAc-glycosylated peptides and catalyze transfer of GalNAc to acceptor substrate sites in addition to those that were previously utilized.

GalNAc-T4 is unique in that it is the only GalNAc-transferase isoform identified so far that in vitro can complete the O-glycan attachment to all of five potential acceptor sites in the tandem repeat sequence (20 amino acids: HGVTSTAPDTR-PAPGSTAPPA, potential O-glycosylation sites underlined) of the human cell membrane mucin, MUC1 (18). GalNAc-T4 was previously shown to transfer GalNAc to two sites (S in -VTS- and T in -PDTR-) not used by other GalNAc-transferase isoforms on the GalNAc4TAP24 glycopeptide (T* APPAHGVTSTAPDTRPAPGSTAPP*, GalNAc attachment sites marked by asterisks) (7). The in vitro kinetic properties of GalNAc-T4 with these glycopeptide substrates appear relatively poor mainly due to low apparent Vₘₐₓ as the apparent Kₘ is low (90 μM with GalNAc4TAP24). Nevertheless, it is clear that an activity such as that exhibited by GalNAc-T4 is relevant time-of-flight; PFPA, pentafluoropropanoic acid; CHO, Chinese hamster ovary.
7.4), 10 mM MnCl2, and 0.25% Triton X-100 with 2–5 M excess of UDP-

Five insect cells grown in serum-free medium. The mutant GalNAc-T-

ghout refers to this MUC1 peptide design) and TAP24 (TAP24S

The mutant GalNAc-T4 is expressed in T47D cells as evaluated by immunocytochemistry with a monoclonal antibody (7, but this isoform is not generally expressed in normal breast tissues or carcinomas. MUC1 is considered a cancer-associated antigen because expression is highly up-regulated in cancers of many tissues including breast and pancreas and the glycosylation is aberrant with short, unbranched O-glycans (21). It has been debated whether O-glycan density of the tandem repeat region is reduced or increased in cancer, but judging from the findings of MUC1 in the cancer cell lines T47D (20) and Colo205 (22), it appears that the latter is most likely.

This study addresses the mechanism by which GalNAc-T4 exerts its GalNAc-glycopeptide substrate specificity in vitro. Initial studies of the substrate specificity of GalNAc-T4 with different glycoforms of the MUC1 peptide indicated that the glycopeptide specificity was independent of the sites of attachment of GalNAc in the peptide sequence. This observation is not in agreement with the findings that most GalNAc-trans-

fersases exhibit rather distinct acceptor substrate specificities governed by the sequence contexts of the peptide substrates (2). Thus, we investigated the hypothesis that a previously identified putative lectin domain found in the C terminus of most GalNAc-transfersases contributed to their function in catalyzing glycosylation of glycopeptides. We evaluated the function of the GalNAc-T4 isoform that displays enzyme activity, which, in addition to showing activity with some peptide substrates, exhibits unique activity with glycopeptides where prior glycosylation is a prerequisite for activity (7, 12). The results clearly demonstrate that the lectin domain of GalNAc-T4 selectively directs the glycopeptide specificity in vitro.

**EXPERIMENTAL PROCEDURES**

**Polypeptide GalNAc-transferase Assays—**Acceptor peptides included five variants of TAP25 (TAPPAGHVTYTSAPDTRAPGSVTSV2APPEM) and TAP24 (TAPPAGHVTYSAPDTRAPGSSTAPP) derived from the human MUC1 tandem repeat (18). The Val-substituted peptides except TAP25V20 were biotinylated. GalNAc glycosylated glycoforms of MUC1-derived peptides were produced using purified recombinant GalNAc-T4 and GalNAc-T2, purified and characterized as described previously. The peptide PSGL1β (Ac-QATEEYLYDLEFPEPPREM) derived from the selectin-ligand-glycoprotein-1 (PSGL-1) (23) was used to monitor GalNAc-T4 activity with unsubstituted peptides. GalNAc-glycopeptides were prepared in reaction mixtures containing 25 mm cocadylate (pH 7.4), 0.25% Triton X-100 with 2–5 M excess of UDP-GalNAc, and 0.4 μg of GalNAc-T4. The amount of GalNAc-transfersases added was determined so that the reaction with the appropriate peptide was near completion in 6 h. Reactions were incubated in the sample carousel of an Applied Biosys-

-tem model HT270 at 30 °C (24). Electrophoretograms were produced every 60 min, and after 6 h the reaction mixtures were separated by reverse-phase HPLC for structural determination. HPLC was performed on a Brawnee ODS column (2.1 mm × 30 mm, 5-μm particle size) (Applied Biosystems, Inc.) equipped with an Agilent 1100 microbore HPLC system (PerkinElmer Life Sciences).

**Reaction Kinetics Monitored by MALDI-TOF—**Reactions were performed in mixtures of 25 μl containing 2.5 nmol of acceptor (glyco)peptide, 40 nmol of UDP-GalNAc, and 0.4 μg of GalNAc-T4. The amount of GalNAc-transfersases added was determined so that the reaction with the GalNAc-TAP25V21 glycopeptide substrate was near completion in 16 h. Sampling of reactions (1 μl) was purified by nanoscale reversed-phase chromatography (Poros R3, PerSeptive Biosystem) and applied directly to the probe with matrix (26). Reactions were incubated at 37 °C in a shaker bath. At 0, 2, and 16 h, a 1-μl aliquot was taken and purified. Mass spectra were acquired on Voyager-DE mass spectrometer equipped with delayed extraction (PerSeptive Biosystem). The matrix used was 2.5-dihydroxybenzoic acid (10 mg/ml, Aldrich) dissolved in a 2.1:1 mixture of 0.1% trifluoroacetic acid in 30% aqueous acetonitrile (Rathburn Ltd.).

**RESULTS**

**GalNAc-T4 Transfer to at Least Three Sites in the MUC1 Tandem Repeat Sequence—**GalNAc-T4 transfers two GalNAc residues to the GalNAc-T4APDG substrate (TAPAPAHGVTYSAPDTRAPGSVTSV2APPEM) at Ser18 in -VTS2A- and Thr14 in -PDT- (7). The kinetics of the reaction of GalNAc-T4 with GalNAc-T4APDG, GalNAc-T4APDG when used in the standard reaction (18 cycles of 95 °C for 45 s; 51 °C for 5 s; 72 °C for 1 min). The PCR product was digested with BstYI, gel-purified, and 5 ng hereof mixed with 10 ng of pAcGP67-GalNAc-T4-sol. The mixture was used to prime a "shuffle

---

2 U. Mandel, H. Hassan, and H. Clausen, unpublished observation.
FIG. 1. CE analysis of in vitro O-glycosylation of the MUC1-derived peptide, TAP24/25, and variants with Val substitutions (T9V and T21V) using purified recombinant secreted human GalNAc-T4 from CHO cells. Refer to "Results" for detailed description. Numbers above peaks refer to numbers of moles of GalNAc incorporated into the peptide as evaluated by MALDI-TOF analysis. The identity of the glycoform produced was in most cases resolved by PFPA hydrolysis combined with mass spectrometry as described under "Experimental Procedures." The substrates and enzymes used in each experiment is indicated above the electropherogram, and the substrate and the product formed in the reactions is depicted in the inserted figure (lines indicate potential O-glycosylation site not occupied, open circles indicate GalNAc residues attached prior to the reaction, gray circles indicate GalNAc residues attached by GalNAc-T1 or T2, and solid circles GalNAc residues transferred by GalNAc-T4.

Lectin Domain of GalNAc-T4 Directs Glycopeptide Specificity
porated from the intermediate time points showed incorporation into both sites, indicating that GalNAc-T4 acts independently on Ser10 in -VT S A- and Thr 14 in -PD T R-. With the substrate GalNAc2TAP25, the product with 1 mol of incorporation did not accumulate, but was converted quickly to a product with 2 mol of GalNAc/mol of peptide (Fig. 1, panel B). Addition of GalNAc-T4 and UDP-GalNAc to long term reactions with GalNAc2TAP25 resulted in appearance of glycoforms with 5 and 6 mol of GalNAc/mol of peptide, and structural analysis confirmed that GalNAc was incorporated at Ser20 in -G S TA- efficiently, in addition to the two sites previously identified (7), provided the substrate has GalNAc residues at Thr9 and/or Thr21. The kinetics of the reaction is such that it is not possible to identify which of the two sites, Thr14 or Ser20, is first glyco-

---

**FIG. 2.** Representative example of the structural analysis of GalNAc-glycosylated MUC1 peptides. Panel I, MALDI-RE-TOF mass spectra of the reverse phase HPLC-purified glycoform of peptide GalNAc2TAP25 glycosylated with GalNAc-T4 (top) (reaction in Fig. 1, panel A). Observed monoisotopic masses are given for (MH)+. Panel II, MALDI-RE-TOF mass spectra of Asp-N digests of the glycosylated TAP25 peptides from panel I. Proteolytic fragments are labeled according to the TAP25 peptide sequence: 1–12 correspond to TAPPAHGVTSAP, and 13–25 to DTIRPAPGSTAPPA. Panel III, MALDI-RE-TOF mass spectra of the Asp-N fragments of glycosylated peptide GalNAc2TAP25 glycosylated with GalNAc-T4 after hydrolysis with 20% PFPA at 90 °C for 1 h. a, Asp-N fragment (1–12)+1GalNAc; b, Asp-N fragment (13–25)+3GalNAc. The identified degradation products are labeled in the spectra according to TAP25 peptide sequence. Asterisk (*) indicates loss of the acetyl group (-42 Da) from the GalNAc residue, observed for all glycosylated hydrolytic fragments. The diagrams under the spectra show observed degradation products for each Asp-N fragment. Asp-N fragment (1–12) has three potential glycosylation sites (underlined in the peptide sequence), i.e. Thr1, Thr9, and Ser10. The glycosylated peptide fragments are detected as non-glycosylated (NG) and glycosylated (+1GalNAc) product ions due to partial cleavage of glycosidic bond during hydrolysis. The presence of a GalNAc residue on the peptide (1–9) unambiguously identifies Thr9 as the glycosylated residue. Asp-N fragment (13–25) has three potential glycosylation sites (underlined in the peptide sequence), i.e. Thr14, Ser20, and Thr21, and based on the observed mass was assumed to carry three GalNAc residue. The observed hydrolytic fragments of the peptide confirmed that all potential sites are occupied. †, calculated monoisotopic molecular masses based on the amino acid sequence. ††, measured monoisotopic masses of all peptide fragments after subtraction of the ionizing proton.
GalNAc2TAP25 (Fig. 1, panel C) is therefore likely to represent GalNAc-T4 activity with the unsubstituted TAP25 peptide (Fig. 1, panel C). Standard initial velocity assays performed in 30–60 min showed no activity over background levels in wild-type CHO cells, but CE analysis showed very low incorporation after 6 h. This may be due to the presence of very low endogenous GalNAc-transferase activity from the medium of the CHO-GalNAc-T4 cells, in which GalNAc-T4 is co-purified. The small peaks at 3- and 4-mol glycoforms observed with GalNAc-T4 in Fig. 1 (panel C) is therefore likely to represent GalNAc-T4 products formed from minor amounts of substrates produced by endogenous co-purified enzymes. Thus, a low endogenous activity will be amplified by GalNAc-T4 activity and hence give activity levels higher than standard controls (7, 12). The enzymes were purified to apparent homogeneity (7, 24), but our purification method does not include affinity steps that would exclude other GalNAc-transferases. As discussed below, the endogenous GalNAc-transferase activity found in media of infected insect cells is higher.

The action of GalNAc-T4 in mixing experiments with GalNAc-T1 and -T2 gave results similar to experiments in which GalNAc-T4 was added to peptide glycoforms produced by GalNAc-T1 and -T2. GalNAc-T1 transfers GalNAc residues to TAP24 in the following order: Thr9, Thr21, and Ser20, but the last residue is slowly glycosylated (Fig. 1, panel D) (24). GalNAc-T2 transfers GalNAcs to TAP24 in the following order: Thr21, Thr9, Ser20, and Thr1, and transfer to Ser20 is more efficient than for GalNAc-T1 (Fig. 1, panel E) (24). Mixing GalNAc-T1 and -T4 (Fig. 1, panel F) initially produced a 2 mol of GalNAc/mol of peptide glycoform that was the major product of GalNAc-T1. This product was slowly converted to a 4 mol of GalNAc/mol of peptide glycoform with similar kinetics to reactions in which GalNAc-T4 catalyzed glycosylation of GalNAc2TAP25 (Fig. 1, panel B). Structural analysis showed that the occupied sites were the same as in Fig. 1 (panel B); Ser10 in -VTSA- was not glycosylated. The kinetics of GalNAc-T1 activity with the third site in TAP24 (Fig. 1, panel D) suggest that GalNAc-T4 catalyzed incorporation of GalNAc into Thr14 and Ser20. Mixing of GalNAc-T2 and -T4 initially produced 2 mol of GalNAc/mol of peptide glycoform similar to the major product seen in reactions with GalNAc-T2 alone, which was then slowly converted to a glycoform with 6 mol of GalNAc incorporated/mol of peptide (Fig. 1, panel G). The order in which the last four residues were incorporated was not determined, but by inference from the previous experiments it is likely that GalNAc-T4 reacted with the 2-mol glycoform.

The MUC1 GalNAc-glycopeptide Specificity of GalNAc-T4 Is Not Dependent on a Specific Glycoform—Enzymatic preparation of TAP24 with 1 mol of GalNAc attached in either Thr9 in -VTSA- or Thr21 in -GSTA- is difficult to prepare as the reactions with GalNAc-T1, -T2, and -T3 proceed to 2 mol of incorporated GalNAc without significant accumulation of the intermediate product with 1 mol of GalNAc incorporated (24). To analyze the importance of the Thr acceptor sites at positions 9 and 21 of the TAP25 repeat peptide, Thr/Val-substituted peptides were analyzed. These peptides include a biotin group in the N terminus, which affected the activities of GalNAc-T1 and -T2. The kinetics of GalNAc-T1 with TAP24 and TAP25V21 (Thr21 is the second site utilized by GalNAc-T1) is shown in Fig. 1 (panels D and H). Surprisingly, the Val21 substitution produced an improved substrate for GalNAc-T1, and 3 mol of GalNAc were efficiently incorporated. Structural analysis revealed that the third mole of GalNAc was incorporated at the N-terminal Thr1, which GalNAc-T1 does not utilize in native TAP24. The kinetics of GalNAc-T2 with TAP24 and TAP25V9 (Thr9 is the second site utilized by GalNAc-T2) is shown in Fig. 1 (panels E and I). The Val9 substitution had adverse effects and resulted in inhibition of incorporation into Ser9 in -GSTA- on the side of the repeat. A similar finding was observed using a synthetic glycopeptide with the core 1 disaccharide (Galβ1–3GalNAcO-Thr) at Thr9 in -VTSA-, for GalNAc-T1,-T2, as well as -T3 (28). As shown in Fig. 1 (panels D and K), neither substitution created substrates that showed activity with GalNAc-T4.

Mixing experiments of GalNAc-T1 and -T4 with TAP25V21 resulted in total incorporation of 5 mol of GalNAc at all available sites (Fig. 1, panel L). Since GalNAc-T1 transfers GalNAc to Thr1, Thr9, and Ser20 in isolated experiments (Fig. 1, panel H), it is concluded that GalNAc-T4 transfers to Ser10 in -VTSA- and Thr14 in -PDTR-. This result showed that GalNAc-T4 shows activity regardless of a GalNAc residue at Thr21. Further, a GalNAc residue at Thr1 in the truncated peptide is required for efficient glycosylation of Ser10 by GalNAc-T4 (compare Fig. 1, panels B and L). Mixing experiments with GalNAc-T2, -T4, and TAP25V9 resulted in incorporation of 4 mol of
GalNAc at the times shown here (Fig. 1, panel M). Structural analysis showed that these were incorporated at Thr1, Thr14, Ser20, and Thr21, but not Ser10. Prolonged incubation with additional enzyme resulted in incorporation into Ser10 (data not shown).

A Val/Val21-substituted peptide, GalNAc-TAP25V9V21, served as a substrate for GalNAc-T1 and -T2, whereas this peptide, as expected, did not serve as a substrate for GalNAc-T4 (data not shown). GalNAc-T2 incorporated 2 mol of GalNAc at Thr1 and Ser20, and this product served as a substrate for GalNAc-T4 (Fig. 1, panel N). GalNAc-T4 transferred GalNAc at Thr1 and Ser20, and this product served as a substrate for GalNAc-T4 (data not shown). GalNAc-T2 incorporated 2 mol of GalNAc at Thr1 and Ser20, and this product served as a substrate for GalNAc-T4 (Fig. 1, panel N). GalNAc-T4 transferred GalNAc at Thr1 and Ser20, and this product served as a substrate for GalNAc-T4 (data not shown).

In summary, analysis of the substrate specificity of GalNAc-T4 with different glycoforms of MUC1 revealed that GalNAc-T4 did not show a requirement for any single site of GalNAc attachment; however, there was a requirement for at least one of the three sites (Thr9, Ser20, and Thr21) done by other GalNAc-transferase isoforms (e.g. GalNAc-T1, -T2, and -T3) to be glycosylated. Thus, substitution of any one of the sites glycosylated in the GalNAc-TAP24/25 glycopeptide by valine did not affect activation of GalNAc-T4 activity for glycopeptides. Catalytic activity with certain sites was affected by site-specific modifications, in particular glycosylation of Ser10. This led us to hypothesize that a triggering event that was independent of the general catalytic activity of the enzyme led to acquisition of specificity for GalNAc-glycopeptides. A likely candidate for the triggering event of glycopeptide activity was the putative lectin domain, which was previously shown by mutational analysis to not significantly affect the activity of GalNAc-T1 with a peptide substrate (14).
The Lectin Domain of GalNAc-T4 Selectively Directs the GalNAc-glycopeptide Specificity but Not the Peptide Specificity—Since GalNAc-T4 exhibits both glycosylation-independent and glycosylation-dependent activities, it offered a model system to analyze the different specificities as separate functions. Hagen et al. (14) originally demonstrated that critical substitutions in the lectin domain of GalNAc-T1 have little effect on catalytic activity (reduction by 10–50%) with peptide substrates, while substitutions in the catalytic domain destroyed activity (Fig. 4, panel A). It was predicted that mutation of an aspartate residue adjacent to a conserved CLD motif in the lectin domain to histidine (D444H in GalNAc-T1 corresponding to D459H in GalNAc-T4) would destroy putative lectin function based on analysis of ricin (29), but mutation of this residue (D444H) in GalNAc-T1 only appeared to reduce activity by approximately 50%. To test if the lectin domain influenced glycopeptide specificity of GalNAc-T4, we prepared recombinant secreted forms of GalNAc-T4459D and -T4459H. These were expressed in High Five insect cells and purified by non-affinity chromatography to apparent homogeneity. Controls included medium from cells infected with an irrelevant viral construct and purified in parallel. Analysis of activities of the purified enzyme preparations with the unglycosylated MUC1 peptides revealed considerable background activity (Fig. 5). The long assay time used for evaluation of product development showed that both GalNAc-T4459D and -T4459H preparations initially (2 h) produced less product than the control; however, once the 1–2-mol GalNAc-glycoform was produced, only the wild-type GalNAc-T4459D converted the substrate to the fully glycosylated GalNAc5TAP25V21 glycopeptide. The background values will vary somewhat with the viral infection efficiency, and this explains the apparent higher endogenous activity in the control. GalNAc-T4459D and -T4459H exhibited essentially the same specific activity with several other unglycosylated peptides where no glycosylation-dependent activities are found. This is illustrated in Fig. 4 (panel B), with the PSGL-1 substrate, which is an unique substrate for GalNAc-T4, and for which insect cells have no endogenous GalNAc-transferase activity (background). The finding that the general catalytic activity of the lectin mutated form is intact is in agreement with the results obtained for GalNAc-T1 (14). In contrast, the glycopeptide specificity of mutant GalNAc-T4459H was selectively affected by the introduced mutation. Glycopeptides derived from tandem repeats of MUC1, MUC2, and MUC5AC (12) were virtually inactive as substrates, as is illustrated in Fig. 4 (panel C), which depicts assays with a GalNAc5TAP25V21 glycopeptide. Essentially identical results were observed with the GalNAc5TAP24 glycopeptide. These results show that the lectin domain is required for the glycopeptide specificity of enzyme activity, but not for activity with naked peptide substrates. This supports the hypothesis that the lectin domain triggers the catalytic domain of GalNAc-T4 to act on GalNAc-glycopeptide substrates by an as yet unknown mechanism. We concluded further that the basic catalytic function and the triggering event are independent properties associated with distinct domains of GalNAc-T4.

The Lectin Domain of GalNAc-T4 Recognize GalNAc—In order to determine if actual carbohydrate recognition contributed to the function of the lectin domain, we analyzed whether triggering of glycopeptide specificity could be blocked by specific carbohydrates in solution. We could not detect direct binding of GalNAc-T2 and -T4 to GalNAc or GalNAc-peptides using conventional binding assays. However, as shown in Fig. 6 (panel A), the glycosylation-dependent specificity of GalNAc-T4 was almost completely inhibited by incubation with 0.23 mM free GalNAc, whereas other sugars, Gal, GlcNAc, or Fuc, failed to show significant inhibition. Assays with 50 mM sugars gave the same pattern, but with less (approximately 50%) inhibition by GalNAc (data not shown). Furthermore, similar inhibition was found with 10 mM α-D-GalNAc-1-benzyl, whereas αGlcNAc-benzyl did not inhibit catalytic activity (data not shown). None of the sugars had significant affects on the glycosylation-independent activities of GalNAc-T4459D or -T4459H, when assayed with naked peptides (Fig. 6, panel B). This provides strong evidence in support of the hypothesis that the lectin domain of GalNAc-T4 binds to GalNAc and contributes to the ability of GalNAc-T4 to catalyze glycosylation of glycopeptides.

**DISCUSSION**

This study demonstrated that previous predictions of a putative lectin domain in the C terminus of most GalNAc-transferases were correct at least in the case of the GalNAc-T4 isoform (15, 16). The lectin domain of GalNAc-T4 confers its unique GalNAc-glycopeptide specificities as evaluated using in
vitro assays by recognition at least in part of GalNAc residues attached to the glycopeptide substrate. At least one additional GalNAc-transferase isoform, GalNAc-T7, exhibits such glycosylation-dependent substrate specificity, but the acceptor substrate specificities of GalNAc-T4 and -T7 with glycopeptides are different (11, 12). The finding that some GalNAc-transferase isoforms selectively or exclusively utilize partially GalNAc-glycosylated substrates in vitro suggests that in vivo functions of these isoforms may be in acting after initiation of O-glycosylation by other enzymes in a “follow-up” role. Thus, we envisage that initiation of O-glycosylation of mucin sequences with high density follow initiation at specific sites as a result of the fine acceptor substrate specificity of the initial acting GalNAc-transferases, and that the products formed serve as substrate for subsequent GalNAc-transferase isoforms which recognize the partially GalNAc-glycosylated substrate. This putative model defines the initiation step of O-glycosylation as a series of ordered actions of GalNAc-transferase isoforms. The data presented here on the unique function of GalNAc-T4 in glycosylation of MUC1 suggest that the glycopeptide specificity of this enzyme is indifferent to the particular GalNAc-MUC1 glycoform, while still showing a restricted acceptor substrate specificity with regard to peptide sequence context of the acceptor substrates/sites. The catalytic unit of GalNAc-transferases is predicted to be functionally distinct from the lectin domain. The lectin-mediated functions are not predicted to govern general unspecific activation since the two GalNAc-T4 and -T7 isoforms exhibit different glycopeptide substrate specificities. For example, GalNAc-T4 transfers GalNAc to the EA2 peptide substrate derived from rat submaxillary mucin, but it is blocked by addition of just 1 mol of GalNAc (12). In contrast, GalNAc-T7 is selectively activated by GalNAc glycosylation of the EA2 peptide substrate (12). Thus, the catalytic unit has distinct acceptor substrate specificity based primarily on peptide sequence context, while the lectin domain, at least for GalNAc-T4, can expand the substrate repertoire of the catalytic unit to include additional specific acceptor sites flanked by GalNAc residues.

Most homologous members of the GalNAc-transferase family, which currently includes eight plus isoforms, appear to have a conserved lectin domain (2, 14). Variation in amino acid sequence among isoforms in this domain is high as compared with the catalytic domain. Preliminary analysis of GalNAc-T7, the only other isoform that has been characterized for glycosylation-dependent activity (11, 12), confirms that the glycopeptide activity can be inhibited by GalNAc.

Functions of lectin domains of other GalNAc-transferase isoforms not showing glycosylation-dependent activities are still unknown, but the finding that most have essential residues conserved suggests that many are functional. A large GalNAc-transferase family with nine distinct genes exists in Caenorhabditis elegans, and interestingly one homologue lacks the lectin domain and instead has the C-terminal HDEL ER retrieval signal (30). A similar isoform has not been identified in animals.

The finding that Gal (or Galβ1–3GalNAc1-benzyl; data not shown) did not produce significant inhibition of GalNAc-T4 compared with GalNAc suggests that the second step of O-glycosylation (extension of the oligosaccharide side chains), which is catalyzed by the β3-galactosyltransferase forming the core 1 structure Galβ1–3GalNAc1-O-Ser/Thr, may block the functional activity of the lectin domain of GalNAc-T4. Thus, once the O-glycan processing step involving elongation to the core 1 structure is accomplished, GalNAc-T4 would not be capable of catalyzing glycosylation of glycopeptides. This suggests that O-glycan elongation/branching and O-glycan density may be regulated by competition among GalNAc-transferases (lectin domain) and the glycosyltransferases involved in O-glycan extension, especially the core 1 synthase β3-Gal-transferase. Such a mechanism is possibly of advantage to complete O-glycan attachments to densely glycosylated regions such as those found in mucin tandem repeats, despite competition with elongation enzymes that appear to block further GalNAc incorporation (28).

What is the mechanism behind the lectin induced activation of glycopeptide activity? One hypothesis predicts that the lectin domain serves to tether the glycopeptide substrate to the enzyme, thereby providing a locally higher acceptor substrate concentration. Evidence in support of this would include a high affinity of the lectin domain for peptide GalNAc residues. Several attempts to demonstrate binding of GalNAc-transferases to GalNAc containing compounds including GalNAc-MUC1 glycopeptides have failed and hence not been able to support this hypothesis; however, since the affinity of the interaction is predicted to be low, this may simply be an experimental problem. Another hypothesis is that glycopeptide binding to the lectin domain induces conformational changes in the catalytic domain and/or lectin domain of the enzyme that activates it to include new substrate specificities. There is no structural information for GalNAc-T4; hence, this remains formally untested. However, molecular modeling analyses of GalNAc-transferases that compare them to known structures, including that of β4Gal-T1 (31), which exhibits limited sequence similarity in the catalytic domain, suggest that the lectin domain exists in a distinct structure in close proximity to the catalytic domain (14, 17). There are no additional mammalian glycosyltransferase structures with identified lectin domains available.

---

3 J. Sobrinho-Simoes, C. A. Reis, and H. Clausen, unpublished observation.

4 C. A. Reis, H. Hassan, and H. Clausen, unpublished observation.
to our knowledge, but recently the structure of a parasite sialyltransferase containing a similar distinct lectin domain was solved (32). Interestingly, it was originally proposed that a putative lectin domain in the C-terminal region of this sialidase/trans-sialyltransferase was involved in the catalytic unit of this enzyme. However, the structure of this enzyme with its substrate clearly demonstrated that the catalytic and lectin domains fold into two distinct tightly associated globular domains, and that the lectin domain is not directly involved in catalysis. The function of this lectin domain in this polypeptide remains unknown.

The polypeptide GalNAc-transferases are distributed throughout the Golgi cisternae, but there is evidence for some differences in distribution among isoforms (33, 34). The O-glycan processing step, involving initiation, elongation, branching, and termination of oligosaccharide chains, is believed to occur throughout the Golgi stacks and in the trans-Golgi network (for review, see Ref. 35). Some glycosyltransferases that are involved in these processing pathways have been found to be differentially located in cisternae of the Golgi, which is consistent with a function in oligosaccharide elongation. This compartmentalization in combination with distinct kinetic properties of the enzymes may be the key mechanism that regulates glycosylation patterns produced in different cell types. The finding that some polypeptide GalNAc-transferase isoforms have lectin domains with specificity for GalNAc-peptides offers new possibilities serving additional roles in Golgi through lectin-like adhesion functions. Lectin chaperones function in earlier parts of the secretory pathway (36). Similarly, ER to Golgi transport may be regulated in part by the ER-Glc-core 3 mannose lectin (36) through the ER-Golgi intermediate compartment. Such functions have not to our knowledge been identified in Golgi so far (35). O-Glycosylation has been shown to be involved in intracellular sorting (37), but the mechanism is unknown. Treatment of cells with millimolar concentrations of GalNAc-O-benzyl inhibits O-glycosylation, presumably by substrate competition with the core 1 β-galactosyltransferase and benzyl-oligosaccharide products are formed (38–41). Furthermore, a recent study demonstrates that benzyl-GalNAc selectively inhibits sialylation of apically sorted sialoglycoproteins (40). The effects of benzyl-GalNAc treatment are generally interpreted as being related to substrate competition with glycosyltransferases elongating O-glycans and the intracellular accumulation of benzyl-oligosaccharide products. The present finding that the function of lectin domains of polypeptide GalNAc-transferases may also be inhibited by similar concentrations provides an additional mechanism for the marked effects exerted by this compound. Interestingly, a parallel in proteoglycan biosynthesis, where exogenously added xylose-benzyl-oligosaccharide products are secreted (42).
The Lectin Domain of UDP-N-acetyl-d-galactosamine:PolypeptideN-acetylgalactosaminytransferase-T4 Directs Its Glycopeptide Specificities
Helle Hassan, Celso A. Reis, Eric Paul Bennett, Ekaterina Mirgorodskaya, Peter Roepstorff, Michael A. Hollingsworth, Joy Burchell, Joyce Taylor-Papadimitriou and Henrik Clausen

J. Biol. Chem. 2000, 275:38197-38205.
doi: 10.1074/jbc.M005783200 originally published online September 12, 2000

Access the most updated version of this article at doi: 10.1074/jbc.M005783200

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 40 references, 25 of which can be accessed free at http://www.jbc.org/content/275/49/38197.full.html#ref-list-1