

Unraveling *O*-Glycan Diversity of Mucins: Insights from SmE Mucinase and Ultraviolet Photodissociation Mass Spectrometry

Amanda Helms,¹ Vincent Chang,² Stacy A. Malaker,² Jennifer S. Brodbelt^{1*}

¹Department of Chemistry, University of Texas, Austin, TX 78712

²Department of Chemistry, Yale University, New Haven, CT 06511

*correspondence to: jbrodbelt@cm.utexas.edu

Introduction

An estimated half of all known proteins are glycosylated in some manner. Glycosylation is a co- or post-translational modification (PTM) in which oligosaccharides are attached to proteins to impart changes in structure or function.¹ *O*-glycosylation of threonine, serine and less commonly tyrosine residues is one common type of glycosylation, and it may stabilize protein structures,²⁻⁴ mediate cell-cell interactions,^{3,5} or modulate activation or inhibition of biological processes.⁶⁻⁸ Some *O*-glycoproteins, such as mucin proteins, contain highly glycosylated regions predominantly populated by serine and threonine residues, with many being capable of hosting glycans.⁹ The microheterogeneity of each glycosite coupled with the macroheterogeneity of the differing occupation of various glycosites further increases the complexity of *O*-glycoproteins,³ creating a multi-level analytical problem for which tandem mass spectrometry has evolved as one of the premier solutions. Owing to the many types of saccharides that contribute to *O*-glycosylation,¹⁰ identification and characterization of all glycans decorating a protein is an equally formidable challenge beyond just site localization. This challenge is arguably more daunting for unravelling *O*-glycosylation than *N*-glycosylation. *N*-glycosylation occurs at asparagine residues found in a well-defined binding motif of N-X-S/T, where X is any amino acid except for proline.¹¹ This motif-specificity facilitates localization of *N*-glycans. Additionally, *N*-glycans may be released from proteins using endoglycosidases, such as PNGase F, thus facilitating analysis of the released glycans and streamlining their compositional characterization.¹¹ In contrast, *O*-glycans, while as prevalent as *N*-glycans, are more complex and not localized to highly specific sequence motifs of proteins,¹² instead can modify any serine, threonine¹³, or (occasionally) tyrosine.¹⁴ Eight core structures have been identified for *O*-glycans, with cores 1-4 being dominant¹ and all exhibiting extensive microheterogeneity. There is currently no universal enzyme for cleaving *O*-glycans from proteins, impeding glycomics analysis.

Digesting *O*-glycoproteins into smaller glycopeptides offers a streamlined strategy for localizing and dissecting the glycans and characterizing the peptide sequences, assuming that sufficient information

can be generated and harvested from MS/MS spectra of the resulting glycopeptides. Several *O*-glycoproteases have been explored for generating *O*-glycopeptides,^{15–18} and a few, aptly named mucinases, are applicable to the extremely complex mucin-type *O*-glycoproteins.^{15,19} Mucinases can navigate the highly glycosylated regions of mucin glycoproteins and cleave between adjacent glycosites, resulting in small peptides that can contain multiple occupied glycosites.^{20,21}

Early discovered mucinases include StcE from *E. coli*^{15,22,23} and BT4244 from *Bacteroides thetaiotaomicron*.^{20,24} BT4244 has been used previously to generate *O*-glycopeptides for analysis by mass spectrometry (MS),^{18,21} and its specificity for *O*-glycoproteins has been examined.²⁴ Similarly, StcE has been used to map *O*-glycosites on mucin proteins^{21,22} and to investigate the “bottlebrush” structure of mucin domains.²⁵ A newly introduced mucinase, SmE from *Serratia marcescens*, has been utilized for characterizing the complex TIM family of mucin domain proteins,¹⁹ whose glycosylation patterns have not been fully cataloged. The utility of SmE for the study of highly glycosylated regions is especially pivotal for examination of mucin proteins, such as the human mucin proteins expressed by epithelial cells in the respiratory, digestive, gastrointestinal, and urogenital tracts (MUC1–22) that are implicated in various diseases.^{9,26,27} Each mucin protein is complex, and notably the aberrant glycosylation of several MUC proteins has been correlated with different cancers,^{23,28} an outcome that could be related to the overexpression or deregulation of the MUC proteins.²⁹ With its proficient ability to cleave between adjacent glycosites, SmE is poised to expand characterization of the most complex glycoproteins.

To capitalize on the proficiency of some of the new mucinases, like SmE, for cleaving glycoproteins, high performance MS/MS methods are needed to characterize the resulting ensembles of glycopeptides. Traditionally, collisionally activation dissociation (CAD) has been used to identify glycan compositions of glycopeptides through the production of B, C, Y, and Z glycosidic fragments. In addition, CAD produces *b/y* ions that provide peptide sequence information,^{26,30,31} but the ability of CAD to localize glycosites is impeded owing to the tendency of CAD to cleave labile PTMs from the peptides. Other MS/MS methods have proven more successful for glycosite localization. Electron transfer dissociation (ETD) generates *c/z* sequence ions that retain PTMs, including glycans, allowing mapping of glycosites of numerous *O*-glycoproteins.³² Combining ETD with supplemental collisional activation enhances the production of fragment ions, and one of these hybrid methods, EThcD (electron transfer higher collision energy dissociation), is a powerhouse for *O*-glycosite localization and characterization of multi-charged glycopeptides.³² With the ability to localize glycans and offer detailed information about their structures, EThcD is widely used for *O*-glycoproteomics, as demonstrated for probing the glycoproteome of human serum,³³ investigating glycoproteins in cerebrospinal fluid,³⁴ deciphering glycosylation of antibodies and their components,^{35,36} and analyzing the *O*-glycan profiles of individual glycoproteins.³⁷

The natural complexity of glycans creates challenges even for the most versatile MS/MS methods like EThcD. Glycans may exist in multiple structural arrangements even if the absolute composition is the same, as exemplified by the structure diversity of a sialylated core 1 glycan, composed of a single N-acetylgalactosamine (GalNAc), galactose (Gal), and sialic acid (Neu5Ac). The three saccharides can be linked sequentially or both the Gal and the Neu5Ac can be attached to the GalNAc. Additionally, the linkage between the Neu5Ac and Gal may be α 2,3 or α 2,6, thus encoding an additional level of structural detail. Assignment of all of these structural features has motivated exploration of other ion activation methods, such as ultraviolet photodissociation (UVPD).³⁸ In addition to generating fragment ions suitable for determining the peptide sequence (*a*, *b*, *c*, *x*, *y*, and *z* ions), including glycosite-localizing ions that retain the glycan, UVPD promotes cross-ring cleavages (A/X ions) of the saccharides that are instrumental for identifying the positions of the glycosidic bonds between saccharides.^{39,40} A difference in linkage position can have profound effects on protein function, its recognition of other cells, disease progression,^{5,12,41} and is imperative for understanding disease pathology.⁴²⁻⁴⁵ Therefore, the combination of mucinase proteolysis and UVPD offers a compelling strategy to advance characterization of *O*-glycoproteins, as detailed in this study.

The performance of UVPD is compared to EThcD and HCD in the present study with a focus on *O*-glycosite localization and differentiation of isomeric glycoforms. One *O*-glycoprotein targeted is TIM-1, a key member of the T-cell/transmembrane, immunoglobulin and mucin (TIM) gene family, whose mucin domain region contains 67 potential glycosites, many of which are adjacent to one another. Additionally, MUC 16 and MUC 1 are a focal point owing to their role in the human microbiota and their aberrant expression in cancer.

Materials and Methods

Materials

All solvents were obtained from VWR (Radnor, PA) unless otherwise stated. Dithiothreitol (DTT) and iodoacetamide (IAM) were obtained from Millipore Sigma. Formic acid was purchased from Fisher Chemical (Waltham, MA). *Serratia marcescens* Enhancin (SmE) was expressed and purified as described previously.¹⁹ The mucin domain proteins TIM-1 (Ser21-Gly295, KI1-H52H3), MUC16 (Pro13810-Pro14451, 76778-648), and MUC1(Met1-Ser380, 77295-302) were purchased from ACROBiosystems (Newark, DE) and VWR (Radnor, PA), respectively.

Glycoprotein Digestion

Each glycoprotein (10 μ g) was added to 50 mM ammonium bicarbonate, and SmE was added in a 1:10 ratio of enzyme:protein prior to incubation overnight at 37 °C. Samples were reduced with DTT (2 mM) at 65 °C for 20 minutes, then alkylated with IAM (5 mM) at room temperature in the dark for 15 min. The reaction was quenched with 1 μ L of formic acid. Samples were then desalted with Pierce C18 spin columns (Thermo Scientific), dried in a Speedvac (Thermo Fisher), and resuspended in 0.1% formic acid to a final concentration of 100 ng/ μ L prior to LCMS analysis.

LC-MS/MS

Peptides from SmE-digested proteins were separated and analyzed using a Dionex UltiMate 3000 high-performance liquid chromatograph (Thermo Fisher Scientific) coupled to an Orbitrap Lumos Fusion mass spectrometer (Thermo Scientific Instruments) modified with an excimer laser (193 nm, 500 Hz) from Coherent, Inc. to enable UVPD in the dual linear ion trap of the instrument. Trap (5 cm length, 3 μ m particles, 120 Å pore size, ID 100 μ m) and analytical columns (20 cm length, 3 μ m particles, 120 Å pore size, ID 75 μ m) were packed in-house with UChrom C18 stationary phase (NanoLCMS Solutions, Oroville, CA). 1-4 μ L of a 100 ng/ μ L digest was injected and separated using a gradient consisting of mobile phase A (water with 0.1% formic acid) and mobile phase B (acetonitrile with 0.1% formic acid). Representative LC traces are shown in **Figures S1-S3**. MS1 scans were acquired at an Orbitrap resolution of 30,000 at m/z 400. The resolution was set to 60,000, and the AGC target was set to 5E4 for all MS2 spectra. If a precursor subjected to HCD produced an oxonium ion common to O-glycans (i.e. m/z 204.08, 292.10, etc.), and these ions were within the top 20 most abundant ion peaks in the spectrum, then a subsequent UVPD spectrum was acquired for the same precursor using 5 laser pulses of 3 mJ energy. This approach was applied to the top three precursor peaks from each MS1 scan. All data were acquired in triplicate.

Data Analysis

MS data were searched with Byonic (v4.3.4) using a list of common mammalian O-glycans to verify O-glycan composition on glycopeptides. **Scheme S1** shows a set of the symbols used for common saccharides and representative glycan linkages. Two glycan modifications were allowed per peptide and up to 6 missed cleavages were allowed. Variable modifications included the oxidation of methionine (+15.99 Da) and deamidation of asparagine and glutamine (+0.98 Da). Precursors were allowed a 10-ppm error tolerance while fragment ions were retained if they fell within a 20-ppm mass tolerance. A 1% FDR was used for every search. Identified glycopeptides by HCD, EThcD, and UVPD were considered valid at a PEP 2D score less than 0.01 and a Byonic score above 300. Glycopeptides identified in Byonic were subject to manual validation using GlycoWorkbench (v1.2.4105), as detailed previously.⁴⁰ Core saccharide

linkages (e.g. Gal β 1-3GalNAc) were not evaluated in this study. Venn diagrams were generated using Venny 2.1.⁴⁶

Results and Discussion

Characterization of TIM-1

T cell/transmembrane immunoglobulin and mucin protein-1 or TIM-1 is expressed on the surface of T cells and plays a role in tumor immunity.⁴⁷ Containing a mucin domain spanning residues 109 to 259, a 151-residue section in total, TIM-1 has 67 potential glycosites, meaning just under 45% of the amino acids are serine or threonine residues. Most of these glycosites are adjacent or proximal to one another; in some cases, four sites exist sequentially. Thus, TIM-1 represents a challenging benchmark protein for characterization. After SmE digestion, 691 glycopeptides were separated by nanoLC and identified and characterized by HCD, EThcD, and UVPD. Examples of the MS/MS spectra are shown for one representative peptide in **Figure S4** (HCD and EThcD) and **Figure 1** (with additional ions labelled in **Figure S5**). Consistent with prior reports, HCD primarily generates *b*- and *y*-type ions and offers limited structural information about the glycan. EThcD produces a more extensive array of fragment ions, including *c*- and *z*-type ions, and is known for localizing *O*-glycosites because the glycans may be preserved on the resulting fragment ions. UVPD yields peptide sequence ions, including ones that retain the entire glycan or partial glycan, in addition to glycan ions which reveal compositional information. For each glycopeptide, the peptide sequence and glycan composition at each site were identified using Byonic.

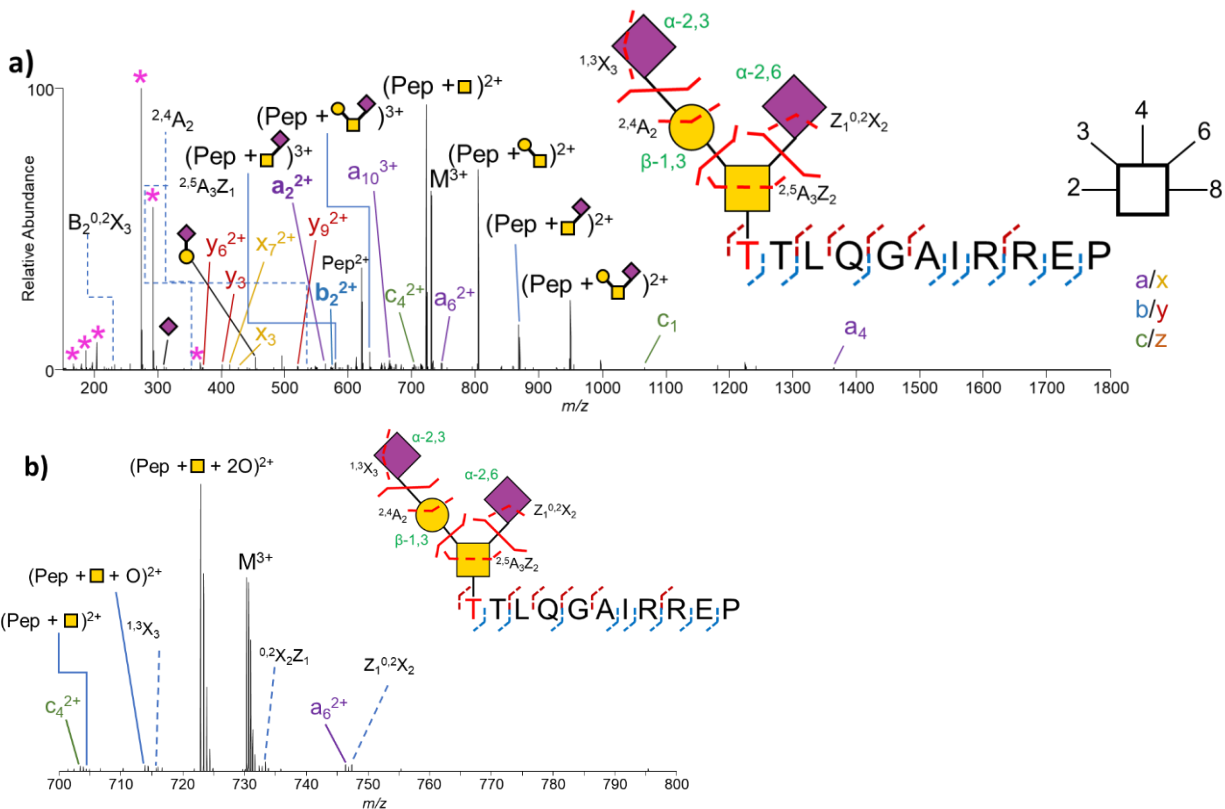


Figure 1. a) UVPD spectrum (3 mJ, 5 pulses) of TT[+947.23]LQGAIIRREP (3+, m/z 730.34) from TIM-1 that shows the capability of UVPD to differentiate glycan linkages. Representative glycosidic cleavages (solid red lines) and cross-ring cleavages (dashed red lines) are shown on the expanded glycan structure with cross-ring cleavages that confirm the linkage positions. Pink asterisks represent oxonium ions. ~ represent fragment ions that are missing complete or partial glycan information. An expanded section of this spectrum is shown in panel b. All fragment ions are summarized in **Table S1**.

Among the 691 glycopeptides analyzed, approximately 20% (136 glycopeptides) were identified by all three MS/MS methods. HCD identified 55 unique glycopeptides, while UVPD and EThcD identified 105 and 245, respectively (**Figure S6**). Inspection of the glycan compositions revealed a considerable number of core 3 glycans (i.e., containing a GalNAc and a GlcNAc on the non-reducing end) (**Figure S7**). For each MS/MS method, core 3 glycans were the most dominant structure identified, outnumbering core 1 and core 2 glycans as well as simple GalNAc glycans. The site that showed the most microheterogeneity was S253, for which EThcD identified 9 different glycan compositions, followed by S171 with 8 glycan structures EThcD (**Figure S8**). For these same two sites, UVPD identified 6 glycan compositions for S253 and 7 for S171, while HCD identified 3 glycan compositions for S253 and 6 for S171.

For some glycan configurations, multiple structures are possible as exemplified by the composition of a trisaccharide containing a single GalNAc, galactose, and sialic acid (NHA), for which the saccharides

may exist as a linear chain or alternatively both the galactose and sialic acid may branch from the GalNAc. Aside from the overall structure, the linkage of the Neu5Ac can also differ, presenting further complications in the complete characterization of a glycoform. The added complexity of linkage heterogeneity convolutes characterization and is often left ambiguous based on MS/MS analysis. Moreover, glycans with the same composition but different structures can exist for the same glycosite. For example, the glycopeptide TTLQGAI RREP has two possible adjacent *O*-glycosites (both T) and a range of possible glycoforms for each site. **Figure 1** showcases the characterization of this glycopeptide by UVPD which produces both peptide sequence ions and glycan compositional ions. A disialylated core 1 glycan is located at T1; the b_1 , a_2^{2+} , b_2^{2+} , and z_{11} ions help localize the glycan. The spectrum shows sequential loss of saccharides, further confirming the glycan composition, along with a few fragment ions that help confirm the linkages of the two Neu5Ac. The first Neu5Ac is α -linked between the second carbon on the Neu5Ac and the third carbon on the Gal. The detection of the $^{2,4}A_2$ ion supports this assignment, as the product only retains the third and fourth carbons of Gal. The second Neu5Ac is normally assumed to be α -linked between the second carbon on the Neu5Ac and the sixth carbon on the GalNAc, should a Gal already be linked to the GalNAc. The combination of a cross-ring cleavage and glycosidic cleavage, yielding the product $^{2,5}A_3Z_2$, helps confirm the $\alpha 2,6$ linkage between Neu5Ac and GalNAc, increasing confidence in the exact structure of the glycan. UVPD differentiates the subtle differences between linkage positions while also providing glycan compositional information and peptide sequence ions.

Isomeric glycopeptides can also be differentiated by UVPD, as demonstrated for the glycopeptide TMSIPT (**Figure S9**). This bis-glycosylated peptide has three glycosites (two Thr and one Ser) and several glycoforms are possible. **Figure S9** depicts the UVPD spectra acquired for the 2+ charge state of the six amino acid peptide with a total glycan composition of HexNAc(3)Hex(1). Based on the fragment ions generated by UVPD, Byonic identified the glycosites as S3 and T6 occupied by (1) GalNAcGlcNAcGal and GalNAc and (2) GalNAcGlcNAc and GalNAcGal, respectively. An array of fragment ions was produced for the peptide, in addition to a few corresponding to the peptide sequence without any glycan contributions in the lower m/z range. The dominant ions in the higher m/z range matched both peptide and glycan ions. Diagnostic ions that were crucial to characterizing these two glycoforms were the y_2 ion (in **Figure S9**) and the b_4 ions identified in both **Figure S9a** and **S9b**. The y_2 ion in confirms a single GalNAc positioned at T6 and the two b_4 ions that have a difference of 162.05 Da, corresponding to galactose, indicating that the spectrum in **Figure S9a** contains a core 2 glycan and the spectrum in **Figure S9b** contains a core 3 glycan at position S3.

Characterization of large glycans of MUC-16

Heavily implicated in ovarian cancer, MUC-16 is a large mucin protein (~1.5 MDa, Accession number: Q8WXI7-1) whose *O*-glycosylation remains under-characterized owing to its vast micro- and macroheterogeneity. The extensive fragmentation afforded by UVPD is crucial when dissecting the *O*-glycosylation of MUC-16, especially for analysis of glycopeptides that contain large glycans for which the glycan moiety makes up a significant portion. Glycopeptides detected in low charge states while containing large glycans are less suited for characterization using conventional methods like HCD and EThcD owing to the complexity of the glycans and low charge density of the ions, thus reinforcing the need for other MS/MS methods like UVPD. To explore this challenge, the *O*-glycopeptides generated upon SmE digestion of the SEA domain of MUC-16 (spanning residues Pro13810-Pro14451 of full-length MUC-16) was investigated with UVPD, HCD, and EThcD.

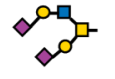



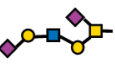
Figure S10 compares the number of unique glycoforms identified for the three MS/MS methods, with UVPD yielding 75 compared to 30 and 31 for HCD and EThcD, respectively. The lower number of glycoforms for all three methods relative to the number identified for TIM1 is due in part to the lower frequency of *O*-glycosylation of the SEA domain of MUC-16.⁴⁸ UVPD identified a higher number of large glycans than that of HCD and EThcD combined (**Figure S11**).

Given the array of large glycans, MUC-16 was explored for deeper analysis of the glycoforms, including site-by-site analysis of glycans with contributions to the overall glycopeptide mass of 859 Da or more (representing glycans containing four or more saccharides). Not only was each glycosite analyzed for different structural patterns of the same glycan composition but also the linkages of identical glycan compositions were dissected. The characterization of large *O*-glycans is notoriously difficult as there are several combinations of glycan-specific ions that correspond to the same *m/z* value, such as cross-ring cleavages associated with localizing Neu5Ac linkages to Gal, resulting in ^{1,3}A₂ and ^{0,4}A₂ ions, each having *m/z* values of 352.12). To illustrate the importance of cross-ring cleavages, **Figure S12** compares UVPD and EThcD spectra for TFLPPLS[+1021.3598]EA (2+, *m/z* 998.44), a glycopeptide containing a pentasaccharide. Unique diagnostic cross-ring cleavage ions were identified for each specific structure in the UVPD spectrum, such as the (^{1,3}X_{2L})²⁺ ion (*m/z* 968.43) of GalNAcGlcNAcGalNeu5Ac that confirms an α2,6 linkage of Neu5Ac to Gal for TFLPPLS[+1021.35]EA (2+, *m/z* 998.44) in **Figure S12a**. In addition, glycosidic fragment ions such as B₄ and C₄ (*m/z* 819.29 and 837.30, respectively) confirm single branch, core 1 glycans with GalNAcGlcNAcGal(2)Neu5Ac composition. For example, the *O*-glycan composition of GalNAcGlcNAcGal(2)Neu5Ac often remains structurally ambiguous after MS/MS analysis owing to the various combinations of branching and linkage patterns that can arise, some of which are illustrated in **Table 1**. The fragmentation pattern generated by UVPD provides sufficient information to elucidate the exact structure of the glycan, thus disambiguating the branching pattern and linkage positions. The

fragmentation pattern from EThcD (**Figure S12b**) is sparser and lacks sufficient fragment ions to deduce linkages between the saccharides. UVPD unlocks the potential for in-depth *O*-glycoproteomics analysis; however, even more sophisticated software tools are needed to streamline data analysis for higher throughput workflows. The current data analysis strategy requires rigorous manual validation for UVPD data.

A sample of the identified *O*-glycans of MUC-16 derived from UVPD analysis are summarized in **Table 1**, which groups the exact structures and compositions of each glycan identified based on linkages and glycosites, yielding comprehensive analysis of large glycoforms of MUC-16 not previously attained (full table is **Table S4**). Glycoforms can be distinguished from one another down to the level of exact linkage patterns as exemplified by the differentiation of GalNAcGlcNAcGal(2)Neu5Ac(2) glycans for TLDRDSLYVNGF (**Table 1**, rows 2 and 3). The number of times each glycopeptide was identified was tallied to investigate whether specific core types were preferred over others at a given glycosite. Among the glycoforms identified, T[+1021.35]FLPPL showed preference for an elongated core 1 structure, with linkages α Neu5Ac2-6Gal1-4GlcNAc1-6Gal1-3GalNAc for this pentasaccharide. This glycan was identified at this glycosite eight times based on Byonic assignment and manual validation. The glycan identities themselves can demonstrate site preferences too. For example, the majority of glycans were localized at T14058 compared to other glycosites. While there is less ambiguity in characterization of smaller glycans, it is still critical to differentiate the linkages which might be crucial correlative biomarkers for disease states. For instance, for the doubly sialylated core 1 glycan, the α Neu5Ac2-6Gal linkage was identified nine times across five glycosites compared to seven times for the α Neu5Ac2-3Gal linkage. The significant diversity identified for the glycans underscores the need for comprehensive characterization, especially in the case of *O*-glycoproteomics.

Table 1. Glycoforms for MUC-16 identified by UVPD. Glycoforms are grouped by their composition, the glycopeptide for which they were identified, and then listed in order of glycosite. Multiple glycoforms were identified at each site. Repeated identification of identical glycopeptides is denoted by the number in the right-most column. Core linkages were assumed (β 1,3 for Gal to GalNAc, β 1,6 for GlcNAc to GalNAc) and α 2,6 linkage was assumed for Neu5Ac to GalNAc. A check mark refers to the presence of that linkage in the glycan and an x means that that linkage was not detected for the glycan. In the “Linkages” column, A refers to sialic acid linkages, H refers to linkages of galactose, and N refers to GlcNAc linkages. ^a glycan position relative to complete protein sequence for MUC-16. Full table is given as Table S4.

Glycopeptide Sequence Glycan Composition Glycosite Position ^a	Linkages						Structure (# Repeats)
	α 2,3 (A)	α 2,6 (A)	β 1,3 (H)	β 1,4 (H)	β 1,3 (N)	β 1,6 (N)	
TGPGLDREQLYLELSQLTHSI GalNAcGlcNAcGal(2)Neu5Ac(2) 13869	✓✓	x	x	✓	NA	NA	 (1)
TLDRDSLYVNGF GalNAcGlcNAcGal(2)Neu5Ac(2) 13896	✓✓	x	x	✓	NA	NA	 (1)
TLDRDSLYVNGF GalNAcGlcNAcGal(2)Neu5Ac(2) 13896	✓	✓	x	✓	NA	NA	 (1)
TFLPPL GalNAcGlcNAcGal(2)Neu5Ac(2) 14058	✓	✓	x	✓	NA	NA	 (2)
TFLPPL GalNAcGlcNAcGal(2)Neu5Ac(2) 14058	✓	x	x	✓	x	✓	 (2)

Examination of *O*-glycosylation of MUC-1

The Tn antigen, a simple glycoconjugate composed of GalNAc alpha-O-linked to Ser or Thr, is expressed in nearly every carcinoma.⁴⁹ In cancer cells, MUC-1 is the most common protein that bears the Tn epitope and is the second most common tumor-associated antigen for cancer vaccine targets.⁵⁰ The glycosylation pattern of MUC-1 diverges significantly in cancer cells, showing an increase in the frequency of the Tn antigen motif.⁵¹ Understanding normal patterns of glycosylation can assist in the determination of glycosites altered in tumor settings and help identify predictive biomarkers. Thus, *O*-glycosylation of MUC-1 was evaluated after SmE digestion, similar to the strategy used for MUC-16. **Figure 2** compares the glycan compositions identified based on analysis of HCD, EThcD, and UVPD spectra. MUC-1 demonstrates more varied *O*-glycosylation than MUC-16 or TIM-1, revealing double the number of glycan compositions. For the SmE digests of TIM-1 and MUC-16, there were some glycopeptides suspected to contain *N*-glycolylneuraminic acid (Neu5Gc), but the MS/MS spectra did not meet the criteria to be

without an oxygen atom (m/z 331.13 and 368.15, respectively), and those that provide both glycan and peptide information such as the doubly charged ion of m/z 964.42, which consists of the peptide sequence, GalNAcGlcNAcFuc and two oxygens from Y-type glycosidic cleavages between GalNAc and Gal and GlcNAc and Gal.

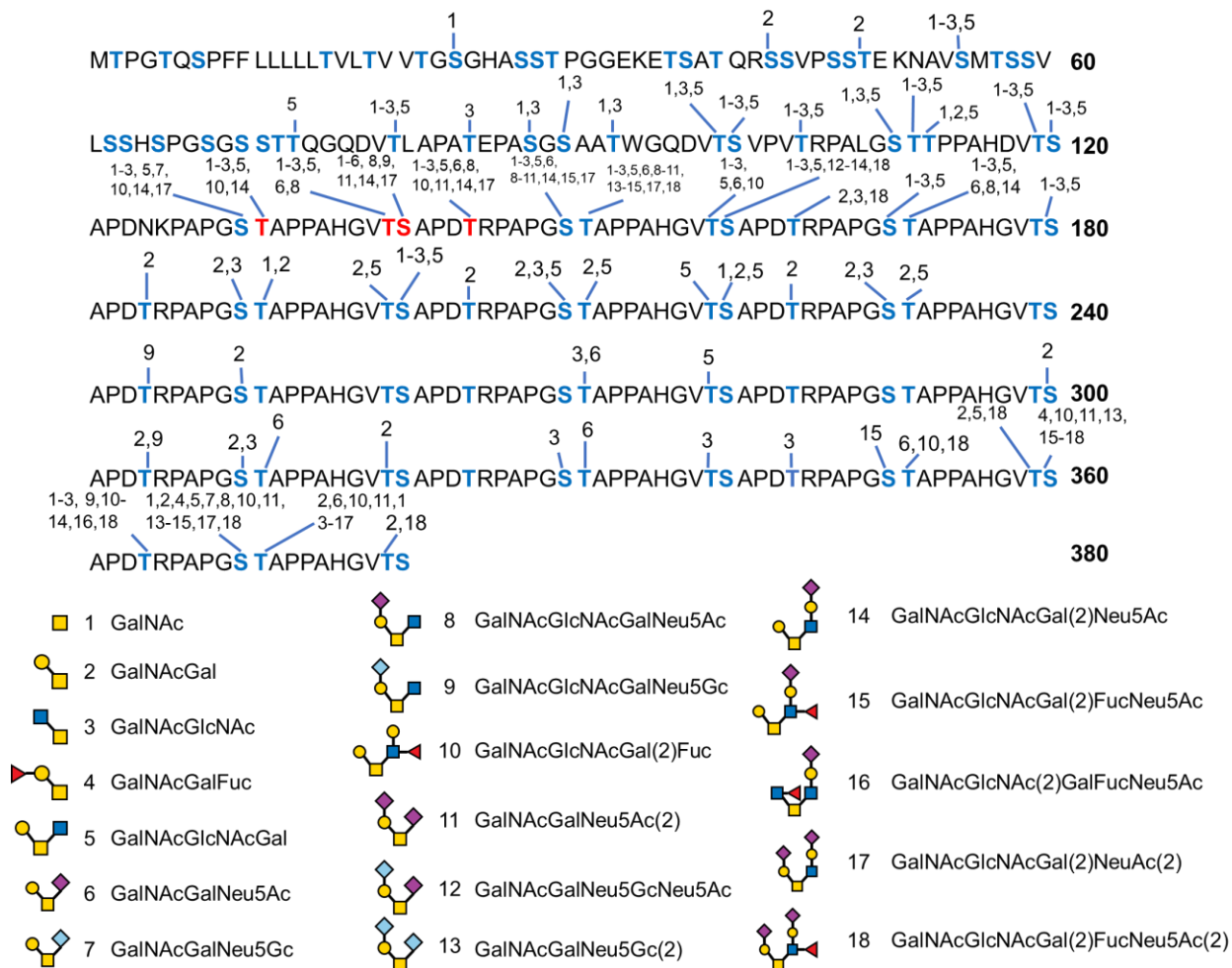


Figure 3. Identified glycosites and their corresponding glycans for MUC-1 with UVPD (2 mJ, 5 pulses). A legend is included at the bottom of the figure, including representative depictions of theoretical structures of the compositions. Amino acids highlighted in red are *O*-glycosites that are confirmed on Uniprot. Those in blue are predicted *O*-glycosites from NetOGlyc 4.0.

Figure 3 details the pattern of *O*-glycosylation of MUC-1 determined by UVPD; similar figures for EThcD and HCD are shown in **Figure S15a** and **15b**, respectively. Based on UVPD, extensive glycosylation is detected from S130 through S180, with some residues shown to be occupied by up to 13 glycan different compositions. This trend of heavy and heterogeneous glycosylation over this span of residues and then again from S360 through T371 is replicated for all three MS/MS methods. The tandem

repeats located between the two heavily glycosylated regions show significantly less glycosylation, which might partly be due to the accessibility of those particular putative glycosites and their ambiguity as repeats within the protein sequence. However, because different glycosites and glycans are identified for the glycopeptides originating from the 20 residue repeats, it suggests there is true site-specific diversity among the 13 consecutive repeats. Simple core 1 glycans were detected throughout the tandem repeats with the greatest frequency between the regions of highest glycodiversity (S130-S180 and S360-T371). Despite the lack of heavy glycosylation through the tandem repeats of MUC-1, the fact that significant glycodiversity was uncovered using the SmE digestion/UVPD strategy suggests that this is a promising approach.

Conclusions

Detailed characterization of *O*-glycoproteins is imperative to answering questions related to how glycosylation varies as a function of disease state. As demonstrated for analysis of glycopeptides produced by SmE digestion of *O*-glycoproteins TIM-1, MUC-16 and MUC-1, UVPD excels in providing structural insight about the linkages of individual *O*-glycans and elucidates overall patterns of glycosylation, even revealing in some cases the number of occurrences of certain glycan linkages. Harvesting the full potential of UVPD mass spectra will require development of more advanced software tools to help streamline the data interpretation workflow, as manual validation of the highly heterogeneous *O*-glycoproteoforms of mucin proteins is incredibly tedious. Nonetheless, the combination of SmE mucinase and UVPD has proved effective for *O*-glycopeptide characterization, allowing confirmation of a variety of structures for large and small glycoforms alike. As new software tools emerge, this integrated workflow should be applicable to analysis of complex *O*-glycoproteomes in a high-throughput manner.

Acknowledgments

This research was supported by the National Institutes of Health (R35GM13965) and the Robert A. Welch Foundation [F-1155]. V.C. is supported by an NSF GRFP award.

Supporting Information:

The supporting information contains additional UVPD spectra of *O*-glycopeptides from TIM-1, MUC-1, and MUC-16, chromatograms from LC-MS experiments, and various tables listing identified peptide fragments and glycan fragments for each spectrum.

References

- (1) Brockhausen, I.; Wandall, H. H.; Ten Hagen; Stanley, P. O-GalNAc Glycans. In *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor NY, 2022.
- (2) Varki, A. Biological Roles of Glycans. *Glycobiology* **2017**, 27 (1), 3–49.
<https://doi.org/10.1093/glycob/cww086>.

- (3) Wandall, H. H.; Nielsen, M. A. I.; King-Smith, S.; de Haan, N.; Bagdonaite, I. Global Functions of O-Glycosylation: Promises and Challenges in O-Glycobiology. *The FEBS Journal* **2021**, *288* (24), 7183–7212. <https://doi.org/10.1111/febs.16148>.
- (4) Li, Z.; Chai, W. Mucin O-Glycan Microarrays. *Curr. Opin. in Struc. Biol.* **2019**, *56*, 187–197. <https://doi.org/10.1016/j.sbi.2019.03.032>.
- (5) Cervoni, G. E.; Cheng, J. J.; Stackhouse, K. A.; Heimbürg-Molinario, J.; Cummings, R. D. O-Glycan Recognition and Function in Mice and Human Cancers. *Biochem. J.* **2020**, *477* (8), 1541–1564. <https://doi.org/10.1042/BCJ20180103>.
- (6) Wu, D.; Robinson, C. V. Understanding Glycoprotein Structural Heterogeneity and Interactions: Insights from Native Mass Spectrometry. *Curr. Opin. in Struc. Biol.* **2022**, *74*, 102351. <https://doi.org/10.1016/j.sbi.2022.102351>.
- (7) Barkeer, S.; Chugh, S.; Batra, S. K.; Ponnusamy, M. P. Glycosylation of Cancer Stem Cells: Function in Stemness, Tumorigenesis, and Metastasis. *Neoplasia* **20** (8), 813–825. <https://doi.org/10.1016/j.neo.2018.06.001>.
- (8) Berger, R. P.; Dookwah, M.; Steet, R.; Dalton, S. Glycosylation and Stem Cells: Regulatory Roles and Application of iPSCs in the Study of Glycosylation-Related Disorders. *BioEssays* **2016**, *38* (12), 1255–1265. <https://doi.org/10.1002/bies.201600138>.
- (9) Cox, K. E.; Liu, S.; Lwin, T. M.; Hoffman, R. M.; Batra, S. K.; Bouvet, M. The Mucin Family of Proteins: Candidates as Potential Biomarkers for Colon Cancer. *Cancers (Basel)* **2023**, *15* (5), 1491. <https://doi.org/10.3390/cancers15051491>.
- (10) Seeberger, P. H. Monosaccharide Diversity. In *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor NY, 2022.
- (11) Stanley, P.; Moremen, K. W.; Lewis, N. E.; Taniguchi, N.; Aebi, M. N-Glycans. In *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor NY, 2022.
- (12) Wilkinson, H.; Saldova, R. Current Methods for the Characterization of O-Glycans. *J. Proteome Res.* **2020**, *19* (10), 3890–3905. <https://doi.org/10.1021/acs.jproteome.0c00435>.
- (13) Hansen, J. E.; Lund, O.; Tolstrup, N.; Gooley, A. A.; Williams, K. L.; Brunak, S. NetOglyc: Prediction of Mucin Type O-Glycosylation Sites Based on Sequence Context and Surface Accessibility. *Glycoconj J.* **1998**, *15*, 115–130. <https://doi.org/10.1023/a:1006960004440>.
- (14) Halim, A.; Brinkmalm, G.; Rüetschi, U.; Westman-Brinkmalm, A.; Protelius, E.; Zetterberg, H.; Blennow, K.; Larson, G.; Nilsson, J. Site-Specific Characterization of Threonine, Serine, and Tyrosine Glycosylations of Amyloid Precursor Protein/Amyloid Beta-Peptides in Human Cerebrospinal Fluid. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (29), 11848–11853. <https://doi.org/10.1073/pnas.1102664108>.
- (15) Malaker, S. A.; Pedram, K.; Ferracane, M. J.; Bertozzi, C. R. The Mucin-Selective Protease StcE Enables Molecular and Functional Analysis of Human Cancer-Associated Mucins. *PNAS* **2019**, *116* (15), 7278–7287. <https://doi.org/10.1073/pnas.1813020116>.
- (16) Riley, N. M.; Malaker, S. A.; Bertozzi, C. R. Electron-Based Dissociation Is Needed for O-Glycopeptides Derived from OPERATOR Proteolysis. *Anal. Chem.* **2020**, *92* (22), 14878–14884. <https://doi.org/10.1021/acs.analchem.0c02950>.
- (17) Haurat, M. F.; Scott, N. E.; Venanzio, G. D.; Lopez, J.; Pluvinaige, B.; Boraston, A. B.; Ferracane, M. J.; Feldman, M. F. The Glycoprotease CpaA Secreted by Medically Relevant Acinetobacter Species Targets Multiple O-Linked Host Glycoproteins. *mBio* **2020**, *11* (5). <https://doi.org/10.1128/mbio.02033-20>.
- (18) Taleb, V.; Liao, Q.; Narimatsu, Y.; García-García, A.; Compañón, I.; Borges, R. J.; González-Ramírez, A. M.; Corzana, F.; Clausen, H.; Rovira, C.; Hurtado-Guerrero, R. Structural and Mechanistic Insights into the Cleavage of Clustered O-Glycan Patches-Containing Glycoproteins by Mucinas of the Human Gut. *Nat Commun* **2022**, *13*. <https://doi.org/10.1038/s41467-022-32021-9>.
- (19) Chongsaritsinsuk, J.; Steigmeyer, A. D.; Mahoney, K. E.; Rosenfeld, M. A.; Lucas, T. M.; Ince, D.; Kearns, F. L.; Battison, A. S.; Hollenhorst, M. A.; Shon, D. J.; Tiemeyer, K. H.; Attah, V.; Kwon, C.; Bertozzi, C. R.; Ferracane, M. J.; Amaro, R. E.; Malaker, S. A. Glycoproteomic Landscape and

- Structural Dynamics of TIM Family Immune Checkpoints Enabled by Mucinase SmE. *Nat Commun* **2023**, *14*, 6169. <https://doi.org/10.1038/s41467-023-41756-y>.
- (20) Sanz-Martinez, I.; Pereira, S.; Merino, P.; Corzana, F.; Hurtado-Guerrero, R. Molecular Recognition of GalNAc in Mucin-Type O-Glycosylation. *Acc. Chem. Res* **2023**, *56* (5), 548–560. <https://doi.org/10.1021/acs.accounts.2c00723>.
- (21) Konstantinidi, A.; Nason, R.; Čaval, T.; Sun, L.; Sørensen, D. M.; Furukawa, S.; Ye, Z.; Vincentelli, R.; Narimatsu, Y.; Vakhrushev, S. Y.; Clausen, H. Exploring the Glycosylation of Mucins by Use of O-Glycodomain Reporters Recombinantly Expressed in Glycoengineered HEK293 Cells. *Journal of Biological Chemistry* **2022**, *298* (4), 101784. <https://doi.org/10.1016/j.jbc.2022.101784>.
- (22) 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.; Iverson, T. M.; Bensing, B. A.; Sullam, P. M.; Varki, A.; de Vries, E.; de Haan, C. A. M.; Vincentelli, R.; Henrissat, B.; Vakhrushev, S. Y.; Clausen, H.; Narimatsu, Y. Display of the Human Mucinome with Defined O-Glycans by Gene Engineered Cells. *Nat Commun* **2021**, *12*, 4070. <https://doi.org/10.1038/s41467-021-24366-4>.
- (23) Yang, Y.; Yin, Y.; Xu, W.; Kang, Y.; Chen, J.; Zou, Y.; Xiao, Z.; Li, Z.; Cao, P. Restoring O-Glycosylation and Expression of MUC2 Limits Progression of Colorectal Cancer. *bioRxiv* **2024**. <https://doi.org/10.1101/2024.01.25.577208>.
- (24) Wardman, J. F.; Sim, L.; Liu, J.; Howard, T. A.; Geissner, A.; Danby, P. M.; Boraston, A. B.; Wakarchuk, W. W.; Withers, S. G. A High-Throughput Screening Platform for Enzymes Active on Mucin-Type O-Glycoproteins. *Nature Chemical Biology* **2023**, *19*, 1246–1255. <https://doi.org/10.1038/s41589-023-01405-3>.
- (25) Park, S.; Colville, M. J.; Shurer, C. R.; Huang, L.-T.; Kuo, J. C.-H.; Paek, J. H.; Goudge, M. C.; Su, J.; DeLisa, M. P.; Lammerding, J.; Zipfel, W. R.; Fischbach, C.; Reesink, H. L.; Paszek, M. J. Mucins Form a Nanoscale Material Barrier against Immune Cell Attack. *bioRxiv* **2022**. <https://doi.org/10.1101/2022.01.28.478211>.
- (26) Kim, J.; Ryu, C.; Ha, J.; Lee, J.; Kim, D.; Ji, M.; Park, C.-S.; Lee, J.; Kim, D.-K.; Kim, H.-H. Structural and Quantitative Characterization of Mucin-Type O-Glycans and the Identification of O-Glycosylation Sites in Bovine Submaxillary Mucin. *Biomolecules* **2020**, *10* (4), 636. <https://doi.org/10.3390/biom10040636>.
- (27) Behera, S. K.; Praharaj, A. B.; Dehury, B.; Negi, S. Exploring the Role and Diversity of Mucins in Health and Disease with Special Insight into Non-Communicable Diseases. *Glycoconj J* **2015**, *32* (8), 575–613. <https://doi.org/10.1007/s10719-015-9606-6>.
- (28) Ratan, C.; K. D., D. C.; Nair, B.; Nath, L. R. MUC Glycoproteins: Potential Biomarkers and Molecular Targets for Cancer Therapy. *Current Cancer Drug Targets* **2020**, *21* (2), 132–152. <https://doi.org/10.2174/1568009620666201116113334>.
- (29) Hansson, G. C. Mucins and the Microbiome. *Annual Review of Biochemistry* **2020**, *89*, 769–793. <https://doi.org/10.1146/annurev-biochem-011520-105053>.
- (30) Sanda, M.; Morrison, L.; Goldman, R. N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal. Chem.* **2021**, *93* (4), 2003–2009. <https://doi.org/10.1021/acs.analchem.0c03173>.
- (31) Liu, D.; Wang, S.; Zhang, J.; Xiao, W.; Miao, C. H.; Konkle; Wan, X.-F.; Li, L. Site-Specific N- and O-Glycosylation Analysis of Human Plasma Fibronectin. *Front. Chem.* **2021**, *9*, 691217. <https://doi.org/10.3389/fchem.2021.691217>.
- (32) Riley, N. M.; Malaker, S. A.; Driessen, M. D.; Bertozzi, C. R. Optimal Dissociation Methods Differ for N- and O-Glycopeptides. *J. Proteome Res.* **2020**, *19* (8), 3286–3301. <https://doi.org/10.1021/acs.jproteome.0c00218>.
- (33) Zhang, Y.; Xie, X.; Zhao, X.; Tian, F.; Lv, J.; Ying, W.; Qian, X. Systems Analysis of Singly and Multiply O-Glycosylated Peptides in the Human Serum Glycoproteome via ETHcD and HCD Mass Spectrometry. *J. Proteomics* **2018**, *170*, 14–27. <https://doi.org/10.1016/j.jprot.2017.09.014>.
- (34) Chen, Z.; Wang, D.; Yu, Q.; Johnson, J.; Shipman, R.; Zhong, X.; Huang, J.; Yu, Q.; Zetterberg, H.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L. In-Depth Site-Specific O-Glycosylation Analysis of

- Glycoproteins and Endogenous Peptides in Cerebrospinal Fluid (CSF) from Healthy Individuals, Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) Patients. *ACS Chem. Biol.* **2022**, *17* (11), 3059–3068. <https://doi.org/10.1021/acscchembio.1c00932>.
- (35) Li, X.; Wilmanowski, R.; Gao, X.; VanAernum, Z. L.; Donnelly, D. P.; Kochert, B.; Schuessler, H. A.; Richardson, D. Precise O-Glycosylation Site Localization of CD24Fc by LC-MS Workflows. *Anal. Chem.* **2022**, *94* (23), 8416–8425. <https://doi.org/10.1021/acs.analchem.2c01137>.
- (36) Hashii, N.; Suzuki, J.; Hanamatsu, H.; Furukawa, J.-I.; Ishii-Watabe, A. In-Depth Site-Specific O-Glycosylation Analysis of Therapeutic Fc-Fusion Protein by Electron-Transfer/Higher-Energy Collisional Dissociation Mass Spectrometry. *Biologicals* **2019**, *58*, 35–43. <https://doi.org/10.1016/j.biologicals.2019.01.005>.
- (37) Shi, J.; Ku, X.; Zou, X.; Hou, J.; Yan, W.; Zhang, Y. Comprehensive Analysis of O-Glycosylation of Amyloid Precursor Protein (APP) Using Targeted and Multi-Fragmentation MS Strategy. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2021**, *1865* (10), 129954. <https://doi.org/10.1016/j.bbagen.2021.129954>.
- (38) Brodbelt, J. S.; Morrison, L. J.; Santos, I. Ultraviolet Photodissociation Mass Spectrometry for Analysis of Biological Molecules. *Chem. Rev.* **2020**, *120* (7), 3328–3380. <https://doi.org/10.1021/acs.chemrev.9b00440>.
- (39) Escobar, E. E.; Wang, S.; Goswami, R.; Lanzillotti, M. B.; Li, L.; McLellan, J. S.; Brodbelt, J. S. Analysis of Viral Spike Protein N-Glycosylation Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2022**, *94* (15), 5776–5784. <https://doi.org/10.1021/acs.analchem.1c04874>.
- (40) Helms, A.; Escobar, E. E.; Vainauskas, S.; Taron, C. H.; Brodbelt, J. S. Ultraviolet Photodissociation Permits Comprehensive Characterization of O-Glycopeptides Cleaved with O-Glycoprotease IMPa. *Anal. Chem.* **2023**, *95* (24), 9280–9287. <https://doi.org/10.1021/acs.analchem.3c01111>.
- (41) Bagdonaite, I.; Malaker, S. A.; Polasky, D. A.; Riley, N. M.; Schjoldager, K. T.; Vakhrushev, S. Y.; Halim, A.; Aoki-Kinoshita, K. F.; Nesvizhskii, A. I.; Bertozzi, C. R.; Wandall, H. H.; Parker, B. L.; Thaysen-Andersen, M.; Scott, N. E. Glycoproteomics. *Nat Rev Methods Primers* **2022**, *2* (48). <https://doi.org/10.1038/s43586-022-00128-4>.
- (42) Bagdonaite, I.; Pallesen, E. M. H.; Nielsen, M. I.; Bennett, E. P.; Wandall, H. H. Mucin-Type O-GalNAc Glycosylation in Health and Disease. In *The Role of Glycosylation in Health and Disease; Advances in Experimental Medicine and Biology*; Springer, 2021; Vol. 1325, pp 25–60.
- (43) Bhat, A. H.; Maity, S.; Giri, K.; Ambatipudi, K. Protein Glycosylation: Sweet or Bitter for Bacterial Pathogens? *Critical Reviews in Microbiology* **2019**, *45* (1), 82–102. <https://doi.org/10.1080/1040841X.2018.1547681>.
- (44) Costa, A. F.; Campos, D.; Reis, C. A.; Gomes, C. Targeting Glycosylation: A New Road for Cancer Drug Discovery. *Trends in Cancer* **2020**, *6* (9), 757–766. <https://doi.org/10.1016/j.trecan.2020.04.002>.
- (45) Kobeissy, F.; Kobaisi, A.; Peng, W.; Barsa, C.; Goli, M.; Sibahi, A.; Hayek, S. E.; Abdelhady, S.; Haidar, M. A.; Sabra, M.; Orešič, M.; Logroscino, G.; Mondello, S.; Eid, A. H.; Mehref, Y. Glycomic and Glycoproteomic Techniques in Neurodegenerative Disorders and Neurotrauma: Towards Personalized Markers. *Cells* **2022**, *11* (3), 581. <https://doi.org/10.3390/cells11030581>.
- (46) Oliveros, J. C. Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams. **2007**.
- (47) Du, P.; Xiong, R.; Li, X.; Jiang, J. Immune Regulation and Antitumor Effect of TIM-1. *Journal of Immunology Research* **2016**, *2016* (8605134). <http://dx.doi.org/10.1155/2016/8605134>.
- (48) Song, Y.; Yuan, M.; Wang, G. Update Value and Clinical Application of MUC16 (Cancer Antigen 125). *Expert Opinion on Therapeutic Targets* **2023**, *27* (8), 745–756. <https://doi.org/10.1080/14728222.2023.2248376>.
- (49) Ju, T.; Otto, V. I.; Cummings, R. D. The Tn Antigen—Structural Simplicity and Biological Complexity. *Angewandte Chemie International Edition* **2011**, *50* (8), 1770–1791. <https://doi.org/10.1002/anie.201002313>.

- (50) Qu, J.; Yu, H.; Li, F.; Zhang, C.; Trad, A.; Brooks, C.; Zhang, B.; Gong, T.; Guo, Z.; Li, Y.; Ragupathi, G.; Lou, Y.; Hwu, P.; Huang, W.; Zhou, D. Molecular Basis of Antibody Binding to Mucin Glycopeptides in Lung Cancer. *International Journal of Oncology* **2016**, *48* (2), 587–594. <https://doi.org/10.3892/ijo.2015.3302>.
- (51) Zhou, D.; Xu, L.; Huang, W.; Tonn, T. Epitopes of MUC1 Tandem Repeats in Cancer as Revealed by Antibody Crystallography: Toward Glycopeptide Signature-Guided Therapy. *Molecules* **2018**, *23* (6), 1326. <https://doi.org/10.3390/molecules23061326>.