



Exploring the glycosylation of mucins by use of O-glycodomain reporters recombinantly expressed in glycoengineered HEK293 cells

Received for publication, January 3, 2022, and in revised form, February 25, 2022. Published, Papers in Press, March 2, 2022.

<https://doi.org/10.1016/j.jbc.2022.101784>

Andriana Konstantinidi¹, Rebecca Nason¹, Tomislav Čaval¹, Lingbo Sun¹, Daniel M. Sørensen¹ , Sanae Furukawa¹, Zilu Ye¹ , Renaud Vincentelli², Yoshiki Narimatsu^{1,3}, Sergey Y. Vakhrushev^{1,*}, and Henrik Clausen^{1,*} 

From the ¹Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark; ²Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, Marseille, France; ³GlycoDisplay ApS, Copenhagen, Denmark

Edited by Gerald Hart

Mucins and glycoproteins with mucin-like regions contain densely O-glycosylated domains often found in tandem repeat (TR) sequences. These O-glycodomains have traditionally been difficult to characterize because of their resistance to proteolytic digestion, and knowledge of the precise positions of O-glycans is particularly limited for these regions. Here, we took advantage of a recently developed glycoengineered cell-based platform for the display and production of mucin TR reporters with custom-designed O-glycosylation to characterize O-glycodomains derived from mucins and mucin-like glycoproteins. We combined intact mass and bottom-up site-specific analysis for mapping O-glycosites in the mucins, MUC2, MUC20, MUC21, protein P-selectin-glycoprotein ligand 1, and proteoglycan syndecan-3. We found that all the potential Ser/Thr positions in these O-glycodomains were O-glycosylated when expressed in human embryonic kidney 293 SimpleCells (Tn-glycoform). Interestingly, we found that all potential Ser/Thr O-glycosites in TRs derived from secreted mucins and most glycosites from transmembrane mucins were almost fully occupied, whereas TRs from a subset of transmembrane mucins were less efficiently processed. We further used the mucin TR reporters to characterize cleavage sites of glycoproteases StcE (secreted protease of C1 esterase inhibitor from EHEC) and BT4244, revealing more restricted substrate specificities than previously reported. Finally, we conducted a bottom-up analysis of isolated ovine submaxillary mucin, which supported our findings that mucin TRs in general are efficiently O-glycosylated at all potential glycosites. This study provides insight into O-glycosylation of mucins and mucin-like domains, and the strategies developed open the field for wider analysis of native mucins.

Mucin-type (GalNAc-type) O-glycosylation is an abundant type of protein glycosylation initiated in the Golgi by a large family of up to 20 polypeptide GalNAc transferase (GALNT) isoenzymes with different kinetic properties and substrate specificities (1, 2). The repertoire of the GALNTs expressed in cells vary, and GalNAc-type glycosylation (hereafter, simply O-glycosylation) is therefore uniquely suited to differentially regulate the positions in proteins being glycosylated in cells (3). O-glycans are found on select Ser and Thr residues (and Tyr) often in clustered motifs with adjacent Pro residues, but no simple consensus sequence motifs have emerged (1, 2, 4, 5). Prediction algorithms for O-glycosylation such as NetOGlyc4.0 (<http://www.cbs.dtu.dk/services/NetOGlyc-4.0/>) (5) and GALNT isoform specific (IsoGlyP) (6) provide valuable tools. Advances in O-glycoproteomics employing genetic engineering for simplification of glycan structural heterogeneity (SimpleCells) (5, 7), improved and novel enrichment strategies (8–11), and enhanced sensitivity and speed of mass spectrometry (MS) (12) have advanced insights into O-glycosites in around 3000 human proteins trafficking the secretory pathway (13, 14). Paradoxically, the classes of proteins predicted to be the most heavily O-glycosylated, that is, mucins and glycoproteins with mucin-like domains comprised of high frequencies of Ser/Thr residues, are those with the least experimental evidence to support the positions where O-glycans are attached (14–21). This conundrum is likely primarily a result of available experimental strategies, where the main obstacle is limited options for proteolytic digestion of O-glycodomains into fragments suitable for MS sequencing because of a characteristic amino acid usage in the domains, generally without charged residues, and the high density of O-glycans (14, 15). Regardless, considerable experimental data for O-glycan positions exist for the human cell membrane mucin, MUC1 (22), the mucin-like lubricin/proteoglycan 4 (23), and fragments of the porcine and canine submaxillary mucins (17–19, 21).

At least 18 distinct human genes encode membrane or secreted mucins (24), and a larger number of genes encode proteins with mucin-like domains. Most mucins contain large O-glycodomains with variable number of more or less

* For correspondence: Henrik Clausen, hclau@sund.ku.dk; Sergey Y. Vakhrushev, seva@sund.ku.dk.

Present address for Lingbo Sun: Medical College of Yan'an University, Yan'an, 716,000, Shaanxi Province, China.

Present address for Zilu Ye: Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Blegdamsvej 3B, Copenhagen, Denmark.

Mucin glycosylation in glycoengineered HEK293 cells

conserved tandem repeat (TR) sequences (25–27), whereas O-glycodomains found in other proteins mainly do not. The classification of mucins may in this respect not be consistent, for example, the P-selectin-glycoprotein ligand 1 (PSGL-1) contains an O-glycodomain with characteristic TRs (28), whereas the cell membrane glycoprotein classified as MUC16 (CA125) has a very large N-terminal O-glycodomain without apparent TRs (29–31). Importantly, TRs in mucins are quite distinct in length and sequence with characteristic spacing of potential Ser/Thr O-glycosites, and these features diverge among closely related mammals (25, 32). We and others have proposed that mucin TRs as well as other O-glycodomains contain unique codes formed by O-glycan clusters and/or patterns that serve as recognition motifs for receptors (33, 34). In order to explore this, we recently developed a cell-based mucin TR array platform to display and produce small fragments (around 200 amino acids) of O-glycodomains derived from the characteristic TR regions (34). This platform relies on a library of stable gene-engineered human embryonic kidney 293 (HEK293) cells with distinct O-glycosylation capacities and expression of a panel of recombinant GFP-tagged mucin TR reporters either as cell membrane retained or secreted proteins. The display of mucin TRs on the cell surface provided a tool to demonstrate that human Siglecs and microbial Siglec-like adhesins appear to recognize their cognate O-glycan ligands with high selectivity for their presentation on mucin TRs (34–36), indicating that it is important to identify the actual O-glycan sites in mucin TRs. We therefore used expression of secreted mucin TR reporters to start characterizing O-glycosylation, and interestingly, we were able to analyze the entire O-glycodomains of several mucin TR reporters with the simplest glycoform (Tn, GalNAc α 1-O-Ser/Thr) by intact MS and demonstrate that most of the mucin reporters were O-glycosylated with exceptionally high fidelity and near complete occupancy of all potential Ser/Thr O-glycosites (34).

Intact MS analysis of glycoproteins has primarily been applied to abundantly available recombinant N-glycoprotein therapeutics including immunoglobulin G and erythropoietin (37–39). More recently, intact MS analysis was applied for direct profiling of human plasma N-glycoproteins, enabling insight into disease states (40, 41) as well as glycoprotein–drug (42) and/or glycoprotein–lectin (43) interactions. Heterogeneity in glycan structures attached to proteins constitutes the main obstacle for use of intact MS for analysis, and this is where genetic glycoengineering can be applied to obtain more homogenous glycoproteoforms (39, 44). Intact MS may be particularly suited for O-glycoproteins with dense O-glycodomains that are poorly accessible to commonly used bottom-up proteomics workflows (34), since intact MS can provide a holistic picture of intact O-glycoprotein macroheterogeneity where the total sum of attached O-glycans may be estimated. One illustrative recent example is the therapeutic chimeric tumor necrosis factor alpha receptor fusion protein (Etanercept) with multiple N-glycans and up to 26 O-glycans attached, where intact MS after removal of N-glycans was used to estimate O-glycan occupancy revealing the presence of 14

to 23 core1 O-glycans (38). However, because of the known issues of mass degeneracy between extended O-glycans and different O-glycan cores, it is very challenging to profile O-glycoproteins at both the microheterogeneity and macroheterogeneity levels if the glycoprotein in question carries more than one type of O-glycan structure (45).

Here, we extended our previous studies of mucin O-glycodomains to include a more comprehensive panel of mucin TR reporters derived from secreted and cell membrane human mucins as well as examples of mucin-like O-glycodomains. When possible, we combined intact MS analysis of excised O-glycodomain reporters with bottom-up analysis to characterize sites of O-glycosylation. Bottom-up site-specific analysis was performed with both select peptidases (Glu-C, trypsin, and Asp-N) as well as glycomucinas (StcE [secreted protease of C1 esterase inhibitor from EHEC], BT4244) (16, 46, 47), which provided further insights into the substrate specificities of bacterial glycoproteases. We begin to address regulation of sites of O-glycosylation in mucin TR domains by the repertoire of GALNTs as well as the elongation process and discovered that the elongation of O-glycans with the core3 structure adversely affects the O-glycan occupancy. Finally, we used the peptidases and glycomucinas for bottom-up analysis of ovine submaxillary mucin (OSM), which led to unambiguous identification of the gene and full coding sequence from a recent genome draft.

Results

The cell-based platform for production of mucin O-glycodomains with rather homogenous O-glycans opens up for detailed structural analysis as outlined in Figure 1 (34, 35). We designed secreted reporters containing TR O-glycodomains derived from a panel of secreted and membrane-bound human mucins as well as mucin-like domains from glycoproteins and expressed these in glycoengineered HEK293 isogenic cells selected to produce different O-glycan structures and densities of O-glycans. The glycoengineered cell lines used were developed previously, and for all genetic glycoengineering designs, multiple clones (usually two to four) were generated and characterized (35). Here, we used one representative clone for expression of the mucin reporters and subsequent analysis. We aimed for intact MS analysis of the isolated O-glycodomains and, when possible and relevant, bottom-up analysis to support interpretation of O-glycan occupancy and identification of O-glycosites (Fig. 1). The design of reporters in most cases enabled release of the C-terminal O-glycodomain by Lys-C digestion and following purification by reverse-phase (C4) HPLC direct analysis by intact MS. For several reporters, we were also able to perform bottom-up analysis by digestion with proteases (Glu-C, trypsin, and Asp-N) and/or glycomucinas (BT4244 and StcE).

Intact MS analysis of mucin TRs and mucin-like O-glycodomains carrying truncated Tn O-glycans

We used HEK293 SimpleCells with KO of either *COSMC* (HEK293^{KO} *COSMC*) or *CIGALT1* (HEK293^{KO} *CIGALT1*) encoding the private chaperone for the core1 synthase or the

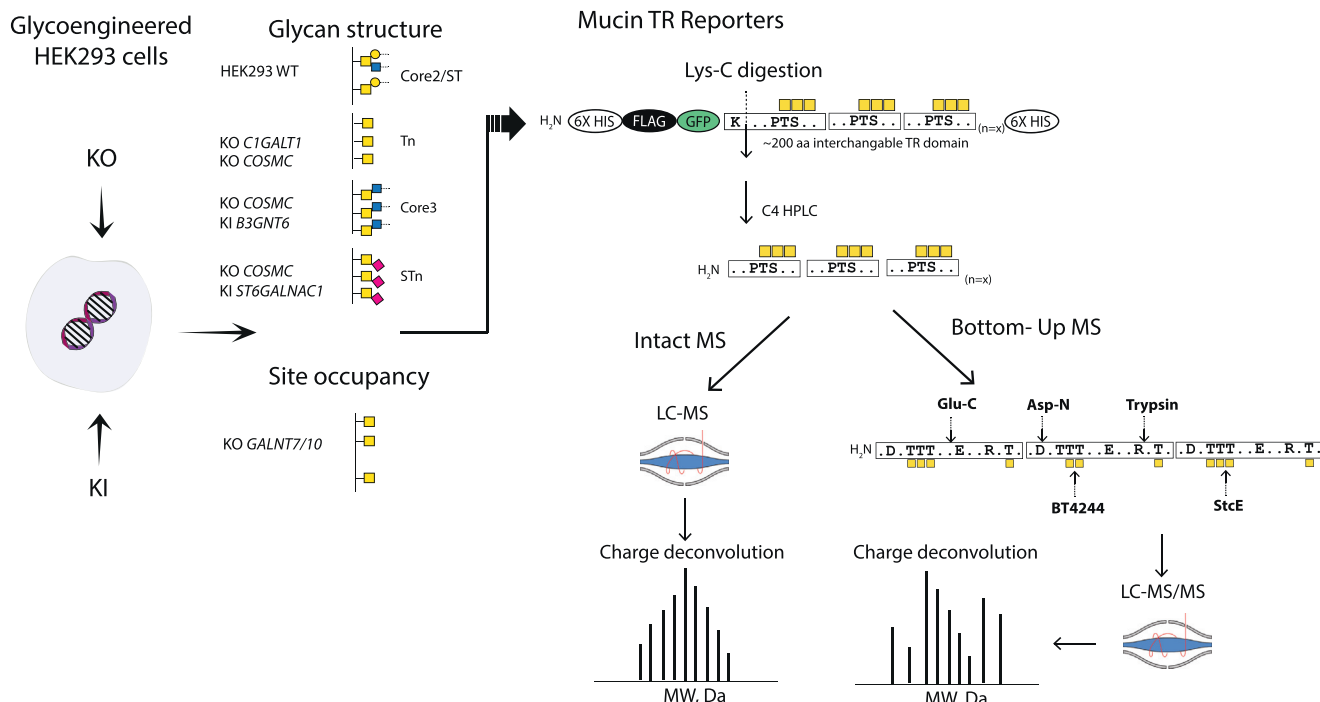


Figure 1. Overview of the cell-based production of mucin TRs and the analytic workflow. Mucin TR reporters were expressed in stably glycoengineered HEK293 cells to produce defined O-glycoforms, including Tn (KO *C1GALT1*), mStA (KO *GCNT1/ST6GALNAC2/3/4*), and core3 (KO *COSMC/KI B3GNT6*) O-glycosylation, and to produce glycoforms with different O-glycan occupancy (KO *GALNT4*; KO *GALNT7/10*) (left panel). The secreted mucin reporters contain N-terminal GFP, multiple tags (6xHis, FLAG tags), and the interchangeable mucin TR O-glycodomains (approximately 200 amino acids). Digestion with Lys-C results in release of the intact O-glycodomain without GFP (except with rare O-glycodomains containing internal Lys residues). This enables intact MS of the isolated O-glycodomains and/or bottom-up analysis after digestion with Glu-C, Asp-N, and trypsin (right panel). Glycan symbols are drawn according to the SNFG nomenclature (94). HEK293, human embryonic kidney 293 cell line; MS, mass spectrometry; SNFG, Symbol Nomenclature for Glycans; TR, tandem repeat.

synthase itself, respectively, resulting in mucin TR reporters with only the simplest O-glycan structure, GalNAc α 1-O-Ser/Thr, also designated Tn. We previously reported intact MS analysis of representative TR domains derived from several secreted (MUC2, MUC5AC, and MUC7) and transmembrane (MUC13 and MUC22) human mucins by the use of mucin TR reporters recombinantly expressed in glycoengineered HEK293 cells (34). Intact MS analysis of these revealed that most, if not all, of the 50 to 100 potential Ser/Thr glycosites were glycosylated (Table 1).

Here, we expanded the intact MS analysis to TRs derived from additional transmembrane mucins and O-glycoproteins with mucin-like domains (Fig. 2 and Table 1). The isolation and intact MS protocols were developed with the Tn-MUC1 glycodomain and demonstrated to yield reproducible results (34), and the intact MS analyses presented were in general performed once. Interestingly, the Tn O-glycan occupancy of select transmembrane mucins was considerably lower than the available Ser/Thr glycosites. In particular, the predominant glycoforms of MUC20 and MUC21 were predicted to have ~50% occupancy. The TR reporters for the transmembrane mucins, MUC1, MUC4, and the mucin-like reporters, PSGL-1 and syndecan-3 (SDC3), were predicted to have close to 100% occupancy, and the predominant glycoforms for MUC17 and MUC22 were predicted to have only slightly lower occupancy than the available total potential glycosites. Thus, the substantially lower O-glycan occupancy found for MUC20 and

MUC21 does not seem to relate to the expression of TRs from membrane-bound mucins in secreted TR reporters, although it should be noted that one study has shown differences in O-glycan processing of the membrane-bound mucin, MUC1, when expressed as a transmembrane protein compared with when expressed as a truncated secreted protein (48). More likely, the lower occupancy for the MUC20 and MUC21 TR domains is related to a greater diversity in amino acid usage and a higher use of Ser than Thr residues in these TRs compared with what is found in secreted mucin TRs. The sequence context of O-glycosites clearly affects O-glycosylation, and Ser O-glycosites are considered poorer acceptor substrates than Thr sites for GALNTs (49), and GalNAc residues attached to Ser and Thr residues attain distinct conformations (50). We are currently not able to analyze TR reporters expressed as membrane-bound proteins.

The TR reporter designs for several mucin O-glycodomains also included potential isolated N-glycosylation sites (MUC7, MUC13, MUC17, and MUC22) (Fig. S1). N-glycosylation consensus sites (NXS/T) are occasionally found in mucin TRs and in particular in cell membrane mucins and mucin-like domains, but whether these are utilized has not been evaluated to our knowledge. The MUC22 TR reporter includes the N-glycan sequon -ETTTNSTTSSE- which provided an opportunity to analyze this (Fig. S2). Intact MS analysis following PNGaseF treatment revealed a characteristic mass shift of ~2500 Da for all major glycoforms leaving a minor

Mucin glycosylation in glycoengineered HEK293 cells

Table 1

Overview of intact MS analysis of mucin TR reporters expressed in HEK293^{KO}C1GALNT1

Tn-mucin TR reporters	Representative mucin TR sequence and number of TRs included in reporter	Pot O-glycosites in TR reporters	Three most abundant HexNAc proteoforms identified ^a	Full range of HexNAc proteoforms identified ^a	Averaged HexNAc incorporated per TR ^b (exp/pot)
MUC2 TR1 ^c	PSPPITTTTTPPPPTT _{10x}	86	79–81	73–86	9/9
MUC2 TR2 ^c	GTQTPPTPITTTTTPPTPT _{7x}	89	87–89	85–90	13/13
MUC5AC ^c	STTSAPTT _{18x}	103	101–103	98–105	6/6
MUC7 ^c	TTAVPPTPSATLDPSSASAPPE _{7x}	67	62–64	58–67	9/9
MUC1 ^c	APDTRPAPGSTAPPAHGVT _{7x}	34	32–34	28–35	5/5
MUC4	PLPVTDTSSASTGHAT _{9x}	65	60–62	49–66	7/7
MUC13 ^c	TSDIITASSPNDGLIT _{9x}	58	51–53	48–55	6/6
MUC17	TSTPSEGSTPFTSMPVSTMPVVTSEAST _{5x}	73	62–64	42–70	13/14
MUC20	SESSASSDGPVITPSRA _{8x}	56	26–28	10–42	3/7
MUC21	SSGASTATNSESTV _{10x}	91	45–47	30–65	5/9
MUC22 ^c	SETTIVTSTAG _{15x}	83	68–71	59–75	5/6
SDC3	^d	52	48–50	45–52	^d
PSGL-1	QTTQPAATEA _{14x}	47	38–40	37–40	3/3

^a HexNAc residues identified by intact MS.

^b Comparison of the experimental (exp) number of HexNAc residues identified and the number of potential (pot) O-glycosites per TR. The values are averaged (Avr) numbers based on most abundant number of HexNAc residues and the number of Ser/Thr O-glycosites available in the most common imperfect TR sequence.

^c Previously reported (34).

^d SDC3 does not contain TRs.

group of glycoforms (*m/z* 32,250–34,000) unchanged (Fig. S2), indicating that most of the TR glycoforms indeed contained a complex-type N-glycan.

Intact MS analysis to probe the effects of the GALNT repertoire on O-glycan occupancy

The initiation step of O-glycosylation is controlled by multiple GALNTs with distinct and partly overlapping acceptor substrate preferences and kinetic properties, and the repertoire of expressed GALNTs in cells vary although several isoenzymes including GALNT1 and GALNT2 are rather ubiquitously expressed (1). Recent *in vitro* and *in cell* studies of the acceptor substrate specificities of GALNTs with detectable activity have demonstrated that these have considerable overlapping functions and only limited nonredundant substrate sites (4, 51–53). For example, analysis of HEK293 cells with individual KO of *GALNT1*, *T2*, *T3*, *T4*, *T7*, and *T10* genes revealed that the nonredundant contributions of these isoenzymes to the cellular O-glycoproteome are quite limited (51). Studies of O-glycosylation of mucin TRs are mainly limited to *in vitro* studies with peptide substrates (54), and these indicate that the typical Pro-Thr-Ser-rich sequences are broad substrates for most GALNTs (1, 6, 55). The so-called follow-up GALNTs (*GALNT4*, *T7*, *T10*, *T12*, and *T17*) selectively serve prior GalNAc-glycosylated substrates (56–60), and it is predicted that these are especially important for mucin TR substrates with dense and clustered glycosites (51). Previously, we showed that the close paralogs *GALNT7/10* are important for glycosylation of dense O-glycodomains (51, 61), and we therefore predicted that dissection of the function of these would be particularly interesting with the mucin TR reporters.

Here, we used intact MS to analyze O-glycans occupancy of mucin TR reporters expressed in HEK293^{KO}C1GALNT1 cells with additional double KO of *GALNT7* and *T10*. Four of the mucin TRs were not substantially affected by the loss of *GALNT7/T10* (Fig. S3), whereas the MUC1, MUC5AC, MUC13, and MUC22 TR reporters showed distinct shifts in mass range corresponding to loss of approximately one HexNAc residue per number of TRs included in reporters (Fig. 3). The lower

number of HexNAcs incorporated in the latter TRs suggested that *GALNT7* and *T10* serve nonredundant functions. We chose to investigate the lower occupancy found for the MUC1 TRs, since this TR is quite conserved and amenable to Asp-N digestion and bottom-up LC-MS/MS analysis, as previously described (62). The *GALNT* follow-up process was originally discovered with the MUC1 TR substrate finding that only *GALNT4* glycosylated two of the five glycosites (Ser in VTSA and Thr in PDTR) in the MUC1 TR and only following prior addition of GalNAc residues by other *GALNTs* (56, 63). However, we recently found that *GALNT4* apparently only directs O-glycosylation at Thr in PDTR in HEK293 cells by analysis of the MUC1 TR reporter expressed in HEK293^{KO}COSMIC/*GALNT4* cells (34). Since KO of *GALNT7/10* resulted in loss of six to seven GalNAc residues for the MUC1 reporter (approximately one per TR), we predicted that these isoforms selectively served the Ser residue in VTSA not glycosylated by *GALNT4* in HEK293 cells. Surprisingly, loss of *GALNT7/T10* did not appear to affect any specific sites as all identified major species contained GalNAc residues at all five sites and all glycoforms contained GalNAc residues at Thr in PDTR and GSTA (Fig. S4). Since the TRs of MUC5AC, MUC13, and MUC22 are less conserved and without obvious digestion strategies for bottom-up analysis, we did not analyze these further.

Interestingly, we observed minor glycoproteoforms with a predicted number of HexNAc residues in excess of the total number of available potential Ser/Thr O-glycosites in the O-glycodomain reporters (Fig. 2), in agreement with our previous study (34). The basis for this is still unclear, but when the corresponding bottom-up analysis of the TR reporters was performed, this did not show evidence of excess HexNAc incorporation. One possible explanation may be a very minor incorporation of HexNAc₂ disaccharides by the *GALNTs* themselves, which when accumulated over the 50 to 100 glycosites produces a small visible mass with a single excess HexNAc. The existence of GalNAcα1–3GalNAcα1–O-Ser/Thr O-glycans has been suggested in human meconium (64).

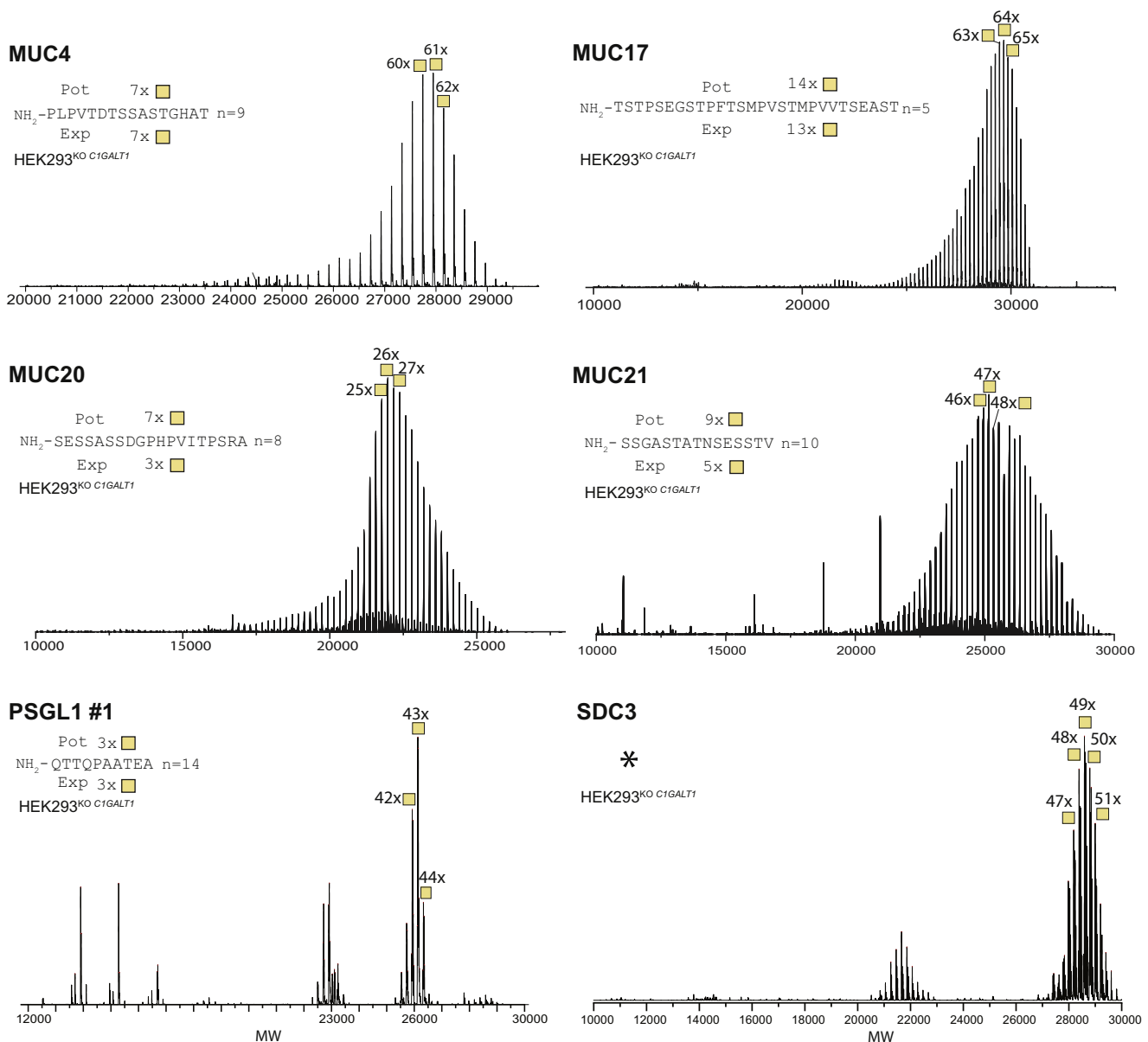


Figure 2. Intact MS analysis of O-glycodomains isolated from mucin TR reporters with Tn O-glycans. Deconvoluted intact mass spectra of isolated O-glycodomains from MUC4, MUC17, MUC20, MUC21, PSGL-1, and SDC3 mucin TR reporters expressed in HEK293^{KO} C1GALT1 cells. The three most abundant masses are annotated with the predicted number of attached HexNAc residues. A representative TR sequence is shown with the number of total potential (Pot) O-glycosylation sites (Ser/Thr residues) and the experimentally (Exp) predicted average number of HexNAc residues per TR. SDC3 does not contain TRs, and the full sequence of the mucin-like domain is shown in Fig. S1. HEK293, human embryonic kidney 293 cell line; MS, mass spectrometry; PSGL-1, P-selectin-glycoprotein ligand 1; SDC3, syndecan-3; TR, tandem repeat.

Intact MS to probe the effect of O-glycan elongation on occupancy

While intact MS analysis of the Tn-glycoforms of several mucin TR reporters was successful, we were unable to obtain interpretable results with more complex glycoforms including STn and T. The only exception was the MUC1 TR, although only after removal of sialic acids by neuraminidase treatment (34). Previously, we used intact MS analysis of the MUC1 TR reporter expressed in HEK293 cells with engineered glycosylation capacities limited to Tn, STn, T, or ST O-glycosylation to demonstrate that the core1 (T/ST) glycosylation pathway did not substantially affect the number of O-glycans attached (34). Interestingly, site-directed knockin (KI) of *ST6GALNAC1* in

HEK293^{KO} COSMC cells to introduce STn also did not substantially affect initiation, in contrast to what was previously found with overexpression (65). Previous studies demonstrated that overexpression of ST6GALNAC1 in cell lines interferes with and overrides normal glycosylation leading to truncated STn O-glycans (65, 66), although potential effect on O-glycan occupancy was not investigated. Here, we were able to extend this to the core3 O-glycosylation pathway directed by B3GNT6 (67). B3GNT6 is only expressed in the normal gastrointestinal tract and downregulated in cancer, and interestingly, B3GNT6 is not expressed in common cancer cell lines (68). Expression of the MUC1 reporter in HEK293^{KO} COSMC, KI B3GNT6 and analysis of the O-glycodymain by intact MS indicated a substantial

Mucin glycosylation in glycoengineered HEK293 cells

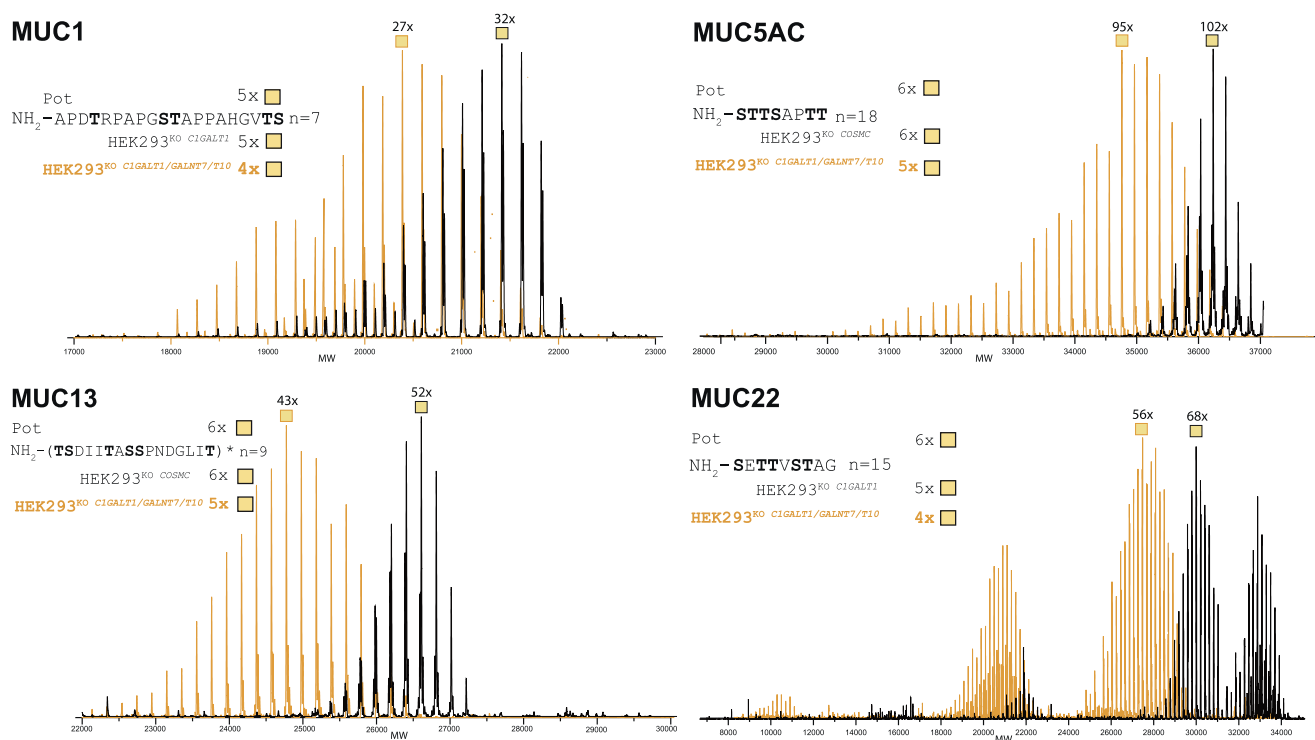


Figure 3. Intact MS analysis of O-glycodomains isolated from mucin TR reporters with altered O-glycan occupancy. Overlay of deconvoluted intact mass spectra of mucin TR reporters produced either in HEK293^{KO} C1GALTI or in HEK293^{KO} COSMC (black contour) and in HEK293^{KO} C1GALTI/GALNT7/T10 (orange contour). The most abundant masses are annotated with predicted number of HexNAc residues. A representative TR sequence is shown with indicated total potential glycosylated Ser/Thr residues (Pot) and the experimentally determined average number of HexNAcs found per each TR domain (HEK293^{KO} C1GALTI or COSMC, HEK293^{KO} C1GALTI/GALNT7/T10). Relative abundances, deconvoluted masses, annotation, and theoretical masses of all peaks above 5% intensity are given in Supporting File S1. For MUC5AC and MUC13 TR reporters, we used the same raw files from previous work (34), but we modified the deconvolution parameters (see Data analysis in the Experimental procedures section). HEK293, human embryonic kidney 293 cell line; MS, mass spectrometry; TR, tandem repeat.

reduction in the number of HexHexNAc₂ O-glycans incorporated (Fig. 4A), predicted from the dominant 32 to 34 to 19 to 21 with a wider range of glycoproteoforms. This interpretation was confirmed by bottom-up analysis revealing rather homogeneous core3 O-glycan structures and demonstrating that the loss of total incorporated O-glycans was due to select loss of O-glycans at Ser in GVTS, Ser in GSTA, and partly Thr in PDTR (Figs. 4B and S5). Since these sites are predicted to be glycosylated through lectin-mediated properties of GALNTs (56, 69), this suggests that the core3 B3GNT6 synthase, in contrast to the core1 C1GALTI synthase, competes with the GALNT follow-up reaction. The initiation and elongation of O-glycosylation takes place in common Golgi compartments, and the elongation process has the potential to compete and interfere with the initiation step orchestrated by GALNTs. Specifically, the initiation step by GALNTs involves so-called follow-up reactions where GALNTs utilize their lectin domains to bind prior incorporated GalNAc residues and efficiently complete glycosylation of adjacent glycosites, and premature elongation of these initial GalNAc residues is predicted to block lectin recognition and efficient glycosylation (69).

Interestingly, we found evidence of residual core1 O-glycans on the MUC1 reporter expressed in HEK293^{KO} COSMC, KI B3GNT6 (Fig. 4B). We were unable to quantify the exact levels, but the core1 structure was found at Thr in PDTR and/or Ser in GSTA, and we also identified glycoforms with a single HexNAc (Tn) at these sites. Based on ELISA assays with lectins and antibodies,

the majority of the O-glycans appears to represent core3 O-glycans since core1 (*Arachis hypogaea* agglutinin [PNA]) and Tn (*Vicia villosa* agglutinin [VVA]) were not or only barely detectable, respectively (34). While the ELISA results are semiquantitative, the results fully support the bottom-up MS analysis. The presence of minor levels of core1 structures is likely because of residual core1 C1GALTI synthase activity, since the chaperone COSMC was KO for this experiment. We originally used KO of the COSMC gene to eliminate core1 elongation in cells (7), but we have noticed the presence of minor levels of residual core1 O-glycopeptides when using Jacalin enrichment of O-glycopeptides instead of the original VVA enrichment. The minor levels of core1 may be due to partial folding of the core1 synthase C1GALTI in the absence of its private COSMC chaperone or the existence of other chaperones. To circumvent this issue, we have subsequently targeted the C1GALTI gene and used HEK293^{KO} C1GALTI cells.

The observed slightly incomplete galactosylation of the core3 disaccharide (HexNAc₂) may be due to HEK293 cells only expressing B4GALTs and not B3GALTs, which may be a preferred pathway for core3 (35).

Bottom-up MS of mucin TRs and O-glycodomains using Glu-C and trypsin

Most mucin TRs are not amenable for protease digestion and bottom-up MS analysis, but TRs in select transmembrane mucins, including MUC20 and MUC21, contain conserved

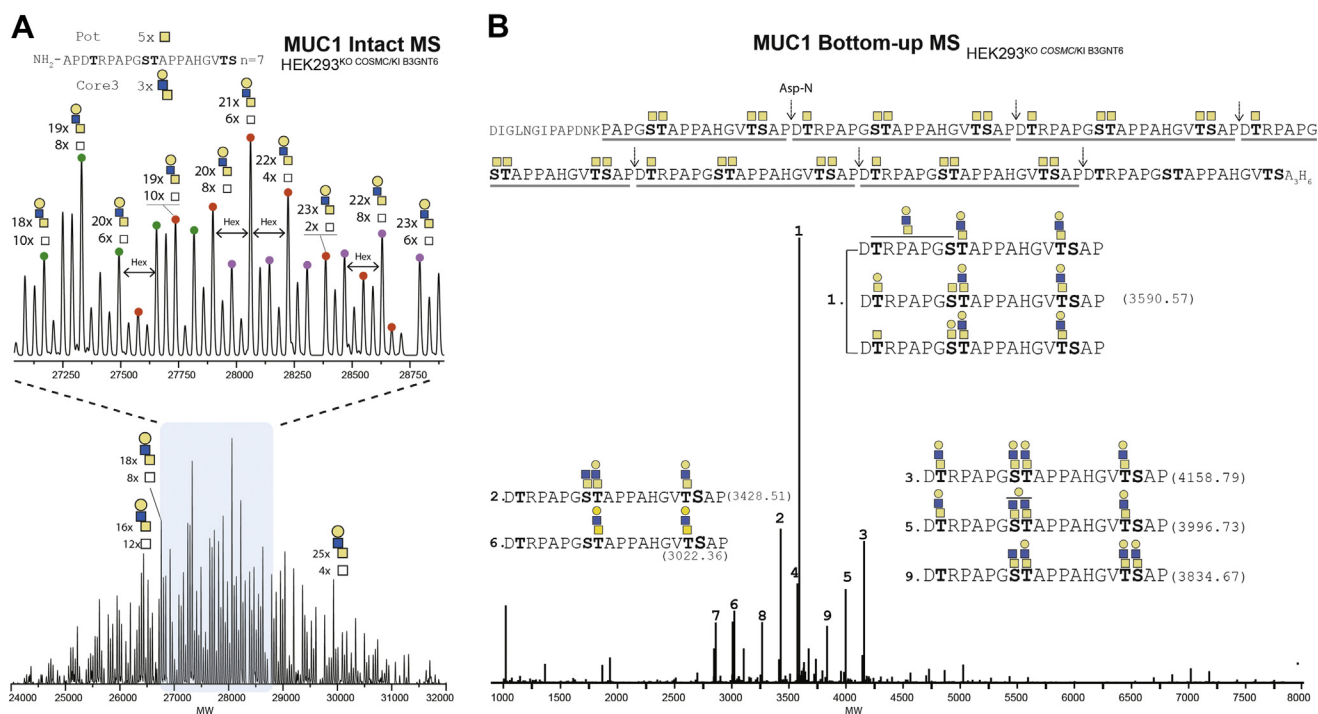


Figure 4. Intact and bottom-up MS analysis of the isolated MUC1 O-glycodomain with core3 O-glycans. A, deconvoluted spectrum of intact MS analysis of the isolated TR O-glycodomain from the MUC1 reporter expressed in HEK293^{KO} COSMIC KI B3GN16. The most abundant masses are annotated with predicted number of HexHexNAc₂ residues. A representative TR sequence is shown with indicated total potential (Pot) O-glycosylation sites (Ser/Thr residues) and the experimentally (Exp) observed average number of HexNAc residues per TR. B, deconvoluted spectrum of bottom-up analysis of the same MUC1 TR O-glycodomain with core3 O-glycans. The full MUC1 TR O-glycodomain sequence shows observed fragments (underlined) and Asp-N cleavage sites (*arrow*). **Bold S/T letters** represent unambiguously annotated glycosites (full sequence), and potential glycosites (peptides) and ambiguous sites are indicated by a *line*. The numbers assigned to each peak from one to nine are given based on decreasing abundance. Experimental masses in dalton for each assigned peak are provided in [Supporting File S1](#). Only peaks with intensity above 10% were assigned. HEK293, human embryonic kidney 293 cell line; MS, mass spectrometry; TR, tandem repeat.

Arg or Glu residues that may be digested by trypsin and Glu-C, respectively (Fig. 1). We therefore analyzed the MUC20 and MUC21 TR reporters expressed in HEK293^{KO} CIGALTI and predicted by the intact MS analysis to have low O-glycan occupancy (Fig. 2). Cumulatively, all potential Ser/Thr glycosites were found with an O-glycan in the identified glycopeptides; however, the individual glycopeptides identified contained GalNAc residues placed in different glycosite combinations, and there was no evidence of apparent preference for Thr or Ser glycosites (Fig. S6). Previous *in vitro* GALNT enzyme studies have demonstrated that the order of incorporation of GalNAc residues at different glycosites in peptides can affect the subsequent incorporations and hence generate different, mutually exclusive glycosylation patterns (70), but the basis for the lower occupancy observed is unknown.

The TR O-glycodomain of PSGL-1 was also amenable for bottom-up analysis using Glu-C, and as shown from the intact MS analysis (Fig. 2), all potential glycosites were found to be occupied (Fig. 5A). We were also able to perform bottom-up analysis of SDC3 with sequential treatment of trypsin and Glu-C to demonstrate that essentially all glycosites were utilized (Fig. 5B). Interestingly, this correlates well with the current accumulated information of utilized O-glycosites in SDC3 as summarized in the GlycoDomainViewer (13) as well as with the results of the intact mass analysis of SDC3 (Fig. 2).

Bottom-up MS analysis of mucin TRs using the glycomucinases StcE and BT4244

Glycomucinasases may offer unique opportunities for use in bottom-up analysis of mucin TRs and other dense O-glycodomains (16, 71); however, they may digest TRs into fragments that are challenging to identify and/or challenging to place in sequence context, and detailed knowledge of the cleavage specificities of these enzymes is still limited (72). StcE is a zinc metalloprotease secreted from the human pathogenic enterohemorrhagic *Escherichia coli* (73). We previously showed that StcE efficiently cleaves the MUC2 and MUC5AC TR reporters (34), and here, we chose to analyze the products generated by StcE cleavage of the MUC2 TR2 reporter with Tn O-glycans (Figs. 6 and S7). We identified Tn-peptides covering most of the sequence and fully confirmed the high occupancy of O-glycosylation demonstrated by intact MS. This also revealed that StcE primarily cleaved in the TTT and TGT sequons and did not cleave in TPT, TQT, and TVT sequons. StcE was proposed to cleave in the S/TX↓S/T motif with an obligate O-glycan at P2 (16).

We also used the glycomucinasase BT4244 suggested to cleave N-terminal to Ser or Thr residues with Tn or T O-glycans attached (46, 47). BT4244 efficiently cleaved the Tn MUC1 reporter, and the predominant cleavage sites identified were in between the diad O-glycans in the GSTA and VTSA

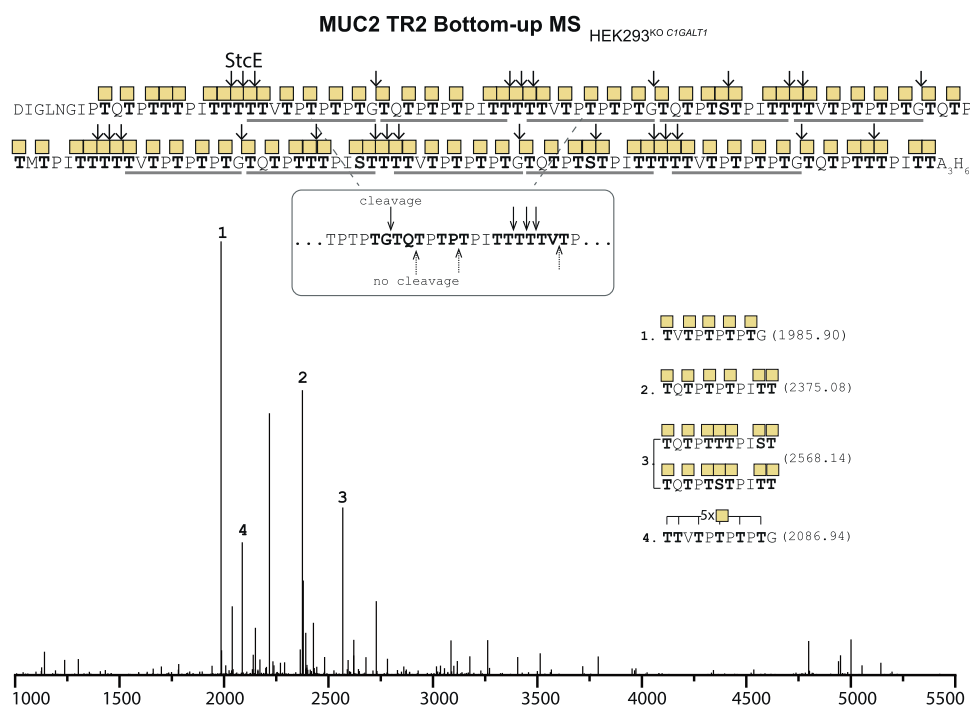


Figure 6. Bottom-up analysis of O-glycodomains with Tn O-glycans using StcE. Deconvoluted spectrum of bottom-up analysis of isolated MUC2 TR2 O-glycodomain from the reporter expressed in HEK293^{KO C1GALT1}. Numbers assigned to each peak from one to four are given based on decreasing abundance. Experimental masses in dalton for each assigned peak are provided in Supporting File S1. Only peaks with intensity above 10% were assigned. HEK293, human embryonic kidney 293 cell line; TR, tandem repeat.

motifs, whereas no cleavage was found in the single PDTR O-glycosite (Figs. 7 and S8). One identified glycopeptide indicated cleavage N-terminal to Ser in the GSTA motif. BT4244 digestion of the PSGL-1 reporter revealed cleavage in between the diad O-glycans in the TT motif, as well as cleavage N-terminal to the single Thr in TEA, and occasionally cleavage N-terminal to the first Thr in QTT (Fig. 7B). While StcE and BT4244 digestion provided useful information on select mucin TRs as shown, it is also clear that their digestion patterns are complex and challenging to analyze, and careful selection of substrates is needed to obtain useful information.

Applying proteases and glycomucinas to characterize OSM

OSM was originally isolated and characterized by Bhargava and Gottschalk (74) and Tettamanti and Pigman (75), and three tryptic peptides were sequenced by Hill *et al.* (76, 77). The OSM gene to our knowledge has not been reported, but the peptide sequences could be identified by BLAST sequence search in a large gene (accession number: XP_027821751) with a putative 333 amino-acid Pro-Thr-Ser-rich TR predicted to represent the OSM gene (Fig. S9). OSM is a widely used isolated mucin that has rather homogenous STn (NeuAc α 2-6-GalNAc α 1-O-Ser/Thr) O-glycans that after neuraminidase treatment is converted to asialo-OSM (AOSM) with Tn O-glycans (78, 79). We reproduced trypsin digestion of AOSM and identified five GalNAc-glycopeptides present in the putative TRs of the OSM gene (Fig. 8B). With StcE and BT4244 digestion, we identified additional GalNAc-glycopeptides (Fig. 8, C and D) that partly overlapped with the peptide sequences originally reported (76). We only identified two

peptides outside the predicted TR region of OSM (Table 2). Interestingly, and in agreement with our studies of the mucin TR reporters, all identified glycopeptides had essentially complete O-glycan occupancy at all possible Ser/Thr residues.

Discussion

The densely O-glycosylated regions of mucin TRs and mucin-like domains remain an analytical challenge to overcome, and the approaches taken here provide strategies and advances towards overcoming this. First of all, our studies of recombinantly expressed reporters for mucin TRs and the isolated OSM mucin suggest that essentially all Ser/Thr residues available in mucin TRs are O-glycosylated and with near complete occupancy, with the notable exception of lower occupancy in a few transmembrane mucin TRs with distinct TR sequences. Furthermore, our preliminary studies of mucin TRs produced in cells with altered repertoire of expressed GALNTs confirm that the repertoire of GALNTs available is important, but given that only minor effects were observed with KO of *GALNT7/10*, these studies also indicate that the O-glycosylation of mucin TRs appears to be covered by considerable functional redundancies among GALNTs. Thus, changes in expression of individual GALNTs are probably less likely to affect the glycosylation of mucin TRs in contrary to what has been found for few select O-glycoproteins and more isolated glycosites (4, 51–53). These results suggest that mucins in contrast to the prevailing prediction may in fact be rather homogeneous molecules, at least with respect to the O-glycan occupancy. Our studies also provided further validation for the use of the glycoengineering and cell-based platform for

Mucin glycosylation in glycoengineered HEK293 cells

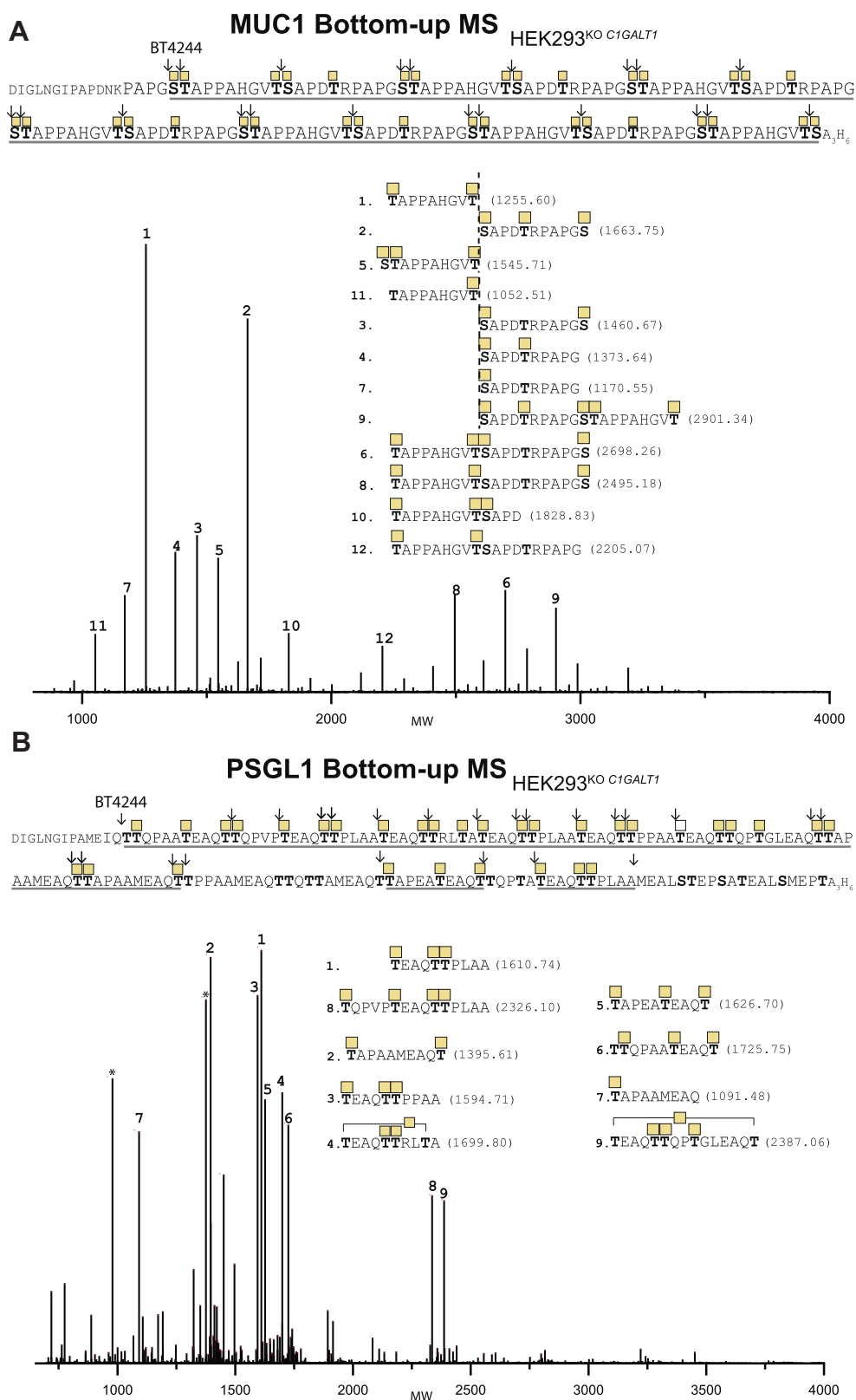


Figure 7. Bottom-up analysis of O-glycodomains with Tn O-glycans using BT4244. A, deconvoluted spectrum of the bottom-up analysis of isolated MUC1 TR O-glycodomains from reporters expressed in HEK293^{KO} C1GALT1. B, spectrum for PSGL-1. Numbers assigned to each peak are based on decreasing abundance. Experimental masses in dalton for each assigned peak are provided in [Supporting File S1](#). Only peaks with intensity above 10% were assigned. HEK293, human embryonic kidney 293 cell line; PSGL-1, P-selectin-glycoprotein ligand 1; TR, tandem repeat.

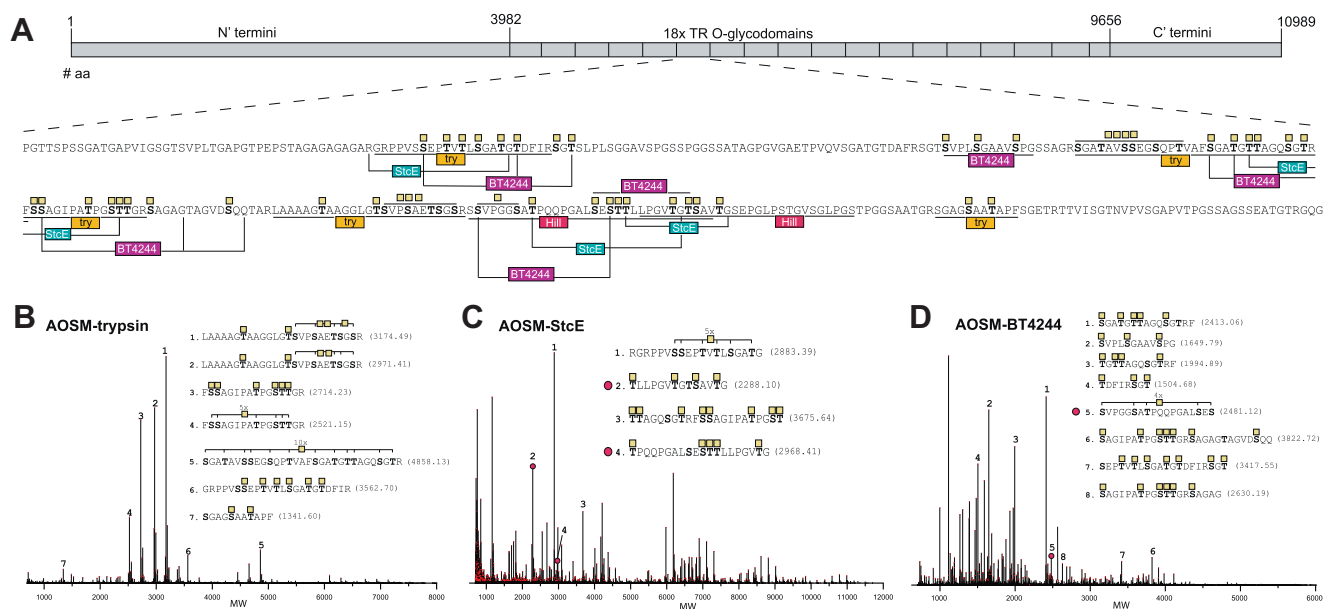


Figure 8. Bottom-up analysis of OSM using trypsin, StcE, and BT4244. A, schematic depiction of the OSM mucin and a representative TR domain (333 amino acids) with a summary of identified glycopeptides indicated. Two peptides originally identified by Hill *et al.* (76, 77) by trypsin digestion of deglycosylated OSM are indicated (*magenta*), and a third peptide (STTQLPGVTGTSAVTGSEPLPSTGVSLPGS) is found in a variable repeat at the C-terminal junction not shown. B–D, deconvoluted spectra of bottom-up analysis of AOSM with homogenous Tn O-glycans digested with *trypsin* (B), *StcE* (C), and *BT4244* (D). Numbers assigned to peaks are based on decreasing abundance. Experimental masses in dalton for each peak are provided in [Supporting File S1](#). Only peaks with intensity above 10% were assigned. AOSM, asialo-OSM; OSM, ovine submaxillary mucin; TR, tandem repeat.

production of mucin O-glycodomains with authentic glycosylation. This is especially important given the scarcity of natural human mucins and difficulties associated with their isolation, characterization, and consistency. Finally, we provided strategies for site-specific analysis of mucin TRs using rational selection and combination of proteases and glyco-proteases, which opens up for more detailed studies of isolated mucins.

Mucins are generally considered highly heterogeneous molecules, and variations may stem from the protein backbone, caused by genetic variations in numbers of TRs encoded by gene alleles, alternative splicing, and degradation (25–27, 80), but the main heterogeneity is predicted to arise from the non-template-driven O-glycosylation process resulting in great diversity in structures and positions of O-glycans (1, 81, 82). However, the notion of great heterogeneity in O-glycosylation of mucins may at least partly be ascribed to difficulties with isolation and characterization of natural mucins from homogeneous sources. The large size of most secreted mucins has largely prevented recombinant expression and analysis from

more homogeneous cell sources. Studies of recombinant MUC1 have shown that O-glycosylation of the five sites in the MUC1 TRs is well occupied (83), although one study suggested that overexpression of GALNT4 was required for efficient glycosylation at the Ser in VTSA motifs (84). Moreover, in the original analysis of isolated OSM, quantitation of the total Ser/Thr and HexNAc content suggested that this native mucin had nearly complete O-glycan occupancy (77). Similarly, studies from Gerken *et al.* (17–19, 21) showed that TRs from porcine submaxillary mucin and canine submaxillary mucin were highly O-glycosylated. Most studies of O-glycans found on mucins use profiling of released O-glycans from isolated mucins derived from large and heterogeneous cell and tissue sources. While these have shown great heterogeneity in the identified structures, these studies do not enable interpretation of the fidelity of O-glycosylation processes in a particular cell. Our studies of mucin TR reporters expressed in glycoengineered HEK293 cells suggest that these can in fact be produced with rather homogeneous O-glycan sites and structures (34).

Table 2
List of (glyco)peptides identified in the non-TR regions of OSM

Position in protein	Amino acid number in protein	Enzyme used	Sequence	Mass (Da)
N-terminal	933–946	Trypsin	AEDDFmSSQNILEK	1641.72
N-terminal	636–644	Trypsin	VSTLSSDYK	998.49
C-terminal	10,138–10,154	Trypsin	ATISGSSHTEATTLIAR	2730.29
C-terminal	10,009–10,019	Trypsin	LGTTVSTDLGLK	1496.75
N/C-terminal	1763–1777/10,282–10,296	StcE	TAGSVGTTGLAGPTF	2351.07
C-terminal	10,642–10,669	StcE	TDFIRSGTRFPVSGGAVSPGSSPGSSA	4261.92
N/C-terminal	1800–1811/10,319–10,330	BT4244	GSTGDTGFRAGG	1284.57
N-terminal	1779–1794	BT4244	SSGRISGSTGVSVSAV	2668.23
C-terminal	10,778–10,789	BT4244	SAAAGTAAGGLG	1308.61

Mucin glycosylation in glycoengineered HEK293 cells

Interestingly, we found that the core3 O-glycosylation pathway may interfere with the initiation process and occupancy of O-glycans at least in the case of the MUC1 TRs (Fig. 4). This is likely because of premature extension of initial GalNAc O-glycans by β 3GlcNAc residues in competition with the follow-up GALNTs that use their C-terminal GalNAc-binding lectins for efficient glycosylation of substrate sites adjacent to initial attached O-glycans (56, 60, 69, 85). Importantly, the core3 trisaccharide structure (Gal β 1-4-GlcNAc β 1-3GalNAc α 1-O-Ser/Thr) on the MUC1 TR reporter was also found to be assembled with high fidelity (Fig. 4).

So far, we have only been able to apply intact MS analysis with the more simple O-glycosylated TR reporters and mainly after removal of sialic acids, and thus, further improvements are needed. In addition, the issues of mass degeneracy preclude confident annotation of O-glycoforms of reporters expressing more than one type of O-glycan structures. However, intact measurements avoid known ionization bias of glycopeptides *versus* peptides, resulting in an underestimation of glycosite occupancy (45, 86). Therefore, a combined approach with intact MS for quantitative assessment of O-glycosylation landscape and bottom-up glycopeptide analysis for O-glycosite identification represents the most promising future avenue to unravel the complexities of mucin and mucin-like domain O-glycoproteins.

The bottom-up analytical strategies for mucin TRs performed here are not universal and ideally involve prior knowledge of the amino acid sequences to design the optimal combination of proteases and glycoproteases (Fig. 1). Traditional proteases are likely only useful for select transmembrane mucins and mucin-like domains as these have greater frequencies of amino acids in typical cleavage sites (Arg, Lys, Glu, and Asp), and it may be expected that adjacent O-glycans interfere with cleavage in variable ways depending on size and sialylation of the O-glycans. Conversely, the bacterial glycoproteases and glycomucinas, in particular StcE, offer a powerful way to cleave densely O-glycosylated domains (72), and here, further knowledge of cleavage sites and O-glycoform dependencies are needed. These enzymes may be dependent on particular O-glycan structures and/or glycopeptide motifs, for example, the OgpA (OpeRATOR) glycoprotease cleaves N-terminal to Ser/Thr residues with attached core1 O-glycans (87). Sialic acid and branching of the O-glycan inhibit cleavage. OgpA provides a valuable tool for general O-glycoproteomics strategies but perhaps is less informative when it comes to mucin TRs with high density of O-glycans leaving too small fragments. In contrast, the StcE glycomucininase was reported to cleave a specific sequence motif (S/TX \downarrow S/T) with requirement of an O-glycan at P2 and option for an O-glycan at P1', but with little influence by the actual structure of the O-glycan(s) (16). However, we found that StcE cleavage appears to be blocked by STn and core3 O-glycans (34), and here, we provided further insight into the peptide sequon for StcE finding that the consensus substrate motif may be refined to S/TX \downarrow S/T, where X \neq P/Q/V. This was particularly useful for bottom-up analysis of the MUC2 TRs that essentially only consist of the broader S/TX \downarrow S/T motifs, since this enabled

release of reasonable-sized glycopeptide fragments suitable for LC-MS/MS analysis (Fig. 6). We also used BT4244 reported to cleave N-terminally to Thr/Ser residues with Tn or T O-glycans (46) and preference for Tn (47). We found that BT4244 exhibits preference toward O-glycan diads (TS, TT, and ST). Mucin TR reporters are valuable tools to dissect the fine specificities of these important glycoproteases not only in terms of protein substrate motifs but also with respect to the role of O-glycan structures given the capabilities to produce distinct glycoforms with the glycoengineered cell-based expression platform.

We extended the bottom-up strategy to a native mucin to further explore our finding that the mucin TR reporters were glycosylated with essentially full occupancy. We selected the mucin OSM because it is known to have homogeneous STn O-glycans, and it is widely used for characterization of antibodies to the cancer-associated STn and Tn (after removal of sialic acids) O-glycans (78, 88, 89). A large fraction of the many monoclonal antibodies available to STn and Tn antigens were elicited with OSM and AOSM as the immunogen, and partially desialylated OSM was used in a clinical trial to stimulate immunity in cancer patients (90). Interestingly though, the gene and sequence of this mucin has not been described in the literature to our knowledge, but with three peptide sequences originally obtained by Hill *et al.* (76, 77), the putative OSM gene was identified by a BLAST search, and by use of proteases and glycomucinas, we could confirm the authenticity of the OSM gene (Figs. 8 and S9). Importantly, our analysis of AOSM revealed that this mucin being naturally glycosylated with STn O-glycans exhibited near complete occupancy, which confirms our findings not only of high occupancy of mucin TRs by analysis of the reporters but also that mucins can be produced with STn O-glycans in normal cells without interference with O-glycan occupancy by premature sialylation.

In summary, our studies provide deeper insights into O-glycosylation of mucins suggesting that these regions carry O-glycans at all potential Ser/Thr residues with near full stoichiometry. We provided further evidence for the cell-based mucin TR array platform as a valid model to study O-glycosylation of mucins and importantly to produce representative mucin fragments with custom-designed glycosylation. There are still considerable challenges in analyzing more complex glycoforms and natural mucins, but the intact MS and bottom-up analytic strategies developed should be of wider use and stimulate further progress.

Experimental procedures

Cell culture

HEK293 cells were cultured in Dulbecco's modified Eagle's medium (Sigma-Aldrich) supplemented with 10% heat-inactivated fetal bovine serum (Sigma-Aldrich) and 2 mM GlutaMAX (Gibco) in a humidified incubator at 37 °C and 5% CO₂. A suspension HEK293 cell line was cultured in an orbital shaker in F17 medium (Gibco) supplemented with 0.1 Kolliphor P188 (Sigma-Aldrich) and 2% Glutamax. All glycoengineered isogenic HEK293 cells used in this study are

available as part of the cell-based glycan array resource (35, 91).

Production and purification of recombinant mucin TR reporters

The design and construction of the mucin TR reporters were previously reported (34, 36), and the full sequences used in this study are shown in Fig. S1. Glycoengineered HEK293 cell lines stably expressing secreted mucin reporters (34) were used for production by seeding cells at 0.25×10^6 cells/ml and culturing for 5 days before harvesting. Secreted reporters were purified by nickel–nitrilotriacetic acid affinity (Qiagen) chromatography (pre-equilibration with 25 mM sodium phosphate, 0.5 M NaCl, 10 mM imidazole [pH 7.4] and eluted with addition of 250 mM imidazole). Purified reporters were analyzed by SDS-PAGE and quantified using a Pierce BCA Protein Assay Kit (Thermo Fisher Scientific).

Isolation of mucin TR O-glycodomains

Purified mucin TR reporters were digested with Lys-C (Roche) (1:35 enzyme/substrate ratio) for 18 h at 37 °C in 50 mM ammonium bicarbonate buffer (pH 8.0). Reactions were heat inactivated at 98 °C for 10 min, and O-glycodomains were isolated by C4 HPLC (Aeris C4; 3.6 μ m, 200 Å, 250 \times 2.1 mm; Phenomenex) using 0% to 100% solvent B during 80 min (A: 0.1% TFA and B: 90% acetonitrile [ACN] in 0.1% TFA) with flow rate set at 0.2 ml/min. The fractions containing the O-glycodomain TR reporters were detected by VVA lectin ELISA. Collected fractions were freeze dried twice, and approximately, 1 μ g was resuspended in 20 μ l 0.1% formic acid (FA) and subjected to intact MS analysis. For MUC1-core3, after C4 HPLC purification, samples were desialylated with 40 mU *Clostridium perfringens* neuraminidase (Sigma–Aldrich) at 37 °C for 5 h in 50 mM sodium acetate buffer (pH 5.0) and subsequently purified by C18 HPLC (Aeris; 3.6 μ m WIDE-PORE XB-C18, 200 Å, 250 \times 2.1 mm; Phenomenex) using same chromatographic conditions as described previously.

ELISA

MaxiSorp 96-well plates (Nunc) were coated with diluted HPLC fractions with concentrations from approximately 1 ng/ μ l. Plates were blocked with PLI-P buffer (PO₄, Na/K, 1% Triton X-100, 1% bovine serum albumin, pH 7.4), incubated with 1 μ g/ml biotinylated-lectin VVA (Vector Laboratories and Lectenz Bio) for 1 h at room temperature, followed by incubation with streptavidin-conjugated horseradish peroxidase (1:5000 dilution) (Dako) for 1 h, and treatment with TMB (3,3',5,5'-tetramethylbenzidine substrate) (Dako) and 0.5 M H₂SO₄ to stop reactions. Absorbance was read at 450 nm.

Intact MS analysis

Intact MS analysis of mucin TRs was performed by EASY-nLC 1200 UHPLC (Thermo Scientific Scientific) interfaced *via* nanoSpray Flex ion source to an on Orbitrap Fusion/Lumos instrument (Thermo Fisher Scientific) using “high” mass range setting in *m/z* range 700 to 4000. Instrument was operated in

“low pressure” mode to provide optimal detection of intact protein masses. MS parameter settings: spray voltage 2.2 kV and source fragmentation energy 35 V. All ions were detected in Orbitrap at a resolution of 7500 (at *m/z* 200). The number of microscans was set to 20. The nLC was operated in a single analytical column set up using PicoFrit Emitters (New Objectives; inner diameter of 75 μ m) packed in-house with C4 phase (Dr Maisch, particle size of 3.0 mm; 16–20 cm column length) or C18 phase (Dr Maisch, particle size of 1.9 mm; column length of 16–20). Each sample was injected onto the column and eluted in gradients from 5 to 30% B in 25 min, from 30 to 100% B in 20 min, and 100% B for 15 min at 300 nl/min (solvent A, 100% H₂O; solvent B, 80% ACN; both having 0.1% [v/v] FA).

Bottom-up MS analysis

LC–MS/MS site-specific O-glycopeptide analysis of mucin TRs was performed on EASY-nLC 1000 UHPLC interfaced *via* nanoSpray Flex ion source to an Orbitrap Fusion MS or EASY-nLC 1000 UHPLC interfaced *via* New Objectives ion source to an Orbitrap Fusion MS (Thermo Fisher Scientific). Briefly, the nLC was operated in a single analytical column set up using PicoFrit Emitters (New Objectives; inner diameter of 75 μ m) packed in-house with Reprosil-Pure-AQ C18 phase (Dr Maisch; particle size of 1.9 μ m, column length of 19–21 cm). Each sample was injected onto the column and eluted in gradients from 3 to 32% B for glycopeptides and 10 to 40% for released and labeled glycans in 45 min at 200 nl/min (solvent A, 100% H₂O; solvent B, 100% ACN; both containing 0.1% [v/v] FA). A precursor MS1 scan (*m/z* 350–2000) of intact peptides was acquired in the Orbitrap at the nominal resolution setting of 120,000, followed by Orbitrap higher-energy collisional dissociation–MS2 and electron-transfer dissociation–MS2 at the nominal resolution setting of 60,000 of the five most abundant multiply charged precursors in the MS1 spectrum; a minimum MS1 signal threshold of 50,000 was used for triggering data-dependent fragmentation events. Targeted MS/MS analysis was performed by setting up a targeted MSn Scan Properties pane.

Data analysis

For intact MS and site-specific bottom-up analysis, raw spectra were deconvoluted to zero charge by BioPharma Finder Software (Thermo Fisher Scientific) as previously described with minor modifications (34). Briefly, sliding windows method was used for chromatography and source spectra with target average spectrum width of 0.1 min. Xtract deconvolution algorithm was used for bottom-up data, and ReSpect deconvolution algorithm was used for intact MS data. Glycoproteoforms were annotated from zero-charge deconvoluted intact MS data by in-house written SysBioWare software (92) using average masses of hexose, HexNAc, and the known predicted mass of the mucin TR reporter sequences. For site-specific glycopeptide identification, the corresponding higher-energy collisional dissociation MS/MS and electron-transfer dissociation MS/MS were analyzed by Proteome Discoverer 2.2 software (Thermo Fisher Scientific).

Mucin glycosylation in glycoengineered HEK293 cells

Mucin TR digestion for bottom-up MS analysis

For bottom-up analysis, C4 HPLC-purified O-glyco-domains were digested with Asp-N (2× [1:35] ratio for 18 h at 37 °C, in 100 mM Tris-HCl buffer [pH 8.0]), Glu-C (1:10 ratio for 18 h at 37 °C in 50 mM ammonium bicarbonate buffer [pH 8.0]), or trypsin (1:8 ratio for 12 h at 37 °C in 50 mM ammonium bicarbonate buffer). Reactions were stopped with the addition of 1 µl concentrated TFA. Samples were desalted with Stage Tips (C18 sorbent-Empore 3M), freeze dried twice, and ~1 µg was resuspended in 20 µl 0.1% FA for nLC-MS/MS. The MUC1-core3 after C4 HPLC purification was desialylated with 40 mU *C. perfringens* neuraminidase (Sigma-Aldrich) at 37 °C for 5 h in 50 mM sodium acetate buffer (pH 5.0) and subsequently purified by C18 HPLC.

Recombinant StcE and BT4244 enzymes were produced in *E. coli* as reported previously (34). BT4244 was produced with a codon-optimized sequence (residues 35–857) cloned into pet28 (kanamycin) (Twist Bioscience). Digestions were performed with secreted mucin reporters and BT4244 at an enzyme to substrate ratio of 1:50 for 3 h at 37 °C in 50 mM ammonium bicarbonate buffer (final volume of 100 µl) and StcE at an enzyme to substrate ratio of 1:50 for 1 h at 37 °C in H₂O (final volume of 10 µl) followed by heat inactivation at 98 °C for 10 min. Peptides were purified by C18 HPLC. Fractions containing Tn-glycopeptides were detected by VVA lectin ELISA. The collected fractions were freeze dried twice, and approximately, 1 µg was resuspended in 20 µl 0.1% FA to be further analyzed by nLC-MS/MS.

Data availability

All data generated or analyzed during this study are included in this article and supporting information files.

The MS data have been deposited to the ProteomeXchange Consortium via the PRIDE (93) partner repository with the dataset identifier PXD029885.

Supporting information—This article contains supporting information. Supporting information includes Supporting Figs. (S1–S9) and one excel file (Supporting File S1) (34, 76, 77).

Author contributions—A. K., Y. N., S. Y. V., and H. C. conceptualization; A. K., Y. N., S. Y. V., and H. C. methodology; A. K., R. N., T. Č., L. S., D. M. S., S. F., Z. Y., R. V., Y. N., S. Y. V., and H. C. formal analysis; A. K., R. N., T. Č., L. S., D. M. S., S. F., Z. Y., R. V., Y. N., S. Y. V., and H. C. investigation; A. K., R. N., T. Č., L. S., D. M. S., S. F., Z. Y., R. V., Y. N., S. Y. V., and H. C. data curation; A. K. and H. C. writing—original draft; A. K., R. N., T. Č., L. S., D. M. S., S. F., Z. Y., R. V., Y. N., S. Y. V., and H. C. writing—review & editing; H. C. funding acquisition.

Funding and additional information—This work was supported by the Neye Foundation, Lundbeck Foundation, the Novo Nordisk Foundation, the European Commission (GlycoImaging H2020-MSCA-ITN-721297, BioCapture H2020-MSCA-ITN-722171), the Mizutani Foundation (to Y. N.), and the Danish National Research Foundation (grant no.: DNRF107).

Conflict of interest—The University of Copenhagen has filed a patent application on the cell-based display platform. GlycoDisplay Aps, Copenhagen, Denmark, has obtained a license to the field of the patent application. Y. N. and H. C. are cofounders of GlycoDisplay Aps and hold ownerships in the company. All the other authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: ACN, acetonitrile; AOSM, asialo-OSM; FA, formic acid; GALNT, GalNAc transferase; HEK293, human embryonic kidney 293 cell line; KI, knockin; KO, knock out; MS, mass spectrometry; OSM, ovine submaxillary mucin; PSGL-1, P-selectin-glycoprotein ligand 1; SDC3, syndecan-3; TR, tandem repeat; VVA, villosa agglutinin.

References

1. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2012) Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **22**, 736–756
2. Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., and Clausen, H. (2020) Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* **21**, 729–749
3. Goth, C. K., Vakhrushev, S. Y., Joshi, H. J., Clausen, H., and Schjoldager, K. T. (2018) Fine-tuning limited proteolysis: A major role for regulated site-specific O-glycosylation. *Trends Biochem. Sci.* **43**, 269–284
4. Schjoldager, K. T., Joshi, H. J., Kong, Y., Goth, C. K., King, S. L., Wandall, H. H., Bennett, E. P., Vakhrushev, S. Y., and Clausen, H. (2015) Deconstruction of O-glycosylation—GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. *EMBO Rep.* **16**, 1713–1722
5. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T. B. G., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Paul Bennett, E., Mandel, U., Brunak, S., Wandall, H. H., et al. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* **32**, 1478–1488
6. Mohl, J. E., Gerken, T. A., and Leung, M. Y. (2021) ISOGlyP: De novo prediction of isoform-specific mucin-type O-glycosylation. *Glycobiology* **31**, 168–172
7. Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T.-B. G., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Lavery, S. B., and Clausen, H. (2011) Mining the O-glycoproteome using zinc-finger nuclease—glycoengineered SimpleCell lines. *Nat. Methods* **8**, 977–982
8. Darula, Z., and Medzihradsky, K. F. (2018) Analysis of mammalian O-glycopeptides - we have made a good start, but there is a long way to go. *Mol. Cell Proteomics* **17**, 2–17
9. Vester-Christensen, M. B., Halim, A., Joshi, H. J., Steentoft, C., Bennett, E. P., Lavery, S. B., Vakhrushev, S. Y., and Clausen, H. (2013) Mining the O-mannose glycoproteome reveals cadherins as major O-mannosylated glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 21018–21023
10. Vakhrushev, S. Y., Steentoft, C., Vester-Christensen, M. B., Bennett, E. P., Clausen, H., and Lavery, S. B. (2013) Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with simplecells. *Mol. Cell Proteomics* **12**, 932–944
11. Riley, N. M., Bertozzi, C. R., and Pitteri, S. J. (2021) A pragmatic guide to enrichment strategies for mass spectrometry-based glycoproteomics. *Mol. Cell. Proteomics* **20**, 100029
12. Ye, Z., Mao, Y., Clausen, H., and Vakhrushev, S. Y. (2019) Glyco-DIA: A method for quantitative O-glycoproteomics with in silico-boosted glycopeptide libraries. *Nat. Methods* **16**, 902–910
13. Joshi, H. J., Jørgensen, A., Schjoldager, K. T., Halim, A., Dworkin, L. A., Steentoft, C., Wandall, H. H., Clausen, H., and Vakhrushev, S. Y. (2018) GlycoDomainViewer: A bioinformatics tool for contextual exploration of glycoproteomes. *Glycobiology* **28**, 131–136
14. Lavery, S. B., Steentoft, C., Halim, A., Narimatsu, Y., Clausen, H., and Vakhrushev, S. Y. (2015) Advances in mass spectrometry driven O-glycoproteomics. *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 33–42

15. Khoo, K. H. (2019) Advances toward mapping the full extent of protein site-specific O-GalNAc glycosylation that better reflects underlying glycomic complexity. *Curr. Opin. Struct. Biol.* **56**, 146–154
16. Malaker, S. A., Pedram, K., Ferracane, M. J., Bensing, B. A., Krishnan, V., Pett, C., Yu, J., Woods, E. C., Kramer, J. R., Westerlind, U., Dorigo, O., and Bertozzi, C. R. (2019) The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7278–7287
17. Gerken, T. A., Owens, C. L., and Pasumathy, M. (1997) Determination of the site-specific O-glycosylation pattern of the porcine submaxillary mucin tandem repeat glycopeptide. Model proposed for the polypeptide: GalNAc transferase peptide binding site. *J. Biol. Chem.* **272**, 9709–9719
18. Gerken, T. A., Tep, C., and Rarick, J. (2004) Role of peptide sequence and neighboring residue glycosylation on the substrate specificity of the uridine 5'-diphosphate- α -N-acetylgalactosamine:polypeptide N-acetylgalactosaminyl transferases T1 and T2: Kinetic modeling of the porcine and canine submax. *Biochemistry* **43**, 9888–9900
19. Gerken, T. A., Gilmore, M., and Zhang, J. (2002) Determination of the site-specific oligosaccharide distribution of the O-glycans attached to the porcine submaxillary mucin tandem repeat: Further evidence for the modulation of O-glycan side chain structures by peptide sequence. *J. Biol. Chem.* **277**, 7736–7751
20. Gerken, T. A., Zhang, J., Levine, J., and Elhammer, Å. (2002) Mucin core O-glycosylation is modulated by neighboring residue glycosylation status: Kinetic modeling of the site-specific glycosylation of the apo-porcine submaxillary mucin tandem repeat by UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases T1 and T2. *J. Biol. Chem.* **277**, 49850–49862
21. Gerken, T. A., Owens, C. L., and Pasumathy, M. (1998) Site-specific core 1 O-glycosylation pattern of the porcine submaxillary gland mucin tandem repeat. Evidence for the modulation of glycan length by peptide sequence. *J. Biol. Chem.* **273**, 26580–26588
22. Hanisch, F. G., Green, B. N., Bateman, R., and Peter-Katalinic, J. (1998) Localization of O-glycosylation sites of MUC1 tandem repeats by QTOF ESI mass spectrometry. *J. Mass Spectrom.* **33**, 358–362
23. Ali, L., Flowers, S. A., Jin, C., Bennet, E. P., Ekwall, A. K. H., and Karlsson, N. G. (2014) The O-glycomap of lubricin, a novel mucin responsible for joint lubrication, identified by site-specific glycopeptide analysis. *Mol. Cell Proteomics* **13**, 3396–3409
24. Corfield, A. P. (2015) Mucins: A biologically relevant glycan barrier in mucosal protection. *Biochim. Biophys. Acta - Gen. Subj.* **1850**, 236–252
25. Hollingsworth, M. A., and Swanson, B. J. (2004) Mucins in cancer: Protection and control of the cell surface. *Nat. Rev. Cancer* **4**, 45–60
26. Hattrup, C. L., and Gendler, S. J. (2008) Structure and function of the cell surface (tethered) mucins. *Annu. Rev. Physiol.* **70**, 431–457
27. Hansson, G. C. (2019) Mucus and mucins in diseases of the intestinal and respiratory tracts. *J. Intern. Med.* **285**, 479–490
28. Wilkins, P. P., Moore, K. L., McEver, R. P., and Cummings, R. D. (1995) Tyrosine sulfation of P-selectin glycoprotein ligand-1 is required for high affinity binding to P-selectin. *J. Biol. Chem.* **270**, 22677–22680
29. O'Brien, T. J., Beard, J. B., Underwood, L. J., Dennis, R. A., Santin, A. D., and York, L. (2001) The CA 125 gene: An extracellular superstructure dominated by repeat sequences. *Tumor Biol.* **22**, 348–366
30. Yin, B. W. T., and Lloyd, K. O. (2001) Molecular cloning of the CA125 ovarian cancer antigen: Identification as a new mucin, MUC16. *J. Biol. Chem.* **276**, 27371–27375
31. Marcos-Silva, L., Narimatsu, Y., Halim, A., Campos, D., Yang, Z., Tarp, M. A., Pereira, P. J. B., Mandel, U., Bennett, E. P., Vakhruşev, S. Y., Lavery, S. B., David, L., and Clausen, H. (2014) Characterization of binding epitopes of CA125 monoclonal antibodies. *J. Proteome Res.* **13**, 3349–3359
32. Lang, T., Hansson, G. C., and Samuelsson, T. (2007) Gel-forming mucins appeared early in metazoan evolution. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16209–16214
33. Irimura, T., Denda, K., Iida, S. I., Takeuchi, H., and Kato, K. (1999) Diverse glycosylation of MUC1 and MUC2: Potential significance in tumor immunity. *J. Biochem.* **126**, 975–985
34. Nason, R., Büll, C., Konstantinidi, A., Sun, L., Ye, Z., Halim, A., Du, W., Sørensen, D. M., Durbesson, F., Furukawa, S., Mandel, U., Joshi, H. J., Dworkin, L. A., Hansen, L., David, L., et al. (2021) Display of the human mucinome with defined O-glycans by gene engineered cells. *Nat. Commun.* **12**, 1–16
35. Narimatsu, Y., Joshi, H. J., Nason, R., Van Coillie, J., Karlsson, R., Sun, L., Ye, Z., Chen, Y. H., Schjoldager, K. T., Steentoft, C., Furukawa, S., Bensing, B. A., Sullam, P. M., Thompson, A. J., Paulson, J. C., et al. (2019) An atlas of human glycosylation pathways enables display of the human glycome by gene engineered cells. *Mol. Cell* **75**, 394–407.e5
36. Büll, C., Nason, R., Sun, L., Van Coillie, J., Sørensen, D. M., Moons, S. J., Yang, Z., Arbitman, S., Fernandes, S. M., Furukawa, S., McBride, R., Nycholat, C. M., Adema, G. J., Paulson, J. C., Schnaar, R. L., et al. (2021) Probing the binding specificities of human Siglecs by cell-based glycan arrays. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–12
37. Yang, Y., Liu, F., Franc, V., Halim, L. A., Schellekens, H., and Heck, A. J. R. (2016) Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nat. Commun.* **7**, 1–10
38. Wohlschläger, T., Scheffler, K., Forstenleher, I. C., Skala, W., Senn, S., Damoc, E., Holzmann, J., and Huber, C. G. (2018) Native mass spectrometry combined with enzymatic dissection unravels glycoform heterogeneity of biopharmaceuticals. *Nat. Commun.* **9**, 1–9
39. Čaval, T., Tian, W., Yang, Z., Clausen, H., and Heck, A. J. R. (2018) Direct quality control of glycoengineered erythropoietin variants. *Nat. Commun.* **9**, 3342
40. Lin, Y. H., Zhu, J., Meijer, S., Franc, V., and Heck, A. J. R. (2019) Glycoproteogenomics: A frequent gene polymorphism affects the glycosylation pattern of the human serum fetuin/ α -2-HS-Glycoprotein. *Mol. Cell Proteomics* **18**, 1479–1490
41. Čaval, T., Lin, Y. H., Varkila, M., Reiding, K. R., Bonten, M. J. M., Cremer, O. L., Franc, V., and Heck, A. J. R. (2021) Glycoproteoform profiles of individual patients' plasma α -1-antichymotrypsin are unique and extensively remodeled following a septic episode. *Front. Immunol.* **11**, 1–14
42. Wu, D., Struwe, W. B., Harvey, D. J., Ferguson, M. A. J., and Robinson, C. V. (2018) N-glycan microheterogeneity regulates interactions of plasma proteins. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 8763–8768
43. Wu, D., Li, J., Struwe, W. B., and Robinson, C. V. (2019) Probing: N-glycoprotein microheterogeneity by lectin affinity purification-mass spectrometry analysis. *Chem. Sci.* **10**, 5146–5155
44. Narimatsu, Y., Büll, C., Chen, Y. H., Wandall, H. H., Yang, Z., and Clausen, H. (2021) Genetic glycoengineering in mammalian cells. *J. Biol. Chem.* **296**, 100448
45. Čaval, T., de Haan, N., Konstantinidi, A., and Vakhruşev, S. Y. (2021) Quantitative characterization of O-GalNAc glycosylation. *Curr. Opin. Struct. Biol.* **68**, 135–141
46. Shon, D. J., Malaker, S. A., Pedram, K., Yang, E., Krishnan, V., Dorigo, O., and Bertozzi, C. R. (2020) An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 21299–21307
47. Noach, I., Ficko-blean, E., Pluvinaige, B., Stuart, C., Jenkins, M. L., Brochu, D., Buenbrazo, N., Wakarchuk, W., Burke, J. E., Gilbert, M., and Boraston, A. B. (2017) Recognition of protein-linked glycans as a determinant of peptidase activity. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E679–E688
48. Engelmann, K., Kinlough, C. L., Müller, S., Razawi, H., Baldus, S. E., Hughey, R. P., and Hanisch, F. G. (2005) Transmembrane and secreted MUC1 probes show trafficking-dependent changes in O-glycan core profiles. *Glycobiology* **15**, 1111–1124
49. Daniel, E. J. P., Las Rivas, M., Lira-Navarrete, E., García-García, A., Hurtado-Guerrero, R., Clausen, H., and Gerken, T. A. (2020) Ser and Thr acceptor preferences of the GalNAc-Ts vary among isoenzymes to modulate mucin-type O-glycosylation. *Glycobiology* **30**, 910–922
50. Corzana, F., Busto, J. H., Jiménez-Osés, G., De Luis, M. G., Asensio, J. L., Jiménez-Barbero, J., Peregrina, J. M., and Avenoza, A. (2007) Serine versus threonine glycosylation: The methyl group causes a drastic alteration on the carbohydrate orientation and on the surrounding water shell. *J. Am. Chem. Soc.* **129**, 9458–9467
51. Narimatsu, Y., Joshi, H. J., Schjoldager, K. T., Hintze, J., Halim, A., Steentoft, C., Nason, R., Mandel, U., Bennett, E. P., Clausen, H., and Vakhruşev, S. Y. (2019) Exploring regulation of protein O-glycosylation

Mucin glycosylation in glycoengineered HEK293 cells

- in isogenic human HEK293 cells by differential O-glycoproteomics. *Mol. Cell. Proteomics* **18**, 1396–1409
52. Bagdonaite, I., Pallesen, E. M., Ye, Z., Vakhrushev, S. Y., Marinova, I. N., Nielsen, M. I., Kramer, S. H., Pedersen, S. F., Joshi, H. J., Bennett, E. P., Dabelsteen, S., and Wandall, H. H. (2020) O-glycan initiation directs distinct biological pathways and controls epithelial differentiation. *EMBO Rep.* **21**, 1–17
 53. Lavrsen, K., Dabelsteen, S., Vakhrushev, S. Y., Levann, A. M. R., Haue, A. D., Dylander, A., Mandel, U., Hansen, L., Frodin, M., Bennett, E. P., and Wandall, H. H. (2018) De novo expression of human polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6) in colon adenocarcinoma inhibits the differentiation of colonic epithelium. *J. Biol. Chem.* **293**, 1298–1314
 54. Kong, Y., Joshi, H. J., Schjoldager, K. T. B. G., Madsen, T. D., Gerken, T. A., Vester-Christensen, M. B., Wandall, H. H., Bennett, E. P., Levery, S. B., Vakhrushev, S. Y., and Clausen, H. (2015) Probing polypeptide GalNAc-transferase isoform substrate specificities by *in vitro* analysis. *Glycobiology* **25**, 55–65
 55. de las Rivas, M., Lira-Navarrete, E., Gerken, T. A., and Hurtado-Guerrero, R. (2019) Polypeptide GalNAc-ts: From redundancy to specificity. *Curr. Opin. Struct. Biol.* **56**, 87–96
 56. Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2000) The lectin domain of UDP-N-acetyl-d-galactosamine: PolypeptideN-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. *J. Biol. Chem.* **275**, 38197–38205
 57. Bennett, E. P., Hassan, H., Hollingsworth, M. A., and Clausen, H. (1999) A novel human UDP-N-acetyl-D-galactosamine:polypeptideN-acetylgalactosaminyltransferase, GalNAc-T7, with specificity for partial GalNAc-glycosylated acceptor substrates. *FEBS Lett.* **460**, 226–230
 58. Kubota, T., Shiba, T., Sugioka, S., Furukawa, S., Sawaki, H., Kato, R., Wakatsuki, S., and Narimatsu, H. (2006) Structural basis of carbohydrate transfer activity by human UDP-GalNAc: Polypeptide α -N-acetylgalactosaminyltransferase (pp-GalNAc-T10). *J. Mol. Biol.* **359**, 708–727
 59. Guo, J.-M., Zhang, Y., Cheng, L., Iwasaki, H., Wang, H., Kubota, T., Tachibana, K., and Narimatsu, H. (2002) Molecular cloning and characterization of a novel member of the UDP-GalNAc:polypeptideN-acetylgalactosaminyltransferase family, pp-GalNAc-T121. *FEBS Lett.* **524**, 211–218
 60. De Las Rivas, M., Paul Daniel, E. J., Coelho, H., Lira-Navarrete, E., Raich, L., Compañón, I., Diniz, A., Lagartera, L., Jiménez-Barbero, J., Clausen, H., Rovira, C., Marcelo, F., Corzana, F., Gerken, T. A., and Hurtado-Guerrero, R. (2018) Structural and mechanistic insights into the catalytic-domain-mediated short-range glycosylation preferences of GalNAc-T4. *ACS Cent. Sci.* **4**, 1274–1290
 61. Steentoft, C., Fuhrmann, M., Battisti, F., Van Coillie, J., Madsen, T. D., Campos, D., Halim, A., Vakhrushev, S. Y., Joshi, H. J., Schreiber, H., Mandel, U., and Narimatsu, Y. (2019) A strategy for generating cancer-specific monoclonal antibodies to aberrant O-glycoproteins: Identification of a novel dysadherin-tn antibody. *Glycobiology* **29**, 307–319
 62. Tarp, M. A., Sørensen, A. L., Mandel, U., Paulsen, H., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2007) Identification of a novel cancer-specific immunodominant glycopeptide epitope in the MUC1 tandem repeat. *Glycobiology* **17**, 197–209
 63. Bennett, E. P., Hassan, H., Mandel, U., Mirgorodskaya, E., Roepstorff, P., Burchell, J., Taylor-Papadimitriou, J., Hollingsworth, M. A., Merckx, G., Van Kessel, A. G., Eiberg, H., Steffensen, R., and Clausen, H. (1998) Cloning of a human UDP-N-acetyl- α -D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase that complements other GalNAc-transferases in complete O-glycosylation of the MUC1 tandem repeat. *J. Biol. Chem.* **273**, 30472–30481
 64. Hounsell, E. F., Lawson, A. M., Feeney, J., Gooi, H. C., Pickering, N. J., Stoll, M. S., Lui, S. C., and Feizi, T. (1985) Structural analysis of the O-glycosidically linked core-region oligosaccharides of human meconium glycoproteins which express oncofoetal antigens. *Eur. J. Biochem.* **148**, 367–377
 65. Marcos, N. T., Pinho, S., Grandela, C., Cruz, A., Samyn-Petit, B., Harduin-Lepers, A., Almeida, R., Silva, F., Morais, V., Costa, J., Kihlberg, J., Clausen, H., and Reis, C. A. (2004) Role of the human ST6GalNAc-I and ST6GalNAc-II in the synthesis of the cancer-associated Sialyl-Tn antigen. *Cancer Res.* **64**, 7050–7057
 66. Sewell, R., Bäckström, M., Dalziel, M., Gschmeissner, S., Karlsson, H., Noll, T., Gätgens, J., Clausen, H., Hansson, G. C., Burchell, J., and Taylor-Papadimitriou, J. (2006) The ST6GalNAc-I sialyltransferase localizes throughout the golgi and is responsible for the synthesis of the tumor-associated sialyl-Tn O-glycan in human breast cancer. *J. Biol. Chem.* **281**, 3586–3594
 67. Iwai, T., Inaba, N., Naundorf, A., Zhang, Y., Gotoh, M., Iwasaki, H., Kudo, T., Togayachi, A., Ishizuka, Y., Nakanishi, H., and Narimatsu, H. (2002) Molecular cloning and characterization of a novel UDP-GlcNAc: GalNAc-peptide β 1,3-N-acetylglucosaminyltransferase (β 3Gn-T6), an enzyme synthesizing the core 3 structure of O-glycans. *J. Biol. Chem.* **277**, 12802–12809
 68. Iwai, T., Kudo, T., Kawamoto, R., Kubota, T., Togayachi, A., Hiruma, T., Okada, T., Kawamoto, T., Morozumi, K., and Narimatsu, H. (2005) Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4572–4577
 69. 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) The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc: Lectin binding to GalNAc-glycopeptide substrates is required for high density GalNAc-O-glycosylation. *Glycobiology* **17**, 374–387
 70. Kato, K., Takeuchi, H., Miyahara, N., Kanoh, A., Hassan, H., Clausen, H., and Irimura, T. (2001) Distinct orders of GalNAc incorporation into a peptide with consecutive threonines. *Biochem. Biophys. Res. Commun.* **287**, 110–115
 71. [preprint] Malaker, S. A., Riley, N. M., Shon, D. J., Pedram, K., Krishnan, V., Dorigo, O., and Bertozzi, C. R. (2021) Revealing the human mucinome. *bioRxiv*. <https://doi.org/10.1101/2021.01.27.428510>
 72. Shon, D. J., Kuo, A., Ferracane, M. J., and Malaker, S. A. (2021) Classification, structural biology, and applications of mucin domain-targeting proteases. *Biochem. J.* **478**, 1585–1603
 73. Lathem, W. W., Grys, T. E., Witowski, S. E., Torres, A. G., Kaper, J. B., Tarr, P. I., and Welch, R. A. (2002) StcE, a metalloprotease secreted by *Escherichia coli* O157:H7, specifically cleaves C1 esterase inhibitor. *Mol. Microbiol.* **45**, 277–288
 74. Bhargava, A. S., and Gottschalk, A. (1966) Studies on glycoproteins. XIII. Preparation of ovine submaxillary gland glycoprotein by gel filtration and its physical, chemical and immunochemical characterization. *BBA - Gen. Subj.* **127**, 223–231
 75. Tettamanti, G., and Pigman, W. (1968) Purification and characterization of bovine and ovine submaxillary mucins. *Arch. Biochem. Biophys.* **124**, 41–50
 76. Hill, H. D., Schwyzer, M., Steinman, H., and Hill, R. L. (1977) Ovine submaxillary mucin. Primary structure and peptide substrates of UDP N acetylgalactosamine mucin transferase. *J. Biol. Chem.* **252**, 3799–3804
 77. Hill, H. D., Reynolds, J. A., and Hill, R. L. (1977) Purification, composition, molecular weight, and subunit structure of ovine submaxillary mucin. *J. Biol. Chem.* **252**, 3791–3798
 78. Kjeldsen, T., Clausen, H., Hirohashi, S., Ogawa, T., Iijima, H., and Hakomori, S. (1988) Preparation and characterization of monoclonal antibodies directed to the tumor-associated o-linked sialosyl-2 \rightarrow 6 α -N-Acetylgalactosaminyl (Sialosyl-Tn) Epitope. *Cancer Res.* **48**, 2214–2220
 79. O'Boyle, K. P., Coatsworth, S., Anthony, G., Ramirez, M., Greenwald, E., Kaleya, R., Steinberg, J. J., Dutcher, J. P., and Wiernik, P. H. (2006) Effects of desialylation of ovine submaxillary gland mucin (OSM) on humoral and cellular immune responses to Tn and sialylated Tn. *Cancer Immun.* **6**, 1–9
 80. Vinall, L. E., Hill, A. S., Pigny, P., Pratt, W. S., Toribara, N., Gum, J. R., Kim, Y. S., Porchet, N., Aubert, J. P., and Swallow, D. M. (1998) Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5. *Hum. Genet.* **102**, 357–366
 81. Cummings, R. D. (2009) The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5**, 1087–1104

82. Čaval, T., Heck, A. J. R., and Reiding, K. R. (2021) Meta-heterogeneity: Evaluating and describing the diversity in glycosylation between sites on the same glycoprotein. *Mol. Cell. Proteomics* **20**, 100010
83. Bäckström, M., Link, T., Olson, F. J., Karlsson, H., Graham, R., Picco, G., Burchell, J., Taylor-Papadimitriou, J., Noll, T., and Hansson, G. C. (2003) Recombinant MUC1 mucin with a breast cancer-like O-glycosylation produced in large amounts in Chinese-hamster ovary cells. *Biochem. J.* **376**, 677–686
84. Olson, F. J., Bäckström, M., Karlsson, H., Burchell, J., and Hansson, G. C. (2005) A MUC1 tandem repeat reporter protein produced in CHO-K1 cells has sialylated core 1 O-glycans and becomes more densely glycosylated if coexpressed with polypeptide-GalNAc-T4 transferase. *Glycobiology* **15**, 177–191
85. de las Rivas, M., Paul Daniel, E. J., Narimatsu, Y., Compañón, I., Kato, K., Hermosilla, P., Thureau, A., Ceballos-Laita, L., Coelho, H., Bernadó, P., Marcelo, F., Hansen, L., Maeda, R., Lostao, A., Corzana, F., *et al.* (2020) Molecular basis for fibroblast growth factor 23 O-glycosylation by GalNAc-T3. *Nat. Chem. Biol.* **16**, 351–360
86. Čaval, T., Buettner, A., Habeger, M., Reusch, D., and Heck, A. J. R. (2021) Discrepancies between high-resolution native and glycopeptide-centric mass spectrometric approaches: A case study into the glycosylation of erythropoietin variants. *J. Am. Soc. Mass Spectrom.* **32**, 2099–2104
87. Trastoy, B., Naegeli, A., Anso, I., Sjögren, J., and Guerin, M. E. (2020) Structural basis of mammalian mucin processing by the human gut O-glycopeptidase OgpA from *Akkermansia muciniphila*. *Nat. Commun.* **11**, 1–14
88. Springer, G. F., and Desai, P. R. (1985) Tn epitopes, immunoreactive with ordinary anti-Tn antibodies, on normal, desialylated human erythrocytes and on Thomsen-Friedenreich antigen isolated therefrom. *Mol. Immunol.* **22**, 1303–1310
89. Numata, Y., Nakada, H., Fukui, S., Kitagawa, H., Ozaki, K., Inoue, M., Kawasaki, T., Funakoshi, I., and Yamashina, I. (1990) A monoclonal antibody directed to Tn antigen. *Biochem. Biophys. Res. Commun.* **170**, 981–985
90. O'Boyle, K. P., Zamore, R., Adhuri, S., Cohen, A., Kemeny, N., Welt, S., Lloyd, K. O., Oettgen, H. F., Old, L. J., and Livingston, P. O. (1992) Immunization of colorectal cancer patients with modified ovine submaxillary gland mucin and adjuvants Induces IgM and IgG antibodies to sialylated Tn. *Cancer Res.* **52**, 5663–5667
91. Büll, C., Joshi, H. J., Clausen, H., and Narimatsu, Y. (2020) Cell-based glycan arrays—a practical guide to dissect the human glycome. *STAR Protoc.* **1**, 100017
92. Vakhrushev, S. Y., Dadimov, D., and Peter-Katalinić, J. (2009) Software platform for high-throughput glycomics. *Anal. Chem.* **81**, 3252–3260
93. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, Ş., *et al.* (2019) The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47**, D442–D450
94. Varki, A., Cummings, R. D., Aebi, M., Packer, N. H., Seeberger, P. H., Esko, J. D., Stanley, P., Hart, G., Darvill, A., Kinoshita, T., Prestegard, J. J., Schnaar, R. L., Freeze, H. H., Marth, J. D., Bertozzi, C. R., *et al.* (2015) Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323–1324