⌘ *Author's Choice*

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 37, pp. 32684–32696, September 16, 2011

Lectin Domains of Polypeptide GalNAc Transferases Exhibit Glycopeptide Binding Specificity*[S]

Received for publication, June 20, 2011, and in revised form, July 9, 2011 Published, JBC Papers in Press, July 15, 2011, DOI 10.1074/jbc.M111.273722

Johannes W. Pedersen[‡], Eric P. Bennett[§], Katrine T.-B. G. Schjoldager[‡], Morten Meldal[¶], Andreas P. Holmér[‡], Ola Blixt[‡], Emiliano Cló[‡], Steven B. Levery[‡], Henrik Clausen[‡], and Hans H. Wandall[‡1]

From the [‡]Department of Cellular and Molecular Medicine and [§]School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N and the [¶]Carlsberg Laboratory and Nano Science Center, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

UDP-GalNAc:polypeptide *α*-*N*-acetylgalactosaminyltransferases (GalNAc-Ts) constitute a family of up to 20 transferases that initiate mucin-type *O*-glycosylation. The transferases are structurally composed of catalytic and lectin domains. Two modes have been identified for the selection of glycosylation sites by GalNAc-Ts: confined sequence recognition by the catalytic domain alone, and concerted recognition of acceptor sites and adjacent GalNAc-glycosylated sites by the catalytic and lectin domains, respectively. Thus far, only the catalytic domain has been shown to have peptide sequence specificity, whereas the primary function of the lectin domain is to increase affinity to previously glycosylated substrates. Whether the lectin domain also has peptide sequence selectivity has remained unclear. Using a glycopeptide array with a library of synthetic and recombinant glycopeptides based on sequences of mucins MUC1, MUC2, MUC4, MUC5AC, MUC6, and MUC7 as well as a random glycopeptide bead library, we examined the binding properties of four different lectin domains. The lectin domains of GalNAc-T1, -T2, -T3, and -T4 bound different subsets of small glycopeptides. These results indicate an additional level of complexity in the initiation step of *O*-glycosylation by GalNAc-Ts.

Mucin-type *O*-glycosylation is a common modification of mammalian proteins passing through the secretory pathway. Biosynthesis is initiated by the transfer of *N*-acetyl-*α*-D-galactosamine (GalNAc) to the hydroxyl group of serine and threonine residues (GalNAc*α*1-O-Ser/Thr), which is catalyzed by a large homologous family of uridine diphosphate (UDP)-*N*-acetyl-*α*-D-galactosamine:polypeptide GalNAc transferases (GalNAc-Ts, EC 2.4.1.41). The GalNAc-T family is evolutionarily conserved, and distinct subfamilies of up to 20 homologs have been identified among vertebrates and invertebrates (1–6). The GalNAc-T isoforms have different kinetic properties and cell and tissue expression patterns, suggesting that they serve nonredundant functions conserved through evolution (1, 3, 7–9).

Nonredundant functions have been verified for several GalNAc-Ts. Studies in *Drosophila melanogaster* have revealed essential functions of individual *Drosophila* GalNAc-Ts for embryonic development, epithelial adhesion, and the secretion of specific extracellular matrix proteins (6, 10, 11). In mice, GalNAc-T1 deficiency causes reduced lymphocyte homing and bleeding disorders (12), as well as defective glycosylation of the bone protein osteopontin (13). In humans, dysfunctional GalNAc-T3 causes the rare autosomal recessive metabolic disorder familial tumoral calcinosis (14, 15), and inactivating mutations have been found in GalNAc-T12 in cancer patients (16). In addition, GalNAc-T14 has been suggested as a negative regulator of apoptotic signaling (17), GalNAc-T16 as a regulator of TGF*β* signaling (18), and GalNAc-T11 as a candidate gene for congenital heart disease with heterotaxy (19). Moreover, genome wide association studies identified the GalNAc-T2 isoform as a candidate gene for dysregulated plasma HDL and cardiovascular risk (20), possibly due to its site-specific *O*-glycosylation of the lipase inhibitor ANGPTL3 (21). Collectively, these findings emphasize that distinct GalNAc-T isoforms have specialized and unique functions.

GalNAc-Ts are structurally composed of a catalytic domain and a lectin domain, which fold into an A type-fold and a *β*-trefoil-fold consisting of three subunits (*α*, *β*, and *γ*), respectively. Both domains are required for efficient glycosylation of mucin domains (22–28). Two different modes by which GalNAc-Ts select glycosylation sites have been identified: (i) independent recognition of acceptor substrate peptide sequence (±3–5 residues of Ser/Thr sites) by the catalytic domain and (ii) concerted recognition of the acceptor substrate by the catalytic domain combined with binding to adjacent *O*-GalNAc sites by the lectin domain (24, 29). Thus, the glycopeptide selectivity of GalNAc-Ts can be mediated by either the catalytic domain alone, as demonstrated by GalNAc-T10 that requires a GalNAc residue at position +1 adjacent to the threonine/serine to be glycosylated (30), or with the help of the lectin domain (22–27).

In the present study, we investigated the fine binding specificities of lectin domains from GalNAc-T1, -T2, -T3, and -T4 to elucidate potential selective binding properties for the adjacent peptide sequence around GalNAc-glycosylated sites. We previously demonstrated that all lectins bound GalNAc, but the question of GalNAc-glycopeptide specificity has remained unanswered. With the introduction of glycopeptide microar-

* This work was supported by the Novo Nordisk Foundation, Danish Medical Research Council, Danish Cancer Research Foundation, Agnes and Poul Friis Foundation, Danish Cancer Society, and the University of Copenhagen (Program of Excellence).

⌘ *Author's Choice*—Final version full access.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Table S1.

[1] To whom correspondence should be addressed: Dept. of Cellular and Molecular Medicine, Blegdamsvej 3, 2200 Copenhagen N, Denmark. E-mail: hhw@sund.ku.dk.

rays (31, 32), we have now been able to address this question. Using a large glycopeptide library based on the sequence of mucins MUC1, MUC2, MUC4, MUC5AC, MUC6, and MUC7 displayed on microarrays, we demonstrate that the lectin domains recognize differential subsets of small glycopeptides. This was further supported by an independent analysis of a random bead glycopeptide library probed with two lectins. The results underscore the complex and coordinated process by which the large polypeptide GalNAc-T enzyme family control protein *O*-glycosylation and provides evidence for a more differentiated role of the diverse sequences of the GalNAc-T lectins domains.

EXPERIMENTAL PROCEDURES

Expression and Purification of GalNAc-Ts—pAcGP67 baculovirus expression constructs containing the soluble coding region of human GalNAc-T1, GalNAc-T2 wild-type, GalNAc-T2 with a mutation in the α-lectin subdomain (GalNAc-T2^{458H}), GalNAc-T2 with a mutation in the D*X*D motif of the catalytic subdomain (GalNAc-T2^{224H}), GalNAc-T3 wild-type, GalNAc-T3 with a mutation in the α-lectin subdomain (GalNAc-T3^{519H}), GalNAc-T4 wild-type, and GalNAc-T4 with a mutation in the α-subdomain (GalNAc-T4^{459H}) were prepared as described previously (14, 24, 26, 33). The GalNAc-T1/T4 construct was generated by fusion of the T1 catalytic domain with the T4 lectin domain. The T1 catalytic domain was amplified by PCR using GalNAc-T1sol as a template and primers EBHC121 (33) and EBHC140 (5′-CCATAAGCTTCCTTTGTAACACCTGGAG-3′; the underlined point mutations create a HindIII restriction site and mutations V391E and D392A. The T4 lectin domain was excised from GalNAc-T4sol (34) using HindIII/BamHI, and the 548-bp fragment was co-subcloned with the BamHI/HindIII-cleaved T1 catalytic domain fragment into the BamHI site of pAcGP67 containing a His$_6$-T7 tag.

Plasmids pAcGP67-T1, -T2, -T2lec^{458H}, -T2^{224H}, -T3, -T3lec^{519H}, -T4, and -T4lec^{459H} were co-transfected with BaculoGold™ DNA (Pharmingen), and the recombinant baculovirus was obtained after two successive amplifications in Sf9 cells as described previously (3, 24). Amplified virus was used to infect High Five™ cells grown in serum-free medium (Invitrogen) in upright roller bottles shaken 140 rpm in water baths at 27 °C. Secreted, soluble recombinant proteins were harvested by centrifugation at 2,000 × *g* and the supernatants were subjected to chromatography on Amberlite® IRA-95, followed by dilution with 25 mM BisTris[2] (pH 6.5). The proteins were further purified by ion-exchange chromatography on 10 ml of SP Sepharose™ Fast Flow (Sigma) and eluted with NaCl in a gradient from 10 mM to 1 M. Fractions containing enzyme were pooled and concentrated using a Centriprep® YM-10 centrifugational filter unit with 10,000 Da cut-off (Millipore). His-tagged enzymes were further affinity purified by nickel-nitrilotriacetic acid purification as described previously (24). Activity and purity were analyzed by standard GalNAc transferase assay (Table 1) and SDS-PAGE stained with Coomassie Blue R-250 using bovine serum albumin as concentration standard.

GalNAc Transferase Assay—Standard assays were performed in 25–50 μl of total reaction mixtures containing 25 mM cacodylate (pH 7.4), 10 mM $MnCl_2$, 0.25% Triton X-100, 200 μM UDP-[^{14}C]GalNAc (1000–3,000 cpm/nmol) (Amersham Biosciences), 0.01–0.5 milliunits of GalNAc transferase, and 40–200 μM acceptor peptide. Activity assays were performed as short time assays and did not exceed 30 min. Enzymatic activity with the acceptor peptides was routinely determined by scintillation counting after Dowex-1 formic acid chromatography. Peptides and products produced by *in vitro* glycosylation were evaluated by mass spectrometry.

MALDI-TOF Mass Spectrometry—Reaction samples (0.5 μl) were diluted 15-fold in 0.1% TFA and applied directly to the probe with matrix (35). The matrix was 2,5-dihydroxybenzoic acid (25 mg/ml, Aldrich) dissolved in a 1:1 mixture of methanol and water. Mass spectra were acquired on a Voyager DE-Pro mass spectrometer equipped with delayed extraction (PerSeptive Biosystems).

Recombinant Mucin Fragments—Recombinant mucins were produced as described previously (36). Briefly, His-tagged mucin fragments covering sequences in MUC2, MUC4, MUC5AC, MUC6, and MUC7 were produced in *Escherichia coli* and purified by nickel chromatography and HPLC. The mucins were *in vitro* glycosylated with GalNAc T1–T4 to Tn (GalNAcα-O), and further elongated with ST6GalNAc-I or β3GnT6 to create STn and Core3. The products were verified by MALDI-TOF and SDS-PAGE.

Peptides—Peptides were synthesized by Schafer-n, Peptides&elephants, and Sigma, *in vitro* glycosylated, and purified using SepPak Vac C_{18} columns (Waters). A glycopeptide library covering different *O*-glycosylation sites in human IgA-hinge peptides was designed and synthesized by robotic multiwell SPPS (Syro II, Biotage) as described (37).

Mucin O-Glycopeptide Array Print and Analysis—Glycopeptides and control structures were printed as described previously (32). Briefly, nonglycosylated and glycosylated peptides and mucin fragments were printed on Schott Nexterion Slide H in quadruplicate using a BioRobotics MicroGrid II spotter (Genomics Solution). Prior to use, the microarrays were blocked for 1 h with 50 mM ethanolamine in 50 mM sodium borate (pH 8.5). Purified recombinant GalNAc-Ts were incubated in a closed container overnight (16 h) at 4 °C or for 1 h at 37 °C, washed three times in PBS, followed by a 2-h incubation with monoclonal antibodies directed against GalNAc-T1 (UH3), GalNAc-T2 (UH4), GalNAc-T3 (UH5) (7), or GalNAC-T4 (5B3).[3] The sample was then washed in PBS and incubated with biotinylated rabbit anti-mouse IgG (DAKO), followed by Cy3-conjugated streptavidin diluted in PBS with 0.05% Tween 20. Binding of catalytically active GalNAc-Ts was also tested using mAbs to the T7 tag or His tag with similar results for GalNAc-T2, -T3, and -T4. The assay, however, was more robust with GalNAc-T-specific mAbs and allowed the detection of GalNAc-Ts when bound to His/T7-tagged mucins.

[2] The abbreviations used are: BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; DMF, *N*,*N*-dimethylformamide; Fmoc, *N*-(9-fluorenyl)methoxycarbonyl; ETD, electron transfer dissociation.

[3] U. Mandel, unpublished data.

TABLE 1

Substrate specificities of GalNAc-T1, GalNAc-T2, GalNAc-T3, GalNAc-T4, GalNAc-T2^{458H}, GalNAc-T3^{519H}, and GalNAc-T4^{459H}, and GalNAc-T1catT4lec

Substrate specificities measured by short time essays not exceeding 30 min as described under "Experimental Procedures."

Acceptor substrate		GalNAc-T1	GalNAc-T2	GalNAc-T3	GalNAc-T4	GalNAc-T2^{458H}	GalNAc-T3^{519H}	GalNAc-T4^{459H}	GalNAc-T1catT4lec
						milliunits/mg			
MUC1	AHGVTSAPDTRPAPGSTAPP	64	212	54	10	155	136	6	20
MUC1a	RPAPGSTAPPA	39	40	44	12	11	112	0	12
MUC1b	AHGVTSAPDTR	0	222	2	0	118	18	0	3
MUC2	PTTTPITTTTTVTPTPTPTPTGTQTPTTTPISTTC	121	192	34	20	74	90	0	32
IgA	VPSTPPTPSPSTPPTPSPSK	133	214	11	12	104	49	0	28
MUC4	PVTYASSASTGDTTPLPVTDTSSVSTGHAT	0	13	14	3	1	34	0	6
OSM fragment	LSESTTQLPGGGPGCA	18	7	0	5	1	0	0	3
Onco-fib-VTHPGY	Ac-PFVTHPGYD	1	6	3	0	4	22	0	0
MUC7	CPPTPSATTPAPPSSAPPETTAA	51	ND[a]	ND	56	ND	ND	31	20

[a] ND, not determined.

Binding was quantified by scanning with a ProScanArray HT Microarray Scanner (PerkinElmer Life Sciences) and image analysis with ProScanArray Express 4.0 software (PerkinElmer Life Sciences). Each spot was measured in four replicates and the mean relative fluorescence intensity used. Data were analyzed and plotted using Microsoft Excel or GraphPad Prism software.

Inhibition Studies—Inhibition studies were performed with Bn-α-GalNAc, Bn-α-GlcNAc (5 mM, respectively), or free GalNAc and GlcNAc (250 mM). In addition, inhibition was performed with selected peptides at a concentration of 0.1–1 mM.

Combinatorial Glycopeptide Synthesis—A combinatorial glycopeptide library was synthesized using biocompatible $PEGA_{1900}$ resin (2 g, 24 ml, 0.24 mmol of NH_2) (38). The resin was carefully washed with *N*,*N*-dimethylformamide (DMF) and distributed evenly into a multiple column (20 columns) library generator (39). The amino groups of the resin were derivatized with the 20 genetically encoded Fmoc-derivative (3 eq), *N*-[1-*H*-benzotriazol-1-yl)-(dimethylamino)methylene]-*N*-methylmetanaminium tetrafluoroborate *N*-oxide (3 eq), and 4-methylmorpholine (3 eq) by pre-mixing the reagents in 10 ml of DMF 2 min prior to addition to the swelled resin. The resin was washed with DMF and the Fmoc group removed with 20% piperidine in DMF. A 11-mer peptide library assembly was created using coupling conditions (Fmoc-AA-OH, *N*-[1*H*-benzotriazol-1-yl)-(dimethylamino)methylene]-*N*-methylmetanaminium tetrafluoroborate *N*-oxide, 4-methylmorpholine, and DMF) and piperidine-mediated deprotection as described above. Between each coupling step, after Fmoc removal, the library generator was half-filled with DMF and tightly closed. It was turned upside down and the resin beads were thoroughly mixed by vigorous agitation of the synthesizer. The generator was turned upright and even redistribution of beads in the wells was controlled. At the sixth coupling of the combinatorial assembly, Fmoc-Ser(Ac_4GalNAc)-OPfp (2 eq) was used, whereas five positions on either side comprised even amounts of the 20 encoded amino acids. The library was transferred to a fritted syringe and washed with DMF and water. Excess liquid was removed. The protecting groups were cleaved with 92% TFA containing ethane dithiol (1%), triisopropylsilane (2%), and water (5%) for 2.5 h and the resin was washed with water, DMF, 20% piperidine/DMF, DMF, water, and plenty of dry methanol in succession. The methanol-swelled resin was covered with NaOMe/methanol (0.1 M, 20 ml) for 30 min, washed, and the treatment repeated for 2 h. The deacetylated glycopeptide library was washed with methanol and water before use in screening. Test sequencing of single library beads using Edman degradation on an Abi-431 sequencer showed that the compounds were as expected and gave the expected sequence result.

Combinatorial Substrate Screening—Purified lectin domains from GalNAc-T4 (GalNAc-T4lec) or -T2 (GalNAc-T2lec) were both reacted with Oregon Green succinimide ester (2.5 eq, Molecular Probes), and the labeled protein was purified using gel permeation chromatography. The above library was divided into two aliquots. An aliquot of 300,000 beads was incubated with labeled GalNAc-T4lec or GalNAc-T2lec. Each mixture was passed through an automated COMPAS bead sorter

(Union Biometrica) and the active beads were removed. The inactive beads from each sorting were incubated with the other protein and positives were isolated. In this manner, selective T2 and T4 ligands were obtained. All labeled beads were subjected to amino acid sequencing (Table 2).

Characterization of O-Glycosylation Sites—Products of *O*-glycosylation reactions were characterized by electrospray ionization-linear ion trap-Fourier transform mass spectrometry (ESI-LIT-FT-MS) in an LTQ-Orbitrap XL hybrid spectrometer (Thermo-Scientific) equipped for both high energy collision-induced dissociation (40) in an external collision cell (41), and electron transfer dissociation (ETD) (42) for peptide sequence analysis by MS/MS (MS^2) with retention of glycan site-specific fragments. Samples were dissolved in methanol/water (1:1) containing 1% formic acid and introduced by direct infusion via a TriVersa NanoMate ESI-Chip interface (Advion BioSystems) at a flow rate of ~100 nl/min and 1.4 kV spray voltage. Mass spectra were acquired in positive ion FT mode at a nominal resolving power of 30,000. Similar to previous studies (21), following acquisition of a survey FT-MS^1 spectrum, ETD-FT-MS^2 product ion spectra were acquired on all observed multiply charged precursor ion clusters including those with a m/z ratio consistent with metalated forms (*i.e.* MH_3K^{4+} and MH_2K^{3+}, as well as MH_4^{4+}, MH_3^{3+}, and MH_2^{2+}), using an isolation width of 5 mass units, normalized collision energy of 35%, activation Q of 0.25, activation times of 100–300 ms, and supplemental activation of 20% (43). Experimental MS spectra were analyzed by comparison with exact m/z values calculated for the observed molecular ion charge states using the known MUC4.1 peptide sequence incremented with 1 and 2 HexNAc residues; the presence of an oxidized Met residue in some +16 mass units precursors was confirmed by acquisition of high energy collision-induced dissociation-MS^2 spectra (isolation width, 5 μm; activation time, 30 ms; activation Q, 0.25; using, normalized collision energy, 30–50%), which allowed straightforward sequencing from efficient generation of completely deglycosylated *b* and *y* ions. ETD-MS^2 spectra were analyzed by comparison with theoretical c- and z$^{\bullet}$- fragment m/z values calculated for all positional combinations of one and two HexNAc residues distributed on the eight potential S and T glycosylation sites in the sequence. Additional calculations were made by replacing one H^+ with K^+, assuming a strong association of the metal ion with a glycosylated residue. Calculations were performed using the web-based Protein Prospector MS Isotope and MS Product software routines.

RESULTS

Direct Binding of GalNAc-T1, -T2, -T3, and -T4 to GalNAc-glycosylated Mucin Peptides—A binding assay was developed on a microarray platform to assess the binding of secreted GalNAc-Ts to various glycopeptides and 12–23-kDa mucin fragments carrying GalNAcα (Tn), Neu Acα2,6GalNAcα (STn), or GlcNAcβ1,3GalNAcα (truncated Core3). The printing efficiency and presentation of the specific glycoforms on peptides and recombinant mucin fragments was verified with Tn- and STn-specific mAbs, as well as lectins (HPA and VVA). All four GalNAc-Ts bound GalNAc-glycosylated peptides and mucin fragments (Fig. 1*B*) via their lectin domains, as determined by assay conditions designed to prevent binding by the catalytic units (omission of Mn^{2+} and UDP) (33, 44). Binding was only observed with MUC1 glycopeptides displaying GalNAc, whereas nonglycosylated peptides or glycopeptides carrying other *O*-glycans (MUC1 carrying either T (Galβ1,3GalNAcα), STn, or truncated Core-3) did not react. We did observe weak reactivity with the larger Core-3 and STn-glycosylated mucin constructs. However, based on MS analysis of these the glycosylation was quite heterogeneous, both in terms of sites and the degree of capping with sialic acid and GlcNAc and we interpret the binding to these STn and Core-3 glycoforms as representing binding to residual uncapped Tn *O*-glycans. This conclusion is supported by similar reactivity with GalNAc reactive lectins and mAbs to these particular constructs. Lectin-mediated binding was confirmed by complete inhibition of binding with 250 mM GalNAc monosaccharide and no inhibition with GlcNAc (not shown). To address the importance of the catalytic domain in selective glycopeptide recognition, we analyzed the binding of GalNAc-Ts with deleterious mutations in the catalytic (GalNAc-T2) or lectin domain (GalNAc-T2, -T3, and -T4) (Fig. 1*A*). GalNAc-T2 with the mutated catalytic domain bound similar to the wild-type enzyme, whereas GalNAc-T2, -T3, and -T4 with mutated lectin domains exhibited complete loss of binding in agreement with our previous studies (24).

The influence of pH and divalent cations on GalNAc-Ts binding to GalNAc peptides was analyzed. Interestingly, the pH optima for all four enzymes was between 6.5 and 7.4 (Fig. 2). A dramatic decrease in binding was seen at pH $<$ 6, which potentially relates to the requirement of releasing GalNAc-Ts from their acceptors in the trans-Golgi network, where pH is 6 or lower (45). Next, metal ion requirements for binding were analyzed and no change in lectin binding was seen when performing the binding assay in the presence of 20 mM EDTA; 5, 10, and 20 mM Ca^{2+}; or Mn^{2+} (not shown).

The Binding of GalNAc-T Lectins Is Influenced by the Peptide Sequence Surrounding the GalNAc Residue—Exploiting the specific glycosylation patterns created by each of the four GalNAc-Ts (T1, T2, T3, and T4), mucin fragments were partially glycosylated on the array by individual GalNAc-Ts and the differentially glycosylated mucins were used as bait to investigate differences in binding between the four GalNAc-T lectins. The most pronounced difference was detected with GalNAc-T4 that preferentially recognized MUC4 over MUC5AC glycosylated with GalNAc-T2, regardless of equal reactivity of the two mucins with GalNAc-recognizing lectins (Fig. 1*C*, *middle panel*). In contrast, GalNAc-T3 preferentially recognized MUC5AC, but not MUC4, when the mucins were glycosylated with GalNAc-T1, -T2, and -T3 (Fig. 1*C*). These findings demonstrate that the sites and patterns of GalNAc decoration of mucin-type peptide substrates influence binding of different GalNAc-T lectin domains.

Next, we analyzed a series of overlapping GalNAc-glycosylated 20-mer peptides covering the MUC4 fragment (supplemental Table S1). Based on our previous findings that the velocity of IgA hinge glycosylation by GalNAc-T2 was dependent on a functional lectin domain, we included a collection of synthetically produced IgA-hinge peptides. The peptides were modi-

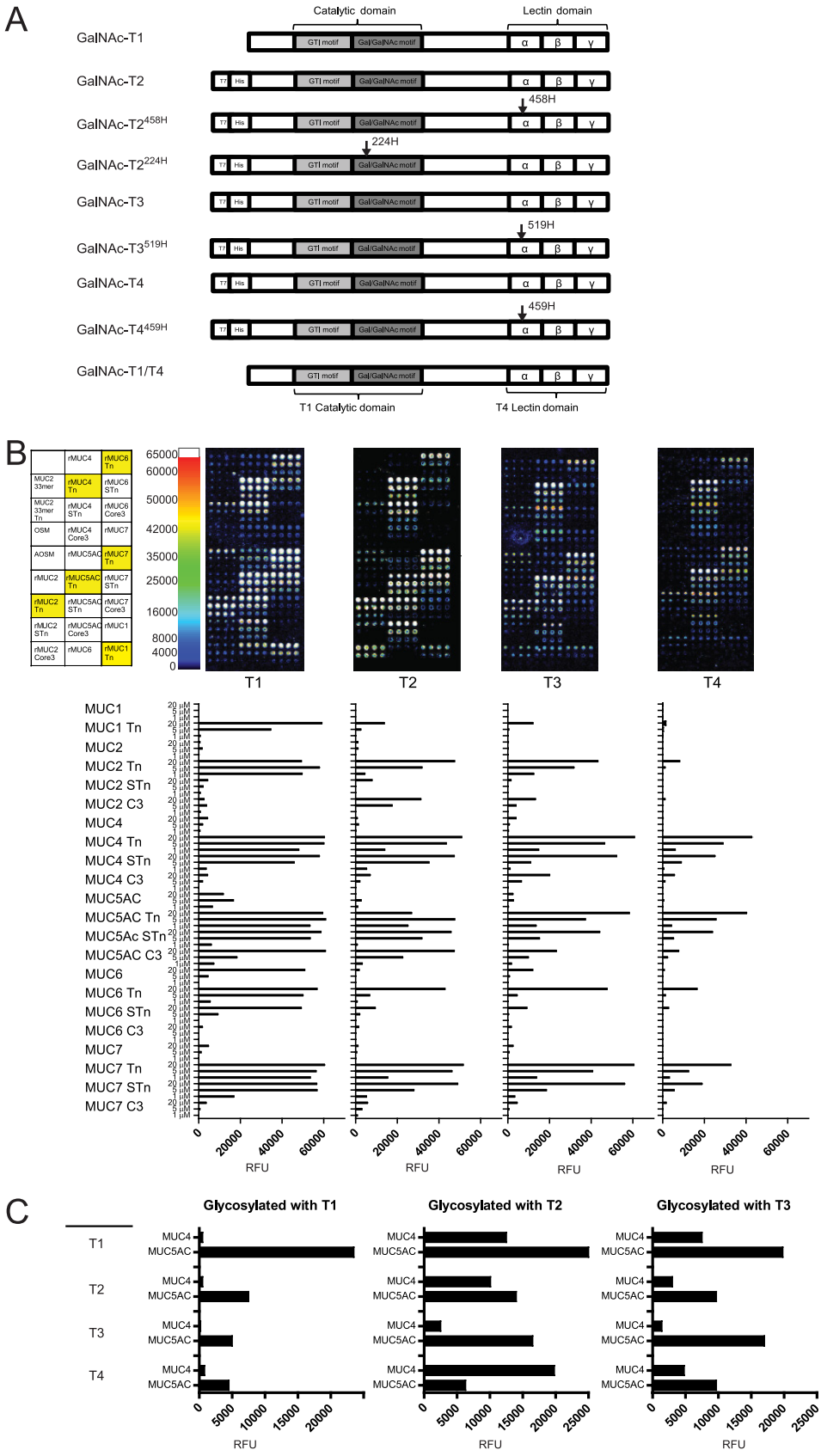


FIGURE 1. **Binding of GalNAc transferase lectins to mucin GalNAc glycopeptides.** *A,* schematic depiction of GalNAc transferases. All GalNAc-Ts are single pass type II transmembrane proteins with catalytic and lectin domains. The localization of mutations in GalNAc-T2^{458H}, GalNAc-T2^{224H}, GalNAc-T3^{519H}, GalNAc-T4^{459H}, and GalNAc-T4lec^{459H} is shown by *arrows*. *B,* fluorescent image of a mucin glycoprotein array probed with GalNAc-T1, -T2, -T3, and -T4. Quantification of binding with the relative fluorescent units is shown on the *x* axis. *C,* GalNAc-T1, -T2, -T3, and -T4 binding to partially glycosylated mucin fragments MUC4 and MUC5AC on the array by individual GalNAc-Ts. Differentially glycosylated mucins were used as baits to probe GalNAc transferase binding.

fied with GalNAc residues attached at different sites (supplemental Table S1) to investigate if a differential binding pattern could be observed between the GalNAc-T isoforms. GalNAc-T1, -T2, -T3, and -T4 also exhibited different binding patterns to this collection of short GalNAc peptides (Fig. 3). A striking result was a particularly strong binding of GalNAc-T1 to one specific MUC4 glycopeptide (GalNAc-MUC4.1, supplemental Table S1) that was only weakly recognized by GalNAc-T2, -T3, and -T4. In contrast, both GalNAc-T3 and -T4 exhibited strong binding to the MUC4.17 glycopeptide. Furthermore, GalNAc-T2 bound IgA-hinge 7 with a single GalNAc in Thr-12 specifically, whereas GalNAc-T1, -T3, and -T4 did not bind this glycopeptide. Similarly, GalNAc-T3 bound IgA-hinge 12 but did not bind other IgA-hinge peptides. Although GalNAc-T4 had a distinct binding pattern different from the other enzymes, we did not find MUC4 or IgA glycopeptides recognized by only GalNAc-T4. GalNAc-Ts with mutated lectin domains exhibited complete loss of binding in agreement with our previous studies.

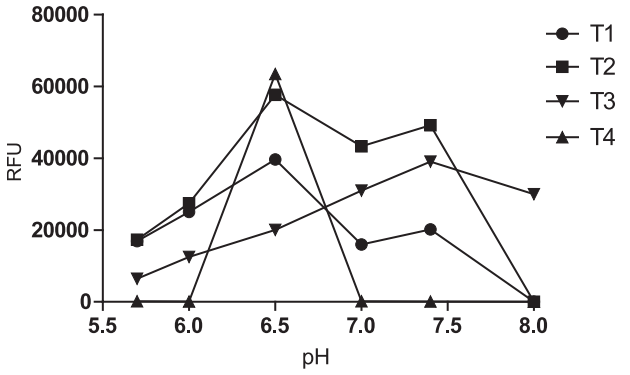


FIGURE 2. **Influence of pH on GalNAc-T binding to MUC1 Tn glycopeptides.** Binding of GalNAc-T1, -T2, -T3, and -T4 to GalNAc-MUC1 at pH 5.7–8.0. Binding is expressed as the mean relative fluorescent activity of MUC1 GalNAc-glycopeptides.

The GalNAc-peptide Specificity Modulates Function of GalNAc-Ts in Follow-up Glycosylation Reactions—We initially attempted to examine the different binding patterns of GalNAc-T1 and -T4 by swapping the lectin domains of GalNAc-T1 and -T4, as well as those of GalNAc-T2 and -T4. Unfortunately, only the GalNAc-T1catT4lec chimeric protein was functional, and that only with very low specific activity (Table 1). GalNAc-T1catT4lec exhibited weak binding to the GalNAc-T1 binding partner GalNAc-MUC4.1, although the binding pattern was not identical to the binding achieved with wild-type GalNAc-T4, suggesting that binding might be influenced by the catalytic domain (not shown). Next, we tested whether the selective binding of GalNAc-T1 to GalNAc-MUC4.1 correlated with the capacity of the enzyme to mediate follow-up glycosylation. Glycosylation of MUC4.1 with GalNAc-T2 resulted in the incorporation of a single GalNAc residue at position Thr-12 (Fig. 4). Only GalNAc-T1 bound GalNAc-MUC4.1, and follow-up glycosylation of GalNAc-MUC4.1 was most effective with GalNAc-T1 (Fig. 4),

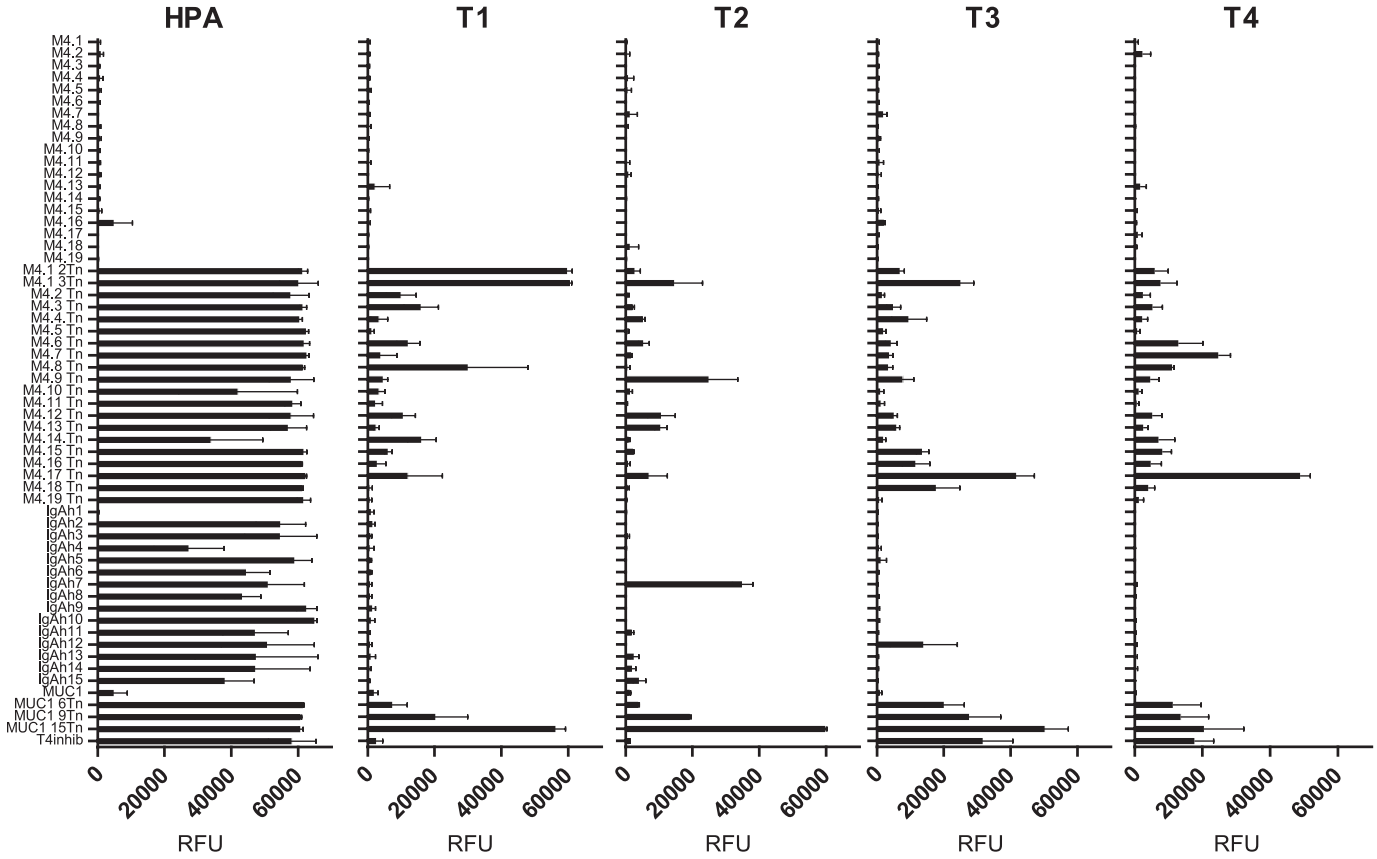


FIGURE 3. **Differential binding of GalNAc transferase lectin domains to GalNAc glycopeptides.** Binding of GalNAc-T1, -T2, -T3, -T4, and HPA to the MUC4 and IgA glycopeptide array. The *bar graph* represents the relative fluorescent unit of the binding of each GalNAc-T. Peptide sequences are listed under supplemental Table S1. *HPA*, helix pomatia agglutinin.

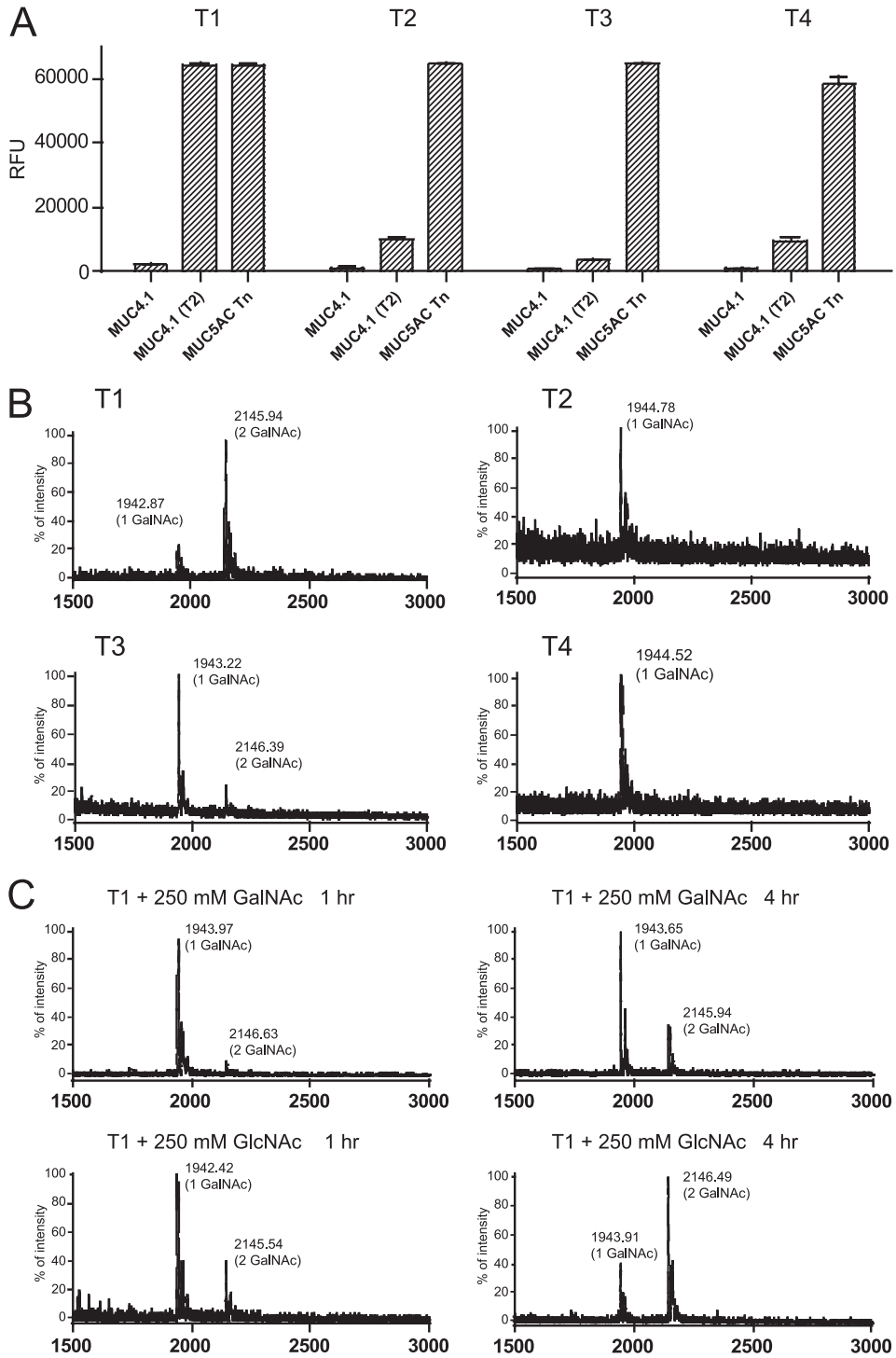
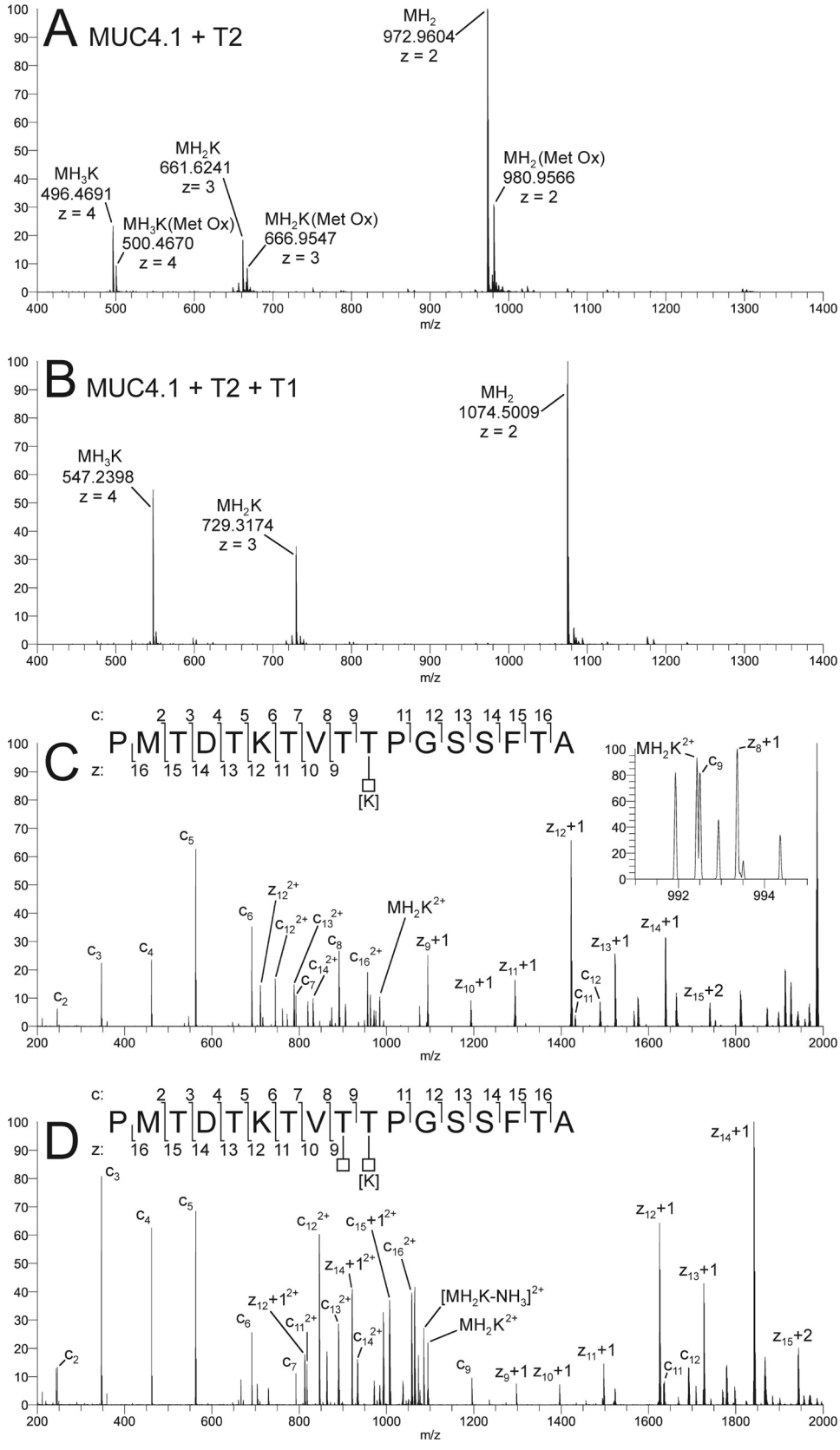


FIGURE 4. **GalNAc-T1 selectively binds and mediates followup glycosylation of GalNAc-MUC4.1.** *A*, binding of GalNAc-T1, -T2, -T3, and -T4 to MUC4.1, GalNAc$^{n\ =\ 1\text{-}pos12}$-MUC4.1 (GalNAc incorporated at Thr-12 by GalNAc-T2, see Fig. 5 for characterization), and GalNAc-glycosylated MUC5AC (recMUC5AC Tn). *B*, glycosylation of GalNAc$^{n\ =\ 1\text{-}pos12}$-MUC4.1 by GalNAc-T1, -T2, -T3, and -T4. *C*, glycosylation of GalNAc$^{n\ =\ 1\text{-}pos12}$-MUC4.1 by GalNAc-T1 for 1 and 4 h in the presence of 250 mM GalNAc (*upper panels*) or GlcNAc (*lower panels*) evaluated by MALDI-TOF MS. Number of GalNAcs incorporated is indicated. *Error bars* indicate standard deviation of four replicates.

incorporating an additional GalNAc residue at Thr-11 adjacent to the pre-existing GalNAc at position Thr-12 (Fig. 5).

We then tested the potential effect of dual binding to substrates through the catalytic and lectin domains. In the presence of Mn^{2+} and UDP GalNAc-Ts bind glycopeptides via both the catalytic and lectin domains. To test whether GalNAc-T binding to glycopeptides is increased when binding is allowed through both domains, a model system with GalNAc-T2 binding to the nonglycosylated and GalNAc-glycosylated recombinant MUC2 fragment was established. Inclusion of Mn^{2+} and UDP in the assay enabled GalNAc-T2 binding, with or without mutations in the lectin domain, to nonglycosylated mucin dem-

onstrating binding via the catalytic domain alone. Inclusion of Mn^{2+} and UDP in the assay did not change binding of GalNAc-T2 to GalNAc-MUC2 (Fig. 6). The contribution of the catalytic domain to the collective binding was assessed by binding GalNAc-T2 lectin mutant (GalNAc-$T2^{458H}$) to GalNAc-MUC2 with the inclusion of Mn^{2+} and UDP in the assay (Fig. 6,



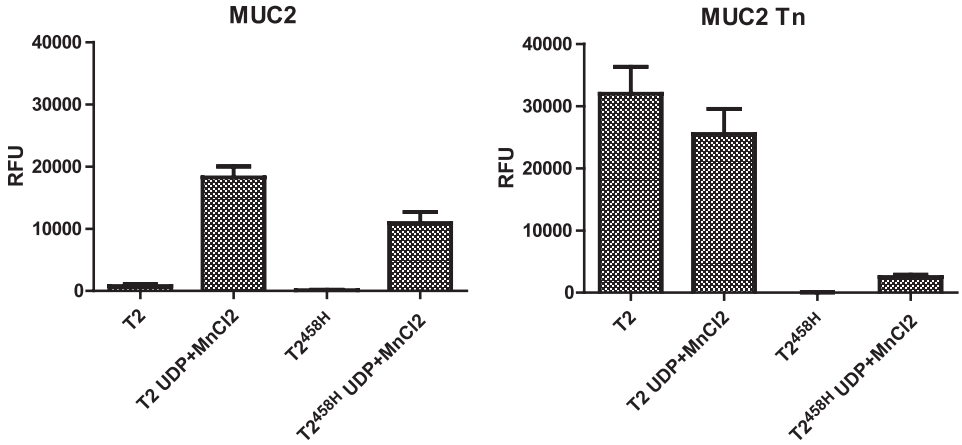


FIGURE 6. **Evaluation of GalNAc-T2 binding to recombinant MUC2.** *Left panel,* binding of GalNAc-T2 and GalNAc-T2^{458H} to nonglycosylated recombinant MUC2. Binding assay included UDP and $MnCl_2$ to allow binding through the catalytic domain. *Right panel,* binding of GalNAc-T2 and GalNAc-T2^{458H} to MUC2 Tn. *Error bars* indicate standard deviation of four replicates.

right panel). This indicated that the catalytic domain under these binding conditions only contributed ~20% of the total binding.

Exploring Binding Motifs of GalNAc-T2 and -T4 Lectin Domains (GalNAc-Tlec) with a Random Tn-glycopeptide Bead Library—Finally, we sought to identify specific glycopeptide sequence motifs for GalNAc-T2lec and -T4lec by screening a random bead library presenting 11-mer GalNAc glycopeptides with a central GalNAc attached to serine at position 6. Seven GalNAc-containing peptides selectively recognized GalNAc-T2lec but not GalNAc-T4lec (Table 2). In addition, 13 GalNAc glycopeptides selectively recognized the GalNAc-T4 lectin domain but not the GalNAc-T2 lectin domain (Table 2). The selected peptides for both GalNac-T2lec and GalNac-T4lec are hydrophobic, consistent with a preference for hydrophobic binding sites, but considerably more so for GalNAc-T4lec. A large bias toward selection of Arg over Lys as a basic amino acid was observed, as well as preference for the +1 site for GalNac-T2lec. Both lectins bound a large proportion of GalNAc peptides with hydrophobic amino acids Ile, Val, and, to some extent, Leu in position +2, although this was most pronounced for GalNAc-T2lec. A generally high occurrence of β-branched amino acids (*e.g.* Val/Ile/Thr) was found, consistent with selectivity for peptides with sequential regions of threonines, such as the mucins. The fact that Val/Ile appear as substitutes for Thr is not surprising, as Thr is uncharged, with some hydrophobic character, as well as shape similarity, conferred by the β-branched methyl group, and this replacement is frequent in natural protein mutations. Some differences were also noted. Glycopeptides containing Phe, Trp, and Ile at positions −3 and −4 were more prevalent among GalNAc-T4lec binders than GalNAc-T2lec binders (position −3, 5/13 for T4 and 1/7 for T2; position −4, 5/13 for T4 and 0/7 for T2). In contrast, GalNAc-

TABLE 2

Differential binding of glycopeptide beads by GalNAc-T2 and -T4

Selection of glycopeptides binding to isolated lectin domains from GalNAc-T2 (T2-Lec) and -T4 (T4-Lec). Glycopeptides were selected from an 11-mer random peptide library surrounding a central serine-GalNAc residue at position six. *A,* peptide sequences of glycopeptides pre-adsorbed with isolated T2 lectin domain and subsequently selected with isolated T4 lectin domain. *B,* peptide sequences of glycopeptides pre-adsorbed with isolated T4 lectin domain and subsequently selected with isolated T2 lectin domain. Red indicates basic amino acid, blue are aromatic, and green are β-branched amino acid. "−" indicates nondetermined amino acid sequences.

A) T2-Lec negative beads incubated with T4-Lec[1]

	-5	-4	-3	-2	-1		+1	+2	+3	+4	+5
1	R	I	T	G	K	S(GalNAc)	M	Y	W	K	P
2	A	N	W	N	P	S(GalNAc)	T	R	I	F	V
3	H	C	F	R	K	S(GalNAc)	L	I	R	I	F
4	W	M	---	---	---	S(GalNAc)	R	I	W	F	D
5	S	Q	P	R	L	S(GalNAc)	F	V	L	R	M
6	R	W	S	F	Y	S(GalNAc)	W	R	A	Y	I
7	Y	A	M	L	Q	S(GalNAc)	V	R	F	F	I
8	D	W	W	R	I	S(GalNAc)	A	L	A	R	V
9	N	T	F	A	I	S(GalNAc)	A	G	R	N	A
10	R	I	L	L	--	S(GalNAc)	R	F	V	R	L
11	W	F	Q	I	--	S(GalNAc)	I	F	I	W	N
12	A	F	M	R	A	S(GalNAc)	N	I	I	Q	W
13	W	K	W	V	Q	S(GalNAc)	I	V	Q	R	M
14	Q	W	K	S	R	S(GalNAc)	F	I	--	I	Y

B) T4-Lec negative beads incubated with T2-Lec[1]

	-5	-4	-3	-2	-1		+1	+2	+3	+4	+5
1	H	P	V	M	Q	S(GalNAc)	R/C	I	F	Q	W
2	S	M	S	L	--	S(GalNAc)	R	L	F	N	F
3	F	R	F	I	N	S(GalNAc)	H	M	H	V	Y
4	M	P	I	Y	I	S(GalNAc)	R	T	F	T	F
5	M	P	K	I	T	S(GalNAc)	R	V	K	I	V
6	V	--	M	Q	N	S(GalNAc)	R	I	L	V	S
7	W	G	S	H	I	S(GalNAc)	V	V	R	A	S

FIGURE 5. **Glycosylation of the MUC4.1 peptide by GalNAc-T2, followed by GalNAc-T1.** Characterization of MUC4.1 products using ESI-LTQ-Orbitrap MS and ETD-MS/MS. *A*, MS of the MUC4.1 + 1Tn product formed by GalNAc-T2; *B*, MS of MUC4.1 + 2Tn product formed by subsequent action of GalNAc-T1; *C*, ETD-MS/MS of $[M + K^+3H]^{4+}$ precursor ion (*m/z* 496.4691, *z* = 4, in *panel A*); *D*, ETD-MS/MS of $[M + K^+3H]^{4+}$ precursor ion (*m/z* 547.2398, *z* = 4, in *panel B*). *Inset* in *C*, expanded region, *m/z* 991–995 from ETD-MS/MS of the $[M + K^+2H]^{3+}$ precursor ion (*m/z* 661.6241, *z* = 3, in *panel A*) of GalNAc-T2, showing a well resolved c_9 ion at a *m/z* consistent with nonglycosylation of Thr-9. The fragmentation patterns in *C* and *D* are consistent with glycosylation of Thr-10, and Thr-9 and Thr-10, respectively, with HexNAc residues (□); *i.e.* abundant c_7^+, c_8^+, c_9^+, c_{11}^+, $z^{\bullet}_9{}^+$, and related ions ($z^{\bullet}_n \pm 1$ u, $c_n \pm 1$ u and/or corresponding doubly charged species) were detected at *m/z* values calculated for HexNAc at these sites. Corresponding fragments consistent with glycosylation elsewhere were not detected.

T2lec more readily bound glycopeptides with the charged amino acid Arg at position +1 (4/7 for T2 and 2/13 for T4) and hydrophobic amino acid Phe at position +3 (3/7 for T2 and 1/13 for T4; Table 2). Based on the identified amino acid motifs important for binding GalNAc-T4lec and GalNAc-T2lec, one glycopeptide (T4inhib) was constructed that was predicted to bind GalNAc-T4. Indeed, the T4inhib GalNAc peptide primarily recognized GalNAc-T4 and not GalNAc-T1 or GalNAc-T2, whereas reactivity was seen with GalNAc-T3 (Fig. 3). Inhibition studies confirmed this reactivity pattern (not shown).

DISCUSSION

GalNAc-Ts are the only eukaryotic glycosyltransferases that have lectin domains in addition to catalytic domains (28). In contrast lectin domains are found on many glycosidases as well as some glycosyltransferases in prokaryotes where they have been shown to support catalytic efficiency by mechanisms of substrate approximation and targeting to substrate sites (46–48). The lectin domain structure is evolutionarily conserved throughout the large GalNAc-T gene family and several GalNAc-T lectins were previously shown to bind GalNAc and modulate the catalytic functions of the enzymes (24, 26). The results presented here demonstrate that the lectin domains furthermore, exhibit differential preferences for GalNAc glycopeptides with different patterns of glycosylation. These findings provide for a role for lectin domains in controlling substrate specificities of GalNAc-Ts and the complex interplay of these in control of protein *O*-glycosylation. This advance was facilitated by application of our glycopeptide libraries displayed on microarrays, which circumvented previous problems with detecting weak binding in ELISA (32).

The important role of the GalNAc-T lectin domains for the function and substrate specificity of the enzymes was originally observed by the requirement of a functional lectin domain of GalNAc-T4 to complete the *O*-glycosylation of the MUC1 tandem repeat (26). Subsequently, we and others demonstrated that other GalNAc-T isoforms also required functional lectin domains for what were referred to as follow-up glycosylation events (23, 29, 49). We have hypothesized that the lectins serve to enhance the efficiency of the initiation step of *O*-glycosylation by providing additional binding to partially completed substrates to complete the process prior to the elongation step of *O*-glycans. Failure to do so could prematurely obstruct the initiation process especially in highly dense *O*-glycan regions such as mucins. In support hereof, a time course study of the *O*-glycosylation of the large secreted MUC5AC mucin showed that glycosylation with GalNAc could be separated in time from that of the *O*-glycan elongation (50). GalNAc-Ts glycosylate peptides with multiple acceptor sites by order of affinity for individual sites. This feature would disfavor glycosylation of low affinity acceptor sites in regions with high density of *O*-glycosylation and potentially result in incomplete decoration of such regions. We envision that greater efficiency in completion of mucin regions with high density of glycosylation can be achieved by the additional recognition of partially glycosylated substrates by the lectin domains. The finding in this report that the lectin domains not only bind GalNAc residues, but preferentially bind certain GalNAc-glycopeptide patterns, further supports this hypothesis.

The glycopeptide array platform provided a sensitive assay to study GalNAc-T lectin specificities, which we previously were unable to do by a conventional enzyme-linked lectin assay (24). We previously failed to demonstrate clear binding with non-labeled transferases (24); however, the sensitive glycopeptide array assay made it possible to detect the weak protein-carbohydrate interactions. This is most likely due to the high density and uniform orientation of glycopeptides on the array. Another possible explanation is stabilization of lectin binding by the addition of secondary antibodies. Similar advances in the microarray format over other binding assays have been noted when analyzing other low affinity protein carbohydrate interactions (51).

Most importantly, we found that the glycopeptide binding specificity of GalNAc-T lectins corresponded with distinct substrate specificities of GalNAc-Ts suggesting that the lectin domains provide more than just binding to GalNAc residues but also influence substrate selection. Selective binding patterns were observed for each GalNAc-T. For example, GalNAc-T1 selectively bound GalNAc-MUC4 peptide corresponding with its lectin domain-dependent follow-up glycosylation. Furthermore, GalNAc-T2 bound selectively to the GalNAc-IgA-hinge peptide with a single GalNAc at Thr-12, which represents one of the initial glycosylation products produced by GalNAc-T2 (52), and this might explain GalNAc-T2 lectin-dependent glycosylation of the IgA-hinge (24, 52). It should be noted, however, that the GalNAc-glycopeptide binding specificity by array analysis is compared with the lectin-dependent function of the enzymes and not the general activity with unglycosylated peptides.

It was surprising that the lectin domain-mediated follow-up glycosylation of GalNAc-MUC4.1 resulted in addition of a GalNAc residue immediately adjacent to the pre-existing GalNAc-glycosylated site in a di-GalNAc motif. Previously, modeling of MUC1 glycopeptides into the crystal structures of GalNAc-T1 (23) has suggested that a specific and significant distance between binding sites of the catalytic and lectin domains is critical in the follow-up glycosylation process. This is in contrast to the follow-up glycosylation mediated by GalNAc-T10, which requires a GalNAc residue at position +1 adjacent to the threonine/serine to be glycosylated (30); explained by a unique GalNAc-O-Ser/Thr-binding site in the catalytic domain of GalNAc-T10 not found in GalNAc-T1 or -T2. Therefore, it was surprising that the GalNAc-T1 lectin domain mediated follow-up glycosylation of GalNAc-MUC4.1 resulting in addition of a GalNAc residue immediately adjacent to the pre-existing GalNAc-glycosylated site in a di-GalNAc motif. Nevertheless, the importance of the lectin domain in this process was confirmed by coincubation with GalNAc that abolished GalNAc-MUC4 binding and inhibited follow-up glycosylation. The result may therefore suggest a mechanism of action of lectin domains different from that previously proposed for GalNAc-T1. One possibility is that molecular flexibility between the catalytic and lectin domains of GalNAc-T1 allows the two binding sites to be in close proximity (22, 23, 25), as seen evidenced by the reciprocal orientation of the catalytic

and hemopexin-like domains of MMP-12 (53). Another possibility is that the lectin domain functions by substrate approximation (increasing the local concentration of the peptide) and creates a selective and lectin domain-dependent advantage for GalNAc-T1 in glycosylation of the preferred substrate-like GalNAc-MUC4.1. In summary the lectin domain could influence substrate specificity differently in the two cases (*i.e.* by substrate approximation in the case of GalNAc-T1 and MUC4.1), and correct spacing between the acceptor and lectin binding sites, as demonstrated with GalNAc-T2/T4 for GalNAc-MUC1.

This could have implications for the methods used to determine the site preferences of individual GalNAc-T isoforms. Important studies of the glycosylation of peptide libraries has revealed unique as well as general preferences for GalNAc-Ts isoforms (2, 30, 54, 55). It has been defined that prolines in position +1 or +3 relative to the glycosylation site are of general importance. The method has also been used to define unique glycosylation sites for individual GalNAc-Ts with tyrosine in position +3 being important for GalNAc-T1 and a valine, isoleucine, or proline in position −3 being important for GalNAc-T2 (54). The use of short synthetic peptides has thus successfully identified consensus sequences for substrate specificity of individual GalNAc-Ts conferred by the catalytic domain. The method does not, however, take into account the potential importance of the lectin domain; a feature that based on the present study could prove to be important for future algorithms of glycosylation prediction.

An interesting question, given the gradual decrease in pH found through the secretory pathway (45), is whether the lectin binding function of GalNAc-Ts is regulated by pH. A number of lectin-mediated functions in intracellular transport are terminated by subtle changes in pH, which result in release of lectin binding. The unique double binding function of GalNAc-Ts via their catalytic and lectin domains may provide these enzymes the ability to compete with *O*-glycan elongation enzymes. Release of substrates after completion of the GalNAc-glycosylation step and initiation of elongation could potentially be due to lack of binding affinity through the catalytic domain, whereas weak interactions with the GalNAc-protein substrate posed by the lectin domain alone could be overcome. However, one could also envision that the lectin binding affinity alone still posed a competitive challenge for the elongation process and a simple mechanism to overcome this would be a pH-induced release of GalNAc-Ts at pH below 6.0 in the late secretory pathway. Interestingly, the differences in the optimum pH for GalNAc-T binding could partly correspond to different localizations of GalNAc-T1, -T2, and -T3 along the secretory pathway, with GalNAc-T1 being localized throughout and GalNAc-T3 in the medial part of the Golgi stack (56).

The random bead library approach helped us to identify high affinity and selective glycopeptide binding candidates to two of the enzymes, GalNAc-T2 and -T4. GalNAc-T lectin binding partners were identified from a library presenting 11-mer peptides synthesized around a central serine-carrying GalNAc. Some general patterns for glycopeptide binding to GalNAc-T2 and -T4 lectins were recognized, although we were unable to extract a specific consensus motif for either of the two lectin domains. The identified glycopeptides contained a large proportion of hydrophobic amino acids, indicative of hydrophobic binding sites in the lectin domain. Furthermore, a high occurrence of β-branched Val/Ile/Thr was found. That Thr can be readily substituted by Val or Ile suggests that the identified peptides reflect lectin binding to peptide regions with high density of threonines, such as those in mucins. Furthermore, both the GalNAc-T2 and -T4 lectin domains were attracted to positively charged peptides carrying one or two arginines, particular in the +1 position upstream from the glycosylation site. This could be due to either general charge-charge interaction between the glycopeptides and the lectin domains leading to increased lectin concentration in the beads or to specific charge-charge interactions in the binding site.

Lectin domains may represent drug targets aimed at reducing efficiency of glycosylation of secreted mucins. Hyperglycosylated mucins are found in chronic inflammatory airway diseases such as asthma, chronic bronchitis, and cystic fibrosis, and inhibition of mucins and their glycosylation could make an important contribution toward controlling disease (57). Targeting the catalytic domains of GalNAc-Ts by UDP-based inhibitors may lead to general inhibition of all *O*-glycoprotein and mucin glycosylation resulting in broad effects on the secretory pathway, the cell membrane, and secretions. Targeting lectin domains selectively could result in a more subtle strategy, which selectively targets mucins and mucin-like proteins with regions of high density of *O*-glycans. This strategy is less likely to affect secretion and general functions of *O*-glycans, but will conceivably result in mucins with less dense *O*-glycosylation, which are more prone to proteolytic digestion and perhaps clearance. Inhibition of mucin biosynthesis could be provided by inhibition of specific GalNAc transferases including their lectin domains. GalNAc peptides selectively inhibiting specific GalNAc-Ts may be useful in combination with inhibitors targeting the catalytic domain. This could create high affinity multivalent and isotype-specific inhibitors of GalNAc-Ts.

In summary, this study demonstrated that GalNAc-T lectins exhibit selective GalNAc glycopeptide binding specificities, and that the pattern and sequence context of GalNAc residues on glycopeptides differentially influence binding properties among the different GalNAc-T isoforms. Thus, the GalNAc-T lectins appear to add additional complexity to the initiation step of *O*-glycosylation.

Acknowledgments—We express our gratitude for the exceptional laboratory support provided by Karin U. Hansen and Lotte Le Fevre Bram.

REFERENCES

1. Clausen, H., and Bennett, E. P. (1996) *Glycobiology* **6,** 635–646
2. Gerken, T. A., Ten Hagen, K. G., and Jamison, O. (2008) *Glycobiology* **18,** 861–870
3. Wandall, H. H., Hassan, H., 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) *J. Biol. Chem.* **272,** 23503–23514
4. Wojczyk, B. S., Stwora-Wojczyk, M. M., Hagen, F. K., Striepen, B., Hang,

H. C., Bertozzi, C. R., Roos, D. S., and Spitalnik, S. L. (2003) *Mol. Biochem. Parasitol.* **131,** 93–107
5. Heise, N., Singh, D., van der Wel, H., Sassi, S. O., Johnson, J. M., Feasley, C. L., Koeller, C. M., Previato, J. O., Mendonça-Previato, L., and West, C. M. (2009) *Glycobiology* **19,** 918–933
6. Schwientek, T., Bennett, E. P., Flores, C., Thacker, J., Hollmann, M., Reis, C. A., Behrens, J., Mandel, U., Keck, B., Schäfer, M. A., Haselmann, K., Zubarev, R., Roepstorff, P., Burchell, J. M., Taylor-Papadimitriou, J., Hollingsworth, M. A., and Clausen, H. (2002) *J. Biol. Chem.* **277,** 22623–22638
7. Mandel, U., Hassan, H., Therkildsen, M. H., Rygaard, J., Jakobsen, M. H., Juhl, B. R., Dabelsteen, E., and Clausen, H. (1999) *Glycobiology* **9,** 43–52
8. Tian, E., and Ten Hagen, K. G. (2006) *Glycobiology* **16,** 83–95
9. Ten Hagen, K. G., Fritz, T. A., and Tabak, L. A. (2003) *Glycobiology* **13,** 1R-16R
10. Tian, E., and Ten Hagen, K. G. (2007) *J. Biol. Chem.* **282,** 606–614
11. Zhang, L., Zhang, Y., and Hagen, K. G. (2008) *J. Biol. Chem.* **283,** 34076–34086
12. Tenno, M., Ohtsubo, K., Hagen, F. K., Ditto, D., Zarbock, A., Schaerli, P., von Andrian, U. H., Ley, K., Le, D., Tabak, L. A., and Marth, J. D. (2007) *Mol. Cell. Biol.* **27,** 8783–8796
13. Miwa, H. E., Gerken, T. A., Jamison, O., and Tabak, L. A. (2010) *J. Biol. Chem.* **285,** 1208–1219
14. Kato, K., Jeanneau, C., Tarp, M. A., Benet-Pagès, A., Lorenz-Depiereux, B., Bennett, E. P., Mandel, U., Strom, T. M., and Clausen, H. (2006) *J. Biol. Chem.* **281,** 18370–18377
15. Topaz, O., Shurman, D. L., Bergman, R., Indelman, M., Ratajczak, P., Mizrachi, M., Khamaysi, Z., Behar, D., Petronius, D., Friedman, V., Zelikovic, I., Raimer, S., Metzker, A., Richard, G., and Sprecher, E. (2004) *Nat. Genet.* **36,** 579–581
16. Guda, K., Moinova, H., He, J., Jamison, O., Ravi, L., Natale, L., Lutterbaugh, J., Lawrence, E., Lewis, S., Willson, J. K., Lowe, J. B., Wiesner, G. L., Parmigiani, G., Barnholtz-Sloan, J., Dawson, D. W., Velculescu, V. E., Kinzler, K. W., Papadopoulos, N., Vogelstein, B., Willis, J., Gerken, T. A., and Markowitz, S. D. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106,** 12921–12925
17. Wagner, K. W., Punnoose, E. A., Januario, T., Lawrence, D. A., Pitti, R. M., Lancaster, K., Lee, D., von Goetz, M., Yee, S. F., Totpal, K., Huw, L., Katta, V., Cavet, G., Hymowitz, S. G., Amler, L., and Ashkenazi, A. (2007) *Nat. Med.* **13,** 1070–1077
18. Herr, P., Korniychuk, G., Yamamoto, Y., Grubisic, K., and Oelgeschläger, M. (2008) *Development* **135,** 1813–1822
19. Fakhro, K. A., Choi, M., Ware, S. M., Belmont, J. W., Towbin, J. A., Lifton, R. P., Khokha, M. K., and Brueckner, M. (2011) *Proc. Natl. Acad. Sci. U.S.A.* **108,** 2915–2920
20. Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., Johansen, C. T., Fouchier, S. W., Isaacs, A., Peloso, G. M., Barbalic, M., Ricketts, S. L., Bis, J. C., Aulchenko, Y. S., Thorleifsson, G., Feitosa, M. F., Chambers, J., Orho-Melander, M., Melander, O., Johnson, T., Li, X., Guo, X., Li, M., Shin Cho, Y., Jin, Go, M., Jin Kim, Y., Lee, J. Y., Park, T., Kim, K., Sim, X., Twee-Hee Ong, R., Croteau-Chonka, D. C., Lange, L. A., Smith, J. D., Song, K., Hua Zhao, J., Yuan, X., Luan, J., Lamina, C., Ziegler, A., Zhang, W., Zee, R. Y., Wright, A. F., Witteman, J. C., Wilson, J. F., Willemsen, G., Wichmann, H. E., Whitfield, J. B., Waterworth, D. M., Wareham, N. J., Waeber, G., Vollenweider, P., Voight, B. F., Vitart, V., Uitterlinden, A. G., Uda, M., Tuomilehto, J., Thompson, J. R., Tanaka, T., Surakka, I., Stringham, H. M., Spector, T. D., Soranzo, N., Smit, J. H., Sinisalo, J., Silander, K., Sijbrands, E. J., Scuteri, A., Scott, J., Schlessinger, D., Sanna, S., Salomaa, V., Saharinen, J., Sabatti, C., Ruokonen, A., Rudan, I., Rose, L. M., Roberts, R., Rieder, M., Psaty, B. M., Pramstaller, P. P., Pichler, I., Perola, M., Penninx, B. W., Pedersen, N. L., Pattaro, C., Parker, A. N., Pare, G., Oostra, B. A., O'Donnell, C. J., Nieminen, M. S., Nickerson, D. A., Montgomery, G. W., Meitinger, T., McPherson, R., McCarthy, M. I., McArdle, W., Masson, D., Martin, N. G., Marroni, F., Mangino, M., Magnusson, P. K., Lucas, G., Luben, R., Loos, R. J., Lokki, M. L., Lettre, G., Langenberg, C., Launer, L. J., Lakatta, E. G., Laaksonen, R., Kyvik, K. O., Kronenberg, F., König, I. R., Khaw, K. T., Kaprio, J., Kaplan, L. M., Johansson, A., Jarvelin, M. R., Janssens, A. C., Ingelsson, E., Igl, W., Kees Hovingh, G., Hottenga, J. J., Hofman, A., Hicks, A. A., Hengstenberg, C., Heid, I. M., Hayward, C., Havulinna, A. S., Hastie, N. D., Harris, T. B., Haritunians, T., Hall, A. S., Gyllensten, U., Guiducci, C., Groop, L. C., Gonzalez, E., Gieger, C., Freimer, N. B., Ferrucci, L., Erdmann, J., Elliott, P., Ejebe, K. G., Döring, A., Dominiczak, A. F., Demissie, S., Deloukas, P., de Geus, E. J., de Faire, U., Crawford, G., Collins, F. S., Chen, Y. D., Caulfield, M. J., Campbell, H., Burtt, N. P., Bonnycastle, L. L., Boomsma, D. I., Boekholdt, S. M., Bergman, R. N., Barroso, I., Bandinelli, S., Ballantyne, C. M., Assimes, T. L., Quertermous, T., Altshuler, D., Seielstad, M., Wong, T. Y., Tai, E. S., Feranil, A. B., Kuzawa, C. W., Adair, L. S., Taylor, H. A., Jr., Borecki, I. B., Gabriel, S. B., Wilson, J. G., Holm, H., Thorsteinsdottir, U., Gudnason, V., Krauss, R. M., Mohlke, K. L., Ordovas, J. M., Munroe, P. B., Kooner, J. S., Tall, A. R., Hegele, R. A., Kastelein, J. J., Schadt, E. E., Rotter, J. I., Boerwinkle, E., Strachan, D. P., Mooser, V., Stefansson, K., Reilly, M. P., Samani, N. J., Schunkert, H., Cupples, L. A., Sandhu, M. S., Ridker, P. M., Rader, D. J., van Duijn, C. M., Peltonen, L., Abecasis, G. R., Boehnke, M., and Kathiresan, S. (2010) *Nature* **466,** 707–713
21. Schjoldager, K. T., Vester-Christensen, M. B., Bennett, E. P., Levery, S. B., Schwientek, T., Yin, W., Blixt, O., and Clausen, H. (2010) *J. Biol. Chem.* **285,** 36293–36303
22. Fritz, T. A., Hurley, J. H., Trinh, L. B., Shiloach, J., and Tabak, L. A. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101,** 15307–15312
23. Fritz, T. A., Raman, J., and Tabak, L. A. (2006) *J. Biol. Chem.* **281,** 8613–8619
24. Wandall, H. H., Irazoqui, F., Tarp, M. A., Bennett, E. P., Mandel, U., Takeuchi, H., Kato, K., Irimura, T., Suryanarayanan, G., Hollingsworth, M. A., and Clausen, H. (2007) *Glycobiology* **17,** 374–387
25. Kubota, T., Shiba, T., Sugioka, S., Furukawa, S., Sawaki, H., Kato, R., Wakatsuki, S., and Narimatsu, H. (2006) *J. Mol. Biol.* **359,** 708–727
26. Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2000) *J. Biol. Chem.* **275,** 38197–38205
27. Tenno, M., Kézdy, F. J., Elhammer, A. P., and Kurosaka, A. (2002) *Biochem. Biophys. Res. Commun.* **298,** 755–759
28. Imberty, A., Piller, V., Piller, F., and Breton, C. (1997) *Protein Eng.* **10,** 1353–1356
29. Raman, J., Fritz, T. A., Gerken, T. A., Jamison, O., Live, D., Liu, M., and Tabak, L. A. (2008) *J. Biol. Chem.* **283,** 22942–22951
30. Perrine, C. L., Ganguli, A., Wu, P., Bertozzi, C. R., Fritz, T. A., Raman, J., Tabak, L. A., and Gerken, T. A. (2009) *J. Biol. Chem.* **284,** 20387–20397
31. Pedersen, J. W., Blixt, O., Bennett, E. P., Tarp, M. A., Dar, I., Mandel, U., Poulsen, S. S., Pedersen, A. E., Rasmussen, S., Jess, P., Clausen, H., and Wandall, H. H. (2011) *Int. J. Cancer* **128,** 1860–1871
32. Wandall, H. H., Blixt, O., Tarp, M. A., Pedersen, J. W., Bennett, E. P., Mandel, U., Ragupathi, G., Livingston, P. O., Hollingsworth, M. A., Taylor-Papadimitriou, J., Burchell, J., and Clausen, H. (2010) *Cancer Res.* **70,** 1306–1313
33. White, T., Bennett, E. P., Takio, K., Sørensen, T., Bonding, N., and Clausen, H. (1995) *J. Biol. Chem.* **270,** 24156–24165
34. 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., Eiberg, H., Steffensen, R., and Clausen, H. (1998) *J. Biol. Chem.* **273,** 30472–30481
35. Mirgorodskaya, E., Hassan, H., Wandall, H. H., Clausen, H., and Roepstorff, P. (1999) *Anal. Biochem.* **269,** 54–65
36. Pedersen, J. W., Blixt, O., Bennett, E. P., Tarp, M. A., Dar, I., Mandel, U., Poulsen, S. S., Pedersen, A. E., Rasmussen, S., Jess, P., Clausen, H., and Wandall, H. H. (2011) *Int. J. Cancer* **128,** 1860–1871
37. Blixt, O., Cló, E., Nudelman, A. S., Sørensen, K. K., Clausen, T., Wandall, H. H., Livingston, P. O., Clausen, H., and Jensen, K. J. (2010) *J. Proteome Res.* **9,** 5250–5261
38. Meldal, M. (1992) *Tetrahedron Lett.* **33,** 3077–3080
39. Meldal, M. (1994) *Methods* **6,** 417–424
40. Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. (2007) *Nat. Methods* **4,** 709–712
41. Zhang, Y., Ficarro, S. B., Li, S., and Marto, J. A. (2009) *J. Am. Soc. Mass Spectrom.* **20,** 1425–1434
42. McAlister, G. C., Berggren, W. T., Griep-Raming, J., Horning, S., Makarov, A., Phanstiel, D., Stafford, G., Swaney, D. L., Syka, J. E., Zabrouskov, V., and

Coon, J. J. (2008) *J. Proteome Res.* **7,** 3127–3136
43. Swaney, D. L., McAlister, G. C., Wirtala, M., Schwartz, J. C., Syka, J. E., and Coon, J. J. (2007) *Anal. Chem.* **79,** 477–485
44. Sørensen, T., White, T., Wandall, H. H., Kristensen, A. K., Roepstorff, P., and Clausen, H. (1995) *J. Biol. Chem.* **270,** 24166–24173
45. Paroutis, P., Touret, N., and Grinstein, S. (2004) *Physiology* **19,** 207–215
46. Dupont, C., Roberge, M., Shareck, F., Morosoli, R., and Kluepfel, D. (1998) *Biochem. J.* **330**, 41–45
47. Fujimoto, Z., Kuno, A., Kaneko, S., Kobayashi, H., Kusakabe, I., and Mizuno, H. (2002) *J. Mol. Biol.* **316,** 65–78
48. Thobhani, S., Ember, B., Siriwardena, A., and Boons, G. J. (2003) *J. Am. Chem. Soc.* **125,** 7154–7155
49. Tenno, M., Saeki, A., Kézdy, F. J., Elhammer, A. P., and Kurosaka, A. (2002) *J. Biol. Chem.* **277,** 47088–47096
50. Sheehan, J. K., Kirkham, S., Howard, M., Woodman, P., Kutay, S., Brazeau, C., Buckley, J., and Thornton, D. J. (2004) *J. Biol. Chem.* **279,** 15698–15705
51. Collins, B. E., Blixt, O., Bovin, N. V., Danzer, C. P., Chui, D., Marth, J. D., Nitschke, L., and Paulson, J. C. (2002) *Glycobiology* **12,** 563–571
52. Iwasaki, H., Zhang, Y., Tachibana, K., Gotoh, M., Kikuchi, N., Kwon, Y. D., Togayachi, A., Kudo, T., Kubota, T., and Narimatsu, H. (2003) *J. Biol. Chem.* **278,** 5613–5621
53. Bertini, I., Calderone, V., Fragai, M., Jaiswal, R., Luchinat, C., Melikian, M., Mylonas, E., and Svergun, D. I. (2008) *J. Am. Chem. Soc.* **130,** 7011–7021
54. Gerken, T. A., Raman, J., Fritz, T. A., and Jamison, O. (2006) *J. Biol. Chem.* **281,** 32403–32416
55. Gerken, T. A., Jamison, O., Perrine, C. L., Collette, J. C., Moinova, H., Ravi, L., Markowitz, S. D., Shen, W., Patel, H., and Tabak, L. A. (2011) *J. Biol. Chem.* **286,** 14493–14507
56. Röttger, S., White, J., Wandall, H. H., Olivo, J. C., Stark, A., Bennett, E. P., Whitehouse, C., Berger, E. G., Clausen, H., and Nilsson, T. (1998) *J. Cell Sci.* **111**, 45–60
57. Voynow, J. A., and Rubin, B. K. (2009) *Chest* **135,** 505–512