



Deciphering Protein O-Glycosylation: Solid-Phase Chemoenzymatic Cleavage and Enrichment

Shuang Yang,^{*,†} Philip Onigman,[‡] Wells W. Wu,[§] Jonathan Sjogren,[‡] Helen Nyhlen,^{||} Rong-Fong Shen,[§] and John Cipollo^{*,†}

[†]Laboratory of Bacterial Polysaccharides, Division of Bacterial, Parasitic and Allergenic Products, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland 20993, United States

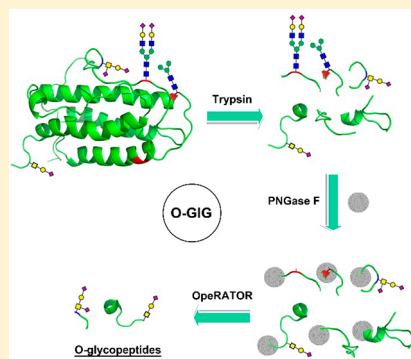
[‡]Genovis Inc., Cambridge, Massachusetts 02142, United States

[§]Facility for Biotechnology Resources, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland 20993, United States

^{||}Genovis AB, Box 790, SE-220 07 Lund, Sweden

S Supporting Information

ABSTRACT: Glycosylation plays a critical role in the biosynthetic-secretory pathway in the endoplasmic reticulum (ER) and Golgi apparatus. Over 50% of mammalian cellular proteins are typically glycosylated; this modification is involved in a wide range of biological functions such as barrier formation against intestinal microbes and serves as signaling molecules for selectins and galectins in the innate immune system. *N*-linked glycosylation analysis has been greatly facilitated owing to a range of specific enzymes available for their release. However, system-wide analysis on *O*-linked glycosylation remains a challenge due to the lack of equivalent enzymes and the inherent structural heterogeneity of *O*-glycans. Although *O*-glycosidase can catalyze the removal of core 1 and core 3 *O*-linked disaccharides from glycoproteins, analysis of other types of *O*-glycans remains difficult, particularly when residing on glycopeptides. Here, we describe a novel chemoenzymatic approach driven by a newly available *O*-protease and solid phase platform. This method enables the assignment of *O*-glycosylated peptides, *N*-glycan profile, sialyl *O*-glycopeptides linkage, and mapping of heterogeneous *O*-glycosylation. For the first time, we can analyze intact *O*-glycopeptides generated by *O*-protease and enriched using a solid-phase platform. We establish the method on standard glycoproteins, confirming known *O*-glycosites with high accuracy and confidence, and reveal up to 8-fold more glycosites than previously reported with concomitant increased heterogeneity. This technique is further applied for analysis of Zika virus recombinant glycoproteins, revealing their dominant *O*-glycosites and setting a basis set of *O*-glycosylation tracts in these important viral antigens. Our approach can serve as a benchmark for the investigation of protein *O*-glycosylation in diseases and other biomedical contexts. This method should become an indispensable tool for investigations where *O*-glycosylation is central.



Protein glycosylation is an enzymatic process in which a glycan is attached to specific amino acid residue side chains of proteins. These post-translationally attached glycans serve many roles such as cellular adhesion,¹ protein folding and quality control,² antigenic masking in viral infection,³ recognition events,⁴ and metabolic roles as well as structural roles.⁵ *N*-linked glycosylation begins in the endoplasmic reticulum (ER) with glycan maturation in the Golgi compartment,⁶ while *O*-linked glycosylation occurs throughout the Golgi. The availability of *O*-glycosyltransferases determines the ranges and amounts of specific *O*-glycans produced.⁷ Notably, it has been recognized that the pattern of glycosylation is highly similar in a given physiological state. Thus, altered glycosylation profiles in disease-associated states can provide diagnostic features.⁸ However, analysis of protein glycosylation is challenging in the extreme due to its complexity and heterogeneity.

N-glycosylation occurs at specific asparagine (Asn or N) residues in the tripeptide motif NXS/T, where X is any amino acid except for proline (S = serine; T = threonine). *O*-glycans typically link to S or T residues through an enzymatic reaction. In general, glycosylation analysis can be carried out in three phases: (a) structural analysis of glycans released from glycoproteins, (b) fingerprinting of glycosylation sites on the peptide backbone after release of resident glycans, and (c) comprehensive profiling of intact glycopeptides. To facilitate the first two analyses, glycans must be released. *N*-glycan release can be achieved by enzymatic digestion such as with PNGase F (peptide-*N*-glycosidase F) or -A, Endo (endoglycosidase) H, and others. The digestion can be performed using a

Received: April 24, 2018

Accepted: June 3, 2018

Published: June 24, 2018

traditional approach in-solution⁹ or in solid-phase with cleavage of the immobilized glycoproteins with the enzyme in column format or beads.¹⁰ The solid-phase approach enables derivatization of N-glycans for improved stabilization, better ionization,^{11,12} and sialic acid linkage determination.¹²

While a range of enzymatic tools exists for cleavage of N-glycans, release of O-glycans is problematic and is limited to short disaccharide versions of core 1/core 3 mucin-type O-glycans.^{13,14} Extended versions of these core disaccharide can only be released subsequent to digestion with glycosidases prior to O-glycanase treatment.¹⁵ Instead of using glycosidase, O-glycans are commonly released using chemical methods such as β -elimination or hydrazine hydrolysis.^{16,17} In terms of glycosylation site identification, N-glycosites can be identified after enzymatic glycan release, while O-glycosite determination is extremely limited since enzymatic releasing options are few and chemical release of O-glycans is likely detrimental to their O-glycosites. N-glycopeptides can be immobilized on the solid support via chemical oxidation of their N-glycans followed by hydrazide coupling to a solid support.¹⁸ The oxidized glycans are then cleaved from the peptide backbone enzymatically to yield the deglycosylated peptides. Alternatively, N-glycopeptides can be enriched by lectin-affinity capture and enzymatically isotope-coded and tagged after release.¹⁹ Again, these N-glycopeptide procedures are facilitated because of the availability of versatile enzymes used to release N-glycans. However, a similarly robust enzyme driven method for analysis of O-linked glycopeptides is not currently available.

Intact N-glycopeptide analysis is essentially straightforward. Standard proteases used in proteomics applications can typically be used alone or in combination to generate N-glycopeptides that are readily analyzable using standard proteomics applications with some modifications.^{20,21} Although some glycopeptide enrichment is often used,²² there has been wide success in the analysis of N-glycopeptides, and this is in part due to the low density of N-glycosites in the majority of N-glycoproteins as compared to that of O-glycopeptides. Many O-glycosylation rich proteins are heavily substituted in S/T rich regions, which prevent efficient proteolysis resulting in glycopeptides that are not easily ionizable. Clearly, new approaches must be developed that can better access the O-glycosylation rich regions while preferably incorporating N-glycopeptide analysis as well.

Substantial progress has been made for profiling of the intact glycopeptides. Metabolic labeling of the glycoproteome shows a promising approach for studying of N-linked and O-linked glycopeptides. This approach utilizes chemical enrichment and isotopic labeling of glycopeptides to select peptides for targeted glycoproteomics using mass spectrometry (MS),²³ which is useful for analysis of specific glycoprotein subtypes. In general, intact N-glycopeptides are comprehensively analyzed by hydrophilic interaction liquid chromatography (HILIC) and MS after tryptic digestion, or enrichment for glycoproteins or glycopeptides by lectin affinity prior to MS analysis. Unfortunately, it is exceptionally challenging to analyze the intact O-glycopeptides using this strategy because no enzymes can release O-glycans or digest O-glycosites. Numerous efforts have been devoted to elucidating the heterogeneity of mucin-type O-glycosylated domains, such as release of O-glycans by chemical reaction coupled with high-performance anion-exchange chromatography with pulsed electrochemical detection (HPAEC-PAD)/LC-ESI-MS (electrospray). However, such approaches often require the use of multiple proteases,

leading to heterogeneity and dispersed signal intensities.²⁴ Recently, zinc finger nuclease gene-engineered SimpleCell lines were used for identification of O-GalNAc glycosites.²⁵ Interestingly, using this method, selective purification of O-glycopeptides and N-glycopeptides is facilitated using ice-cold acetone. The rationale is that O-glycopeptides have specific behavior in ice-cold acetone compared to their counterparts such as N-glycopeptides and peptides.²⁶ HILIC enrichment of PNGase F-treated O-glycopeptides in combination with lectin affinity such as that of Jacalin and WGA (wheat germ agglutinin) can specifically enrich O-glycopeptides. However, the complexity of peptide sequence in mucin-type domains can often result in highly complex data sets since no enzyme can digest the mucin domain. Therefore, it is a challenge to analyze O-glycopeptides absent of a strategy that can cleave and enrich the O-glycosylated sites such as mucin domain. Application of such a simplified strategy could advantage MS-based methods.^{27,28}

Recent O-glycosite analytical approaches have achieved some success. O-glycosite identification is determined using reductive β -elimination with NaOH and NaBD₄.^{29,30} The reaction converts the O-glycosylated threonine residue into dehydroamino-2-butyric acid, which is further reduced to 2-aminobutyric acid deuterated in the C₂ position. However, the peptide backbone tends to degrade in the presence of reductant NaBD₄.²⁹ To prevent this issue, reagents used for Michael addition are introduced in addition to the β -elimination; yet, nonspecific dehydration of unmodified or phosphorylated serine and threonine can lead to ambiguity as to the status of the serine/threonine residues as glycans modified, phosphate modified, or unmodified are possible.³⁰ Other reagents such as methylamine have been incorporated into reductive β -elimination, resulting in site specific amination and O-glycosylation site identification. Unfortunately, the glycan information is lost.^{31,32} Identification of the relatively simple O-GlcNAc modification site on O-glycopeptides can be identified by mild β -elimination followed by Michael addition using dithiothreitol (DTT), deuterated DTT, or biotin pentylamine to tag O-GlcNAc sites.³³ Overall, these methods are highly nonspecific for O-glycopeptides. Moreover, glycan heterogeneity information are lost.³⁴ No robust method has been reported for analysis of intact O-glycosylation.

Here, we report the development of a novel method, *O*-Glycopeptide Immobilization for *O*-Glycosylated peptide enrichment (*O*-GIG), for chemoenzymatic digestion of O-glycosylated peptides, profiling N-glycans, identifying N-glycosites, determining linkages of sialyl O-glycopeptides, and mapping heterogeneous O-glycosylation on the peptide backbone. This method is facilitated via a solid support. Glycopeptides after tryptic digestion (Figure 1a) are immobilized on aldehyde-activated resin via reductive amination on N-termini of peptides (Figure 1b). N-glycans are cleaved by PNGase F digestion (Figure 1c), with an option for sequential derivatization of sialic acids prior to digestion.¹² The solid support immobilizes peptides, deamidated peptides (from N-glycan release), and O-glycopeptides, which are distinctly digested by the recently discovered O-protease, *OperATOR*, within a pH range of 6.0 to 8.8 (Figure 1d). The O-protease cleaves the peptide bond N-terminal to serine or threonine that is substituted with O-glycan, while non-O-glycosylated serine/threonine remains on the solid support. The O-glycopeptides are analyzed by LC-MS (Figure 1e) for identification. *O*-GIG is an invaluable chemoenzymatic platform for deciphering

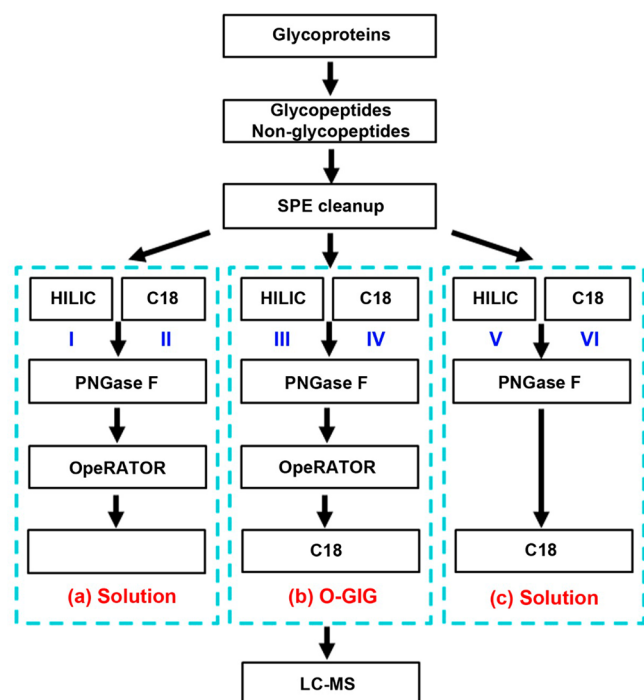


Figure 1. The schematic diagram of *O*-glycosylation analysis using solid-phase chemoenzymatic *O*-GIG. (a) Glycoproteins were digested for C-18 cleanup or HILIC (hydrophilic interaction liquid chromatography). (b) Glycopeptides were immobilized to the aldehyde-functionalized resins via reductive amination. (c) *N*-linked glycans were released by PNGase F or A. Optionally, sialic acids can be sequentially modified for stabilization and differential linkages. (d) The *O*-glycan specific enzyme (OpeRATOR) releases *O*-linked glycopeptides by hydrolyzing *N*-terminal of the glycosylated S (serine) or T (threonine). Peptides containing S or T without *O*-linked glycan attached remain on the resin. (e) The *O*-linked glycopeptides were quantitatively analyzed by LC-MS.

complex *O*-glycosylation in addition to *N*-glycosylation with high specificity and confidence. For the first time, we are capable of elucidating *O*-glycosylation and *N*-glycosylation using a chemoenzymatic solid-phase method.

EXPERIMENTAL PROCEDURES

Sample Preparation. Proteins were extracted from cells or tissues using denaturing buffer, which contains 8 M urea in 1 M NH_4HCO_3 (all chemicals were purchased from Sigma-Aldrich unless otherwise noted). Samples were sonicated for 30 s, then cooled in an ice bath for 30 s, repeated four times (note: standard glycoproteins are used as received). Proteins were reduced in 12 mM of dithiothreitol (DDT) for 1 h at 37 °C, followed by alkylation by adding 16 mM of iodoacetamide (IAA) for 30 min at room temperature in the dark.³⁵ Samples were diluted 1:5 in HPLC water to reduce the concentration of urea to <2 M. Trypsin (Promega) was added at a minimum concentration of 12.5 ng/ μL and 1:50–1:100 trypsin/protein, incubating at 37 °C for overnight. After the digest was cooled to room temperature, formic acid (FA) (100%, MS grade; Fisher) was added to acidify the digest with pH < 2.5. Samples were centrifuged at 13 000g for 10 min at room temperature, and the pellet is discarded. Peptides are purified by solid-phase extraction (SPE) cartridge.

Enzymatic Cleave of *O*-Glycosylated S or T. OpeRATOR (Genovis AB; 2000 units) was dissolved in 2000 μL of

HPLC water to have a concentration of 1 unit/ μL . The enzyme was aliquoted and stored in –20 °C prior to use. For samples on the resin, enzymes (20–50 μL) were added with reaction buffer (20 mM Tris, pH 6.8; Fisher) to 300 μL ; for peptides or glycopeptides, samples were dried in Speed-Vac (Thermo) before being resuspended in reaction buffer (the total volume was up to 300 μL). The sample–enzyme mixtures were incubated at 37 °C overnight. The digest was purified by SPE cartridge for LC-MS/MS analysis.

***O*-Glycosylation Analysis Method.** Different methods were established to compare the performance of *O*-GIG versus the in-solution approach. Figure 2 schematically shows three different approaches for analysis of *O*-glycosylation. SPE cleanup was performed on C18 to use global peptides for these methods, while HILIC was also compared to use glycopeptides. C18 cleanup was conducted using tryptic digestion samples, while HILIC enrichment required the drying of tryptic digestion samples first before resuspending in 80% ACN (acetonitrile) in 0.1% TFA (trifluoroacetic acid). Both global peptides and HILIC-enriched glycopeptides were tested using three methods. *O*-GIG immobilizes peptides on the resin via *N*-termini after peptides or glycopeptides were dissolved in 475 μL of pH 10 binding buffer (50 mM sodium carbonate and 100 mM sodium citrate, pH 10), with incubation for 3 h.¹¹ After adding 25 μL of 1 M NaCNBH_3 in DI (distilled water), peptide conjugation proceeded for another 3 h before washing with 500 μL 1 $\times$ PBS twice. The reaction was continued in 50 mM NaCNBH_3 in 500 μL 1 $\times$ PBS buffer for 3 h. After blocking aldehyde sites on the resin in 50 mM NaCNBH_3 in 1 M Tris-HCl, samples were washed by 1 M NaCl and HPLC water three times sequentially. To remove *N*-glycans, 0.5 μL of PNGase F (or 250 units; New England BioLabs Inc.) was added to the resin in 300 μL of 25 mM NH_4HCO_3 , incubated at 37 °C for 3 h. The resin was washed by 500 μL of 20% ACN and DI water three times sequentially before *O*-glycosites were digested by OpeRATOR (30 μL ; Figure 2a). To compare performance with OpeRATOR digestion in solution, SPE cleaned peptides or glycopeptides were resuspended in 300 μL of 25 mM NH_4HCO_3 and 0.5 μL of PNGase F, incubated at 37 °C for 3 h. After C18 cleanup, samples were resuspended in 300 μL of 20 mM Tris (pH 6.8) and 30 μL of OpeRATOR, incubated at 37 °C overnight. Samples were purified by C18 or HILIC, depending on whether they were C18 or HILIC enriched before PNGase F digestion (Figure 2b). To determine performance of OpeRATOR on *O*-glycopeptides, samples after SPE cleanup were also only digested by PNGase F without the use of OpeRATOR. These samples were also purified by C18 or a HILIC SPE cartridge for LC-MS/MS analysis (Figure 2c).

LC-MS/MS. Peptides (1 μg) were analyzed by LC/MS/MS using a Thermo Fisher Ultimate LC and Fusion Orbitrap MS (San Jose, CA). Briefly, peptides were first loaded onto a trap cartridge (Thermo Fisher PepMap, C18, 5 μm , 0.3 $\times$ 5 mm), then eluted onto a reversed phase Easy-Spray column (Thermo Fisher PepMap, C18, 3 μm , 100 Å) using a linear 120 min gradient of ACN (2–50%) containing 0.1% formic acid at 250 nL/min flow rate. The eluted peptides were sprayed into the Fusion Orbitrap. The data-dependent acquisition (DDA) mode was enabled, and each FTMS MS1 scan (120 000 resolution) was followed by linear ion-trap MS2 scans using top speed (acquire as many MS2 scans as possible within one second cycle time). Precursor ion fragmentation took place in the HCD cell with CE energies of 33 for glycopeptides. Automatic gain control (AGC) targets were 2.0×10^5 and 1.0×10^4 ,

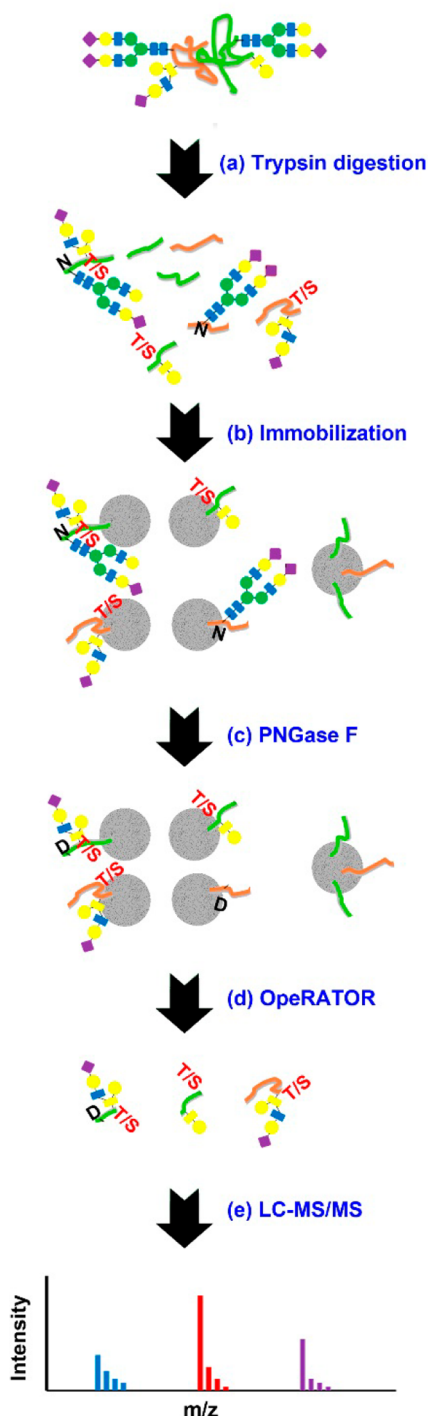


Figure 2. Different methods for analysis of O-glycosylation. (a) In solution digestion with PNGase F + OpeRATOR: N-glycans were released by PNGase F. After C18 SPE cleanup, peptides were digested in solution using OpeRATOR, followed by C18 cleanup of O-glycopeptides. (b) After SPE cleanup (HILIC and C18 respectively), O-GIG immobilizes peptides to the resin, followed by PNGase F release of N-glycans. The O-glycopeptides were digested and enriched from the resin using OpeRATOR. (c) In solution digestion with PNGase F only: N-glycans were removed by PNGase F, followed by C18 cleanup of O-glycopeptides.

respectively, for MS1 and MS2. The spray voltage and ion transfer tube temperature were set at 1.8 kV and 250 °C,

respectively. MS spectra were analyzed by Byonic and Byologic software (Protein Metrics, CA).

Protein O-Glycosylation 3D Molecular Structure. The peptide sequence of glycoprotein was acquired from Uniprot (www.uniprot.org). The known O-glycosites were based on the references in Uniprot “amino acid modifications,” together with references cited on the Web site. The sequence was uploaded to Swiss-model (<https://swissmodel.expasy.org>) for searching the glycoprotein template.³⁶ This generates a protein model in PDB format, which can be uploaded to Glycam (glycam.org) for incorporating glycosylation in the model, including N-glycans and O-glycans. The model with glycosylation added can be output in PDB format, which was used in CCP4MG for postprocessing.

RESULTS

Anticipated Enzymatic Digestion of O-Glycosites. O-glycoproteins or O-glycopeptides were usually digested in solution. The enzyme activity was demonstrated in the Supporting Information (Figures S2 and S3). Going forward, the term “in solution” is used to distinguish from “solid-phase.” To illustrate the digestion process, we use a hypothetical peptide with O-glycans attached (hypothetical O-glycopeptide), $\text{H}_2\text{N}-\text{ST}[\text{NeuAcHexHexNAc}]\text{PIGQPS}[\text{HexHexNAc}]\text{IPGGPVRLT}-\text{COOH}$. If OpeRATOR is the only enzyme used, the digestion products will include $\text{H}_2\text{N}-\text{S}$, $\text{T}[\text{NeuAcHexHexNAc}]\text{PIGQP}$, and $\text{S}[\text{HexHexNAc}]\text{IPGGPVRLT}-\text{COOH}$ (Figure 3a). This enzyme cleaves the N-terminal of S or T with O-glycan attached. Conversely, when the peptide is immobilized on the solid support (O-GIG), it generates the same peptide and glycopeptides; however, $\text{H}_2\text{N}-\text{S}$ is immobilized on the resin (Figure 3b). This reveals that O-GIG will reduce the number of non-O-glycopeptides in the digest. Further, if in solution the peptide is digested by trypsin first, there are two peptides in the solution, including $\text{H}_2\text{N}-\text{ST}[\text{NeuAcHexHexNAc}]\text{PIGQPS}[\text{HexHexNAc}]\text{IPGGPVR}$ and $\text{LT}-\text{COOH}$; after OpeRATOR digestion, four peptides are present in solution in the final product, i.e., $\text{H}_2\text{N}-\text{S}$, $\text{T}[\text{NeuAcHexHexNAc}]\text{PIGQP}$, $\text{S}[\text{HexHexNAc}]\text{IPGGPVR}$, and $\text{H}_2\text{N}-\text{LT}-\text{COOH}$ (Figure 3c). Note that half of the peptides yielded are non-O-glycosylated. However, if peptides are used for conjugation, each peptide is immobilized to the resin; further digestion with OpeRATOR generates two O-glycopeptides in the flow-through, including $\text{T}[\text{NeuAcHexHexNAc}]\text{PIGQP}$ and $\text{S}[\text{HexHexNAc}]\text{IPGGPVR}$ (Figure 3d). Therefore, solid-phase enzymatic digestion yields simplified products due to retention of the N-terminal peptide. This illustration demonstrates the anticipated O-GIG specific enrichment for O-glycopeptides.

Quantity of OpeRATOR. To determine the amount of enzyme that is required to digest O-glycosites from glycopeptides, we used increasing amounts of OpeRATOR for testing the tryptic fetuin (from bovine serum). Figure 4 shows the relative ion abundance of known O-glycosites of bovine fetuin liberated using different amounts of enzyme. Six reported O-glycosites from fetuin are $\text{S}[271]$, $\text{T}[280]$, $\text{S}[282]$, $\text{S}[296]$, $\text{T}[334]$, and $\text{S}[341]$ (Supporting Information Table S1). The other potential O-glycosites are estimated using NetOGlyc 4.0, where a score of 0.5 higher is predicted as “positive” for glycosylation.³⁷ When no or minimal enzyme is applied, fewer O-glycosites and lower abundance for each site are observed (Figure 4a,b). When 100 units of enzymes are added to 300 μL of buffer (20 mM Tris, pH 6.8), all six O-glycosites are detected with substantially increased relative

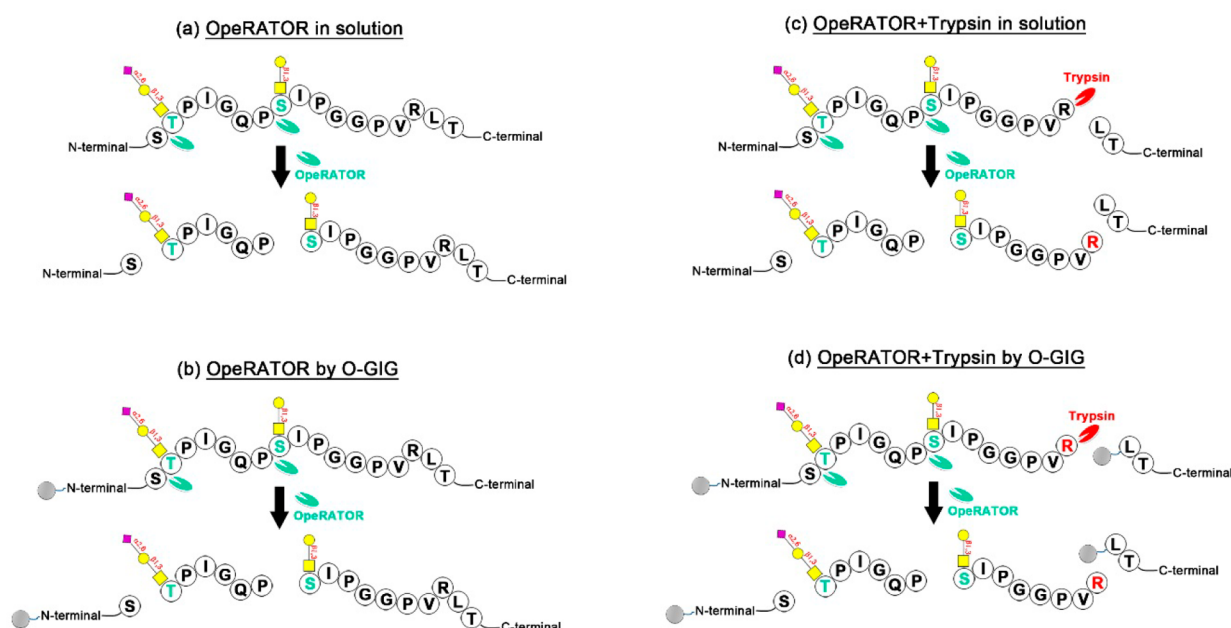


Figure 3. The schematic digestion of *O*-protease on glycopeptides in solution or on the solid support. (a) OpeRATOR digestion of *O*-glycoproteins in solution. Only *O*-glycosylated S or T can be cleaved by enzymes, resulting in N-terminal peptides, C-terminal peptides, and *O*-glycopeptides in solution. (b) OpeRATOR digestion of *O*-glycoproteins by O-GIG. *O*-glycosylated S or T peptides remain in the sample, while N-terminal peptides immobilize on resins. (c) OpeRATOR digestion in solution on tryptic peptides. Enzymatic cleavage on glycosylated S or T results in N-terminal peptides, *O*-glycosylated peptides, and C-terminal peptides. (d) OpeRATOR digestion by O-GIG on tryptic peptides. N-terminal and *O*-terminal peptides were removed by remaining on resins and washing solution, respectively.

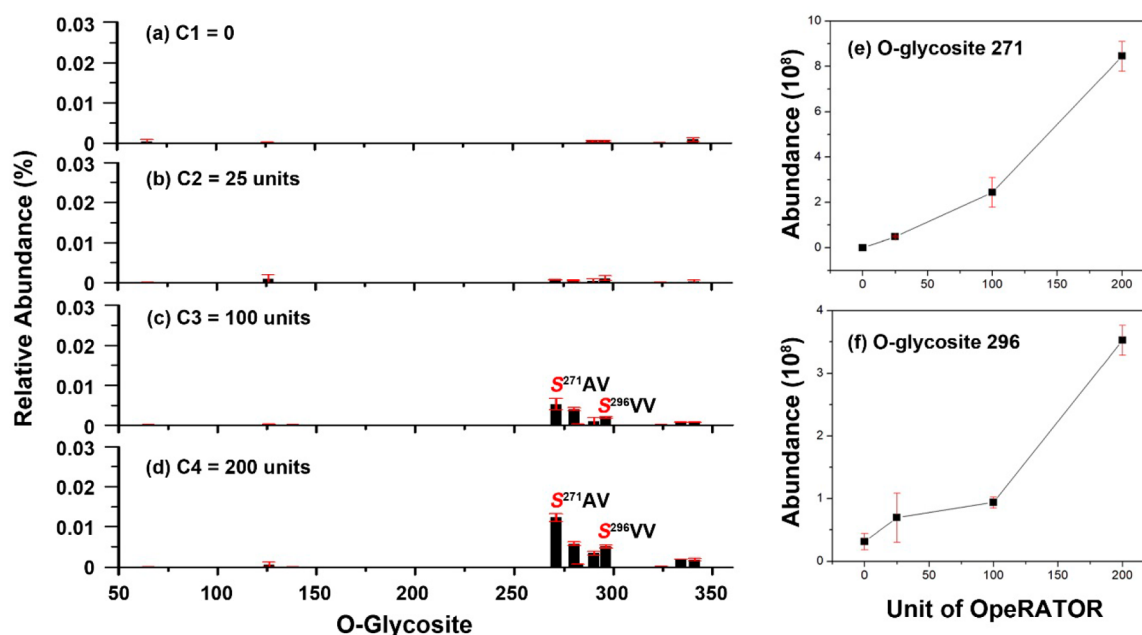


Figure 4. The effect of OpeRATOR concentration on the digestion of *O*-glycosites. (a) Concentration (C_1) = 0, relative abundance of *O*-glycosites in the identified peptides on different sites; (b) C_2 = 25 units; (c) C_3 = 100 units; (d) C_4 = 200 units; (e) *O*-glycosites at 271; (f) *O*-glycosites at 296. The error bar represents the standard deviation.

abundance (Figure 4c,d; Supporting Information Table S2). With the increased amount of enzyme, more *O*-glycopeptides are observed. For example, the relative abundance of *O*-glycosites, S[271] or S[296], is substantially increased in the identified peptides (Figure 4e,f), revealing that the enhanced digestion on *O*-glycosites has been successfully achieved. Additionally, other potential *O*-glycosites are also identified by O-GIG analysis (see Supporting Information Table S2).

O-GIG Specificity and Enrichment on O-Glycosylation Analysis. Performance of O-GIG analysis of *O*-glycosylation was assessed. All samples are processed with PNGase F (PF). Solid and solution phases, with HILIC or C18 enrichment, and with or without OpeRATOR (Op) were tested. As shown in Figure 5, conditions were as follows. I = HILIC/Solution/PF/Op: peptides are enriched by HILIC first, followed by PNGase F (PF)/OpeRATOR digestion in solution. II = C18/Solution/

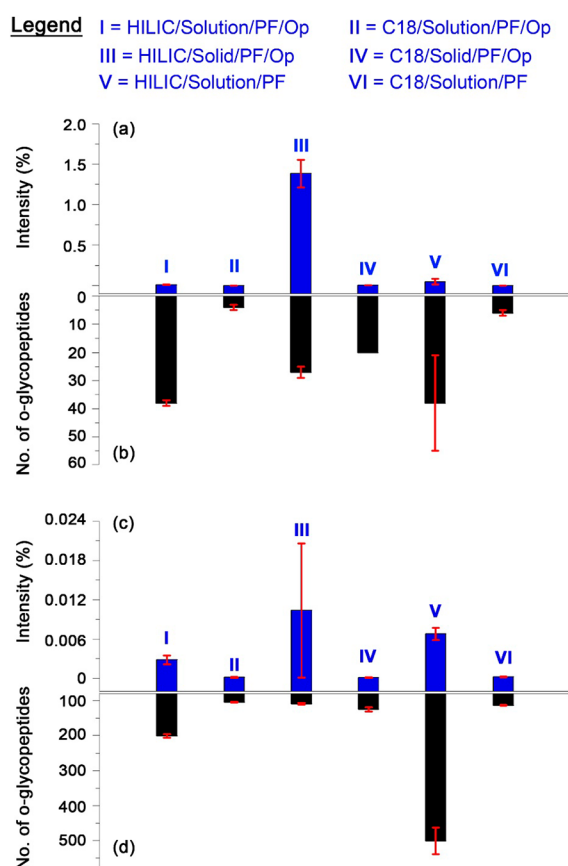


Figure 5. Comparison of O-GIG performance of O-glycopeptide identification with in-solution methods with and without OpeRATOR. (a) Relative abundance of O-glycopeptides with single O-glycosite. I = HILIC/Solution/PF/Op: PNGase F/OpeRATOR digests HILIC enriched O-glycopeptides in solution. II = C18/Solution/PF/Op: PNGase F/OpeRATOR digests C18 purified O-glycopeptides in solution. III = HILIC/Solid/PF/Op: PNGase F/OpeRATOR digests HILIC enriched O-glycopeptides on solid-phase (O-GIG). IV = C18/Solid/PF/Op: PNGase F/OpeRATOR digests C18 purified O-glycopeptides on solid-phase (O-GIG). V = HILIC/Solution/PF: O-glycopeptides by HILIC after PNGase F and trypsin digestion. VI = C18/Solution/PF: O-glycopeptides after PNGase F and trypsin digestion. (b) Number of O-glycopeptides with single O-glycosite identified by different methods. (c) Relative abundance of O-glycopeptides with multiple O-glycosites. (d) Number of O-glycopeptides with multiple O-glycosites. The error bar represents the standard deviation.

PF/Op: peptides are purified by C18, followed by PNGase F/OpeRATOR digestion in solution. III = HILIC/Solid/PF/Op: peptides are enriched by HILIC, followed by PNGase F/OpeRATOR digestion on the solid support. IV = C18/Solid/PF/Op: peptides are purified by C18, followed by PNGase F/OpeRATOR digestion on the solid support. V = HILIC/Solution/PF: peptides are enriched by HILIC only in solution. VI = C18/Solution/PF: peptides are purified by C18 only in solution. Figure 5a and b show the relative abundance and number of glycopeptides containing one glycosite. We plotted the relative abundance of O-glycosites from these methods, as well as the number of O-glycosites identified. Overall, HILIC enriched glycopeptides produced better identification of single O-glycosite O-glycopeptides both in relative abundance (Figure 5a III) and number (Figure 5b III) of O-glycosites. Although a higher number of O-glycosites are identified in solution (I)

compared to O-GIG (III), more missed-cleavage O-glycosite products are detected in solution. The relative abundance of O-glycopeptides is considerably lower for the in-solution approach (I or II). Without the use of OpeRATOR (V, HILIC; VI, C18), there are an appreciable number of O-glycosites detected; however, the relative abundance of O-glycopeptides are very low in both approaches (V and VI). The OpeRATOR can specifically cleave the N-terminal of O-glycosylated S or T, and this feature of the enzyme resulted in fewer O-glycopeptides with multiple O-glycosites. Figure 5c and d show the abundance and number, respectively, of glycopeptides with multiple glycosites. This is evidenced by the low number and intensity of such peptides where OpeRATOR is used (I, II, III, and IV). PNGase F only digested peptides show approximately a 5-fold increase of the number of multiple O-glycosites in the same O-glycopeptide (V). The minimal number of O-glycopeptides with multiple O-glycosites is detected by the O-GIG method (III). The MS/MS fragmentation on O-glycopeptide, S-[NeuAcHexHexNAc]IPGGPVR, is compared on these methods (Supporting Information Figure S4). These results confirm that O-GIG is a robust method specifically for digestion of O-glycosites and enrich the O-glycopeptides.

Mucin-Type O-Glycosylation. The O-GIG technique was used to study bovine submaxillary gland mucin (MBS) O-glycosites. The MBS, encoded by the Muc1 gene, contains 580 amino acids (Uniprot ID: Q8WML4). Using NetOGlyc 4.0 for O-glycosite prediction, a maximum of 181 potential O-glycosites are present with 149 sites scoring 0.500 or above, indicating that these sites have a higher probability for O-glycosylation. However, only seven O-glycosites have been confirmed based on Uniprot. These known O-glycosites are centered around S[77]PGH from S[73] to S[89]. Most O-glycosites remain to be investigated. The predicted O-glycosites are listed in Supporting Information Table S3. O-GIG detects 56 O-glycosites from 95 unique O-glycopeptides. Forty O-glycosites identified are predicted with high probability by NetOGlyc 4.0 (>0.50). Sixteen O-glycosites with a low NetOGlyc probability score (<0.50) were clustered in the MBS N-terminal domain. We compared the identification of O-glycosites with and without sialic acids by treating with neuraminidase (Supporting Information Table S3). Results reveal that sialic acid removal improved the identification of O-glycosites, probably due to the negative charge from sialic acids that repel the enzyme binding to O-glycan, negatively impacting O-glycosite digestion. After removal of sialic acids, a total of 56 O-glycosites were detected; conversely, only 37 O-glycosites were found in MBS with intact sialic acids.

Figure 6 depicts the 3D structure of MBS without O-glycans (a) and with O-glycans decorated (b). Muc1 consists of transmembrane, SEA (sperm protein, enterokinase and agrin), and N-terminal (Muc1-N) domains.^{38,39} Muc1-N contains a heavily O-glycosylated tandem repeat, whose O-glycosylation has not been fully characterized (Figure 6). We identified many O-glycosites in this region using O-GIG. To compare how O-glycans decorated on the MBS Muc1-N domain, the 3D structure of MBS was generated by Swiss-Model, GLYCAM, and CCP4MG. Without O-glycans modeled to the protein backbone, the Muc1-N domain is oriented as a random coil (Figure 6a). Using detected O-glycosylation data, O-glycans were added to their respective glycosites (note: we only added those highly O-glycosylated sites for illustration; Figure 6b). The N-terminal region from Serine 157 through Threonine 576

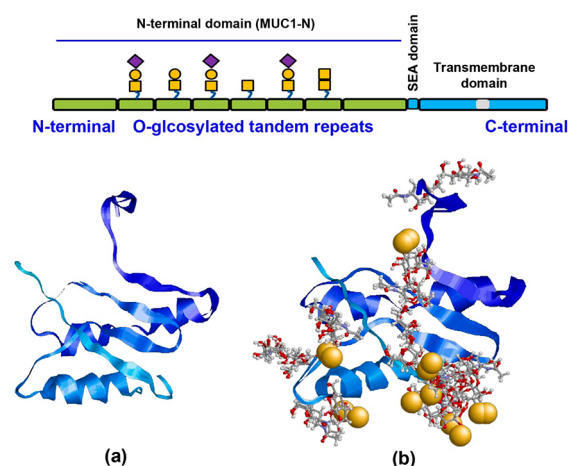


Figure 6. Depiction of highly *O*-glycosylated N-terminal domain using O-GIG. The short C-terminal is embedded in the membrane as a transmembrane domain, whereas the SEA domain connects the C-terminal to its N-terminal domain. (a) 3D structure of MBS without *O*-glycan attached; (b) 3D structure of MBS with *O*-glycans attached.

was highly decorated by *O*-glycans, in agreement with previous reports.³⁹

Zika Virus *O*-Glycosylation and 3D Structure of Zika Envelope and NS1 Glycoproteins. Next, we studied *O*-glycosylation of the Zika virus (ZIKV) strain glycoprotein. Glycoproteins, ZIKV-E/envelope protein (domain III; 40543-V08H; Sino Biological) and ZIKV-NS1 protein (40544-V07H; Sino Biological, China), are used as a model for mapping *O*-glycosites. Both ZIKV-E and ZIKV-NS1 are recombinant glycoproteins, with a DNA sequence encoding the Zika virus (strain Zika SPH2015) E/Envelope (domain III; from Val[593] to Lys [699]) and the Zika virus (strain Zika SPH2015) NS1 (from Val[796] to Leu[1157]) of the polyprotein (Zika virus; Supporting Information Table S6). Several observations were made for ZIKV-E *O*-glycosylation as revealed by O-GIG analysis: (1) Glycosite T[643] is likely heavily *O*-glycosylated and accounts for 52.3% of *O*-glycosite relative ion abundance (Supporting Information Table S5); glycosite T[615] occupies 17.4% in all peptides identified by O-GIG-LC-MS. (2) The relative abundance of T[643][HexNAc(3)Fuc(1)] is most dominant in glycosite relative abundance with 38% intensity. (3) *O*-glycans are asialyl type with only 0.71% present as sialyl forms. ZIKV-NS1 glycoprotein is more complex than ZIKV-E with more *O*-glycosites and *O*-glycans (Supporting Information Tables S5 and S6). Eighty-one unique *O*-glycopeptides are identified in ZIKV-NS1, consisting of 32 *O*-glycosites and 41 *O*-glycans (10 sialyl *O*-glycans and 31 asialyl *O*-glycans). The relative abundance of *O*-glycopeptides in Zika NS1 is 53% of total detected intensity, in which S[1137][HexNAc(3)]NLVR is up to 43.8%. The asialyl *O*-glycopeptides are approximately 52% of total intensity among all peptides identified. These results demonstrate that ZIKV-NS1 is considerably glycosylated by asialyl *O*-glycans, with a minimal number of *O*-sialylated glycosites. Therefore, both ZIKV-NS1 and ZIKV-E are minimally sialylated.

On the basis of the *O*-glycosylation on both ZIKV-E and ZIKV-NS1, we built models to schematically illustrate the distribution of *O*-glycosylation on E and NS1 (Figure 7). To show the sites of *O*-glycans on ZIKV glycoprotein, we plot the 3D structure E or NS1 without *O*-glycans (Figure 7a and c) and the most dominant *O*-glycans (Figure 7b and d). In general, *O*-

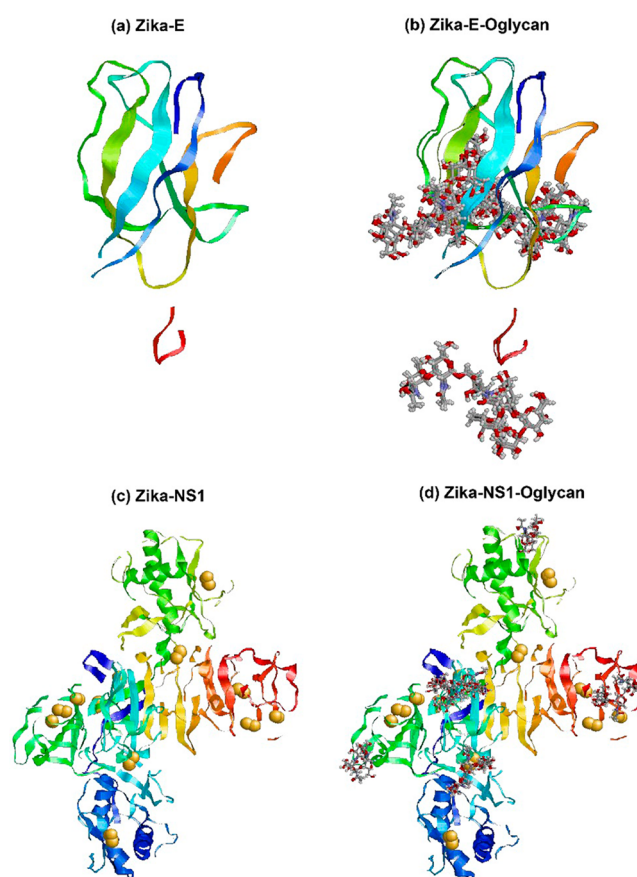


Figure 7. *O*-glycosylation on Zika virus strain SPH2015 glycoprotein. The *O*-glycosylation of ZIKV (Zika virus) was characterized by the O-GIG chemoenzymatic method. (a) ZIKV-E: strain Zika SPH2015 envelope protein without *O*-glycan attached to the protein. (b) ZIKV-E with high abundance of *O*-glycans on surface. *O*-glycosite depicted: T[643][HexNAc(3)Fuc(1)] (38.4%), T[615][HexNAc(2)Hex(2)] (17.4%), T[641][HexNAc(1)Hex(1)] (13.7%), T[617][HexNAc(2)-Hex(2)] (4.9%), and T[696][HexNAc(3)Hex(2)Fuc(2)] (1.4%). (c) ZIKV-NS1: strain Zika SPH2015 protein without *O*-glycans attached. (d) ZIKV-NS1 with high abundance of *O*-glycans attached. *O*-glycosites include S[1137][HexNAc(3)] (43.8%), S[1040]-[HexNAc(2)Hex(1)] (1.4%), T[923][HexNAc(1)Hex(1)] (0.77%), T[1003][HexNAc(3)Hex(1)Fuc(1)] (0.54%), and T[1087]-[HexNAc(1)Hex(1)NeuAc(1)] (0.26%).

glycans are decorated on the outer surface of ZIKV glycoproteins for both E and NS1. The *O*-glycosylation of ZIKV-E and ZIKV-NS1 is listed in the Supporting Information Tables S5 and S6.

DISCUSSION

Although mass spectrometry has greatly facilitated the advancement of glycoprotein analysis, the lack of robust strategies for the analysis of *O*-glycosylation makes such analyses challenging. This is largely due to the absence of *O*-glycosylation specific glycosidases and proteinases/peptidases. The heterogeneity of *O*-glycan compounds discovery needed enzymes since many core configurations are possible for this post translational modification. For example, mucin-type *O*-glycans alone contains eight different core structures, which can subsequently be extended to give mature linear or branched *O*-glycans. Until now, no single enzyme was available to digest all eight core types at the glycan or protein/peptide level.

Moreover, mucin-type domains could have tens of closely spaced S and T glycosylation sites, which may be heavily decorated with an assortment of complex O-GalNAc glycans. This situation can lead to steric issues regarding enzyme access. Although sequential digestion using a range of glycosidases can increase access to glycosites, a multienzyme degradation approach can be complex. For instance, mucin-type O-glycans could contain Gal, GlcNAc, Fuc, and Sia in a range of linkage configurations, and therefore, the choice of enzyme sets may not be straightforward in many cases. Gal and GlcNAc also may be modified by sulfation, whereas Sia could be modified by O-acetylation.⁴⁰ The use of proteases with unique O-protease properties can be an effective approach to decipher the complexity of protein O-glycosylation since the O-glycan can remain intact with all its complexity, and peptide information can allow mapping of the resident glycan(s) within the protein.

In this work, we have utilized a recently discovered O-protease, OpeRATOR, an O-glycan specific endoprotease, discovered in *Akkermansia muciniphila* and expressed in *E. coli*. We have developed a novel approach to study protein O-glycosylation using this enzyme in a chemoenzymatic strategy. The enzyme binds specifically to O-glycans on native glycoproteins or glycopeptides and digests the peptide backbone at the N-terminal end of O-glycan substituted S and T residues. It is critical that this enzyme has no activity on nonglycosylated S or T, as well as N-glycosylated N (asparagine). Therefore, after release of N-glycans, we can specifically accumulate glycopeptides with O-glycosylated S and T. This technique allows us to study complex O-glycosylation, even in densely O-glycosylated regions such as that found in mucin-type domains.

Several strategies can be executed for digestion of O-glycopeptides by OpeRATOR. (1) Glycoproteins vs glycopeptides: the digestion of O-glycosylated sites can be carried out using glycoproteins or glycopeptides. (a) OpeRATOR using glycoproteins: OpeRATOR digestion of O-glycosylated S and T on glycoproteins yields O-glycopeptides and non-O-glycopeptides. The resulting glycopeptides may contain a long length of amino acids since the two adjacent O-glycosites can be far away from each other (Figure 3a). Where multiple K/R may exist between them, a trypsin strategy may be introduced (Figure 3b). The OpeRATOR digest can be further cleaved by trypsin, producing shorter, charged glycopeptides containing C-terminal K or R for better ionization. (b) OpeRATOR using glycopeptides: Glycopeptides also can be generated using trypsin first followed by OpeRATOR digestion. This will generate nonglycosylated peptides, N-glycopeptides, and peptides with O-glycans and/or N-glycans. The identification is improved through glycopeptide enrichment using HILIC SPE. (2) In solution vs solid-phase: the enzymatic digestion can be performed in solution or on a solid phase. (a) In solution, when glycoprotein is digested by trypsin followed by OpeRATOR cleavage. Products will include nonglycopeptides, N-glycopeptides, and O-glycopeptides. To remove N-glycopeptides, PNGase F or A is required to cleave N-glycans. To collect the O-glycopeptides, multiple chromatography cleanup and enrichment steps are required, including C18 and HILIC. (b) On the solid phase, glycopeptides can be immobilized for conducting sequential chemical or enzymatic reactions (Figure 3d). The removal of chemicals or enzymes is facilitated by extensive washing steps.^{11,12} Products eluted are simplified since some nonglycopeptides remain linked to the solid phase. This approach can be useful for analyzing O-

glycosylation since multiple chemical and enzymatic treatments are required and washing steps are simplified on the solid phase format.

The solid-phase chemoenzymatic method (O-GIG) is specific for enrichment of O-glycopeptides. Use of glycopeptides illustrates this unique feature in comparison to the use of glycoproteins. All peptides are immobilized on resin via reductive amination through their N-termini. Once conjugated to resin, the glycopeptides can be treated by sequential reactions including the derivatization of Sia linkage,¹² stabilization of Sia,^{41,42} release of N-glycans,⁴¹ or cleavage of O-glycans.^{11,43} The sequential derivatization of Sia linkage has been demonstrated in our previous studies, which not only differentiate $\alpha 2,3$ and $\alpha 2,6$ but also neutralize negatively charged sialic acid.¹² It is essential to digest O-glycosylated S and/or T on the solid support, because only those O-glycopeptides can be snipped from their O-glycosites, while nonglycosylated S and/or T remains on the resin (Figure 3c,d). Thus, the OpeRATOR digest produces highly abundant O-glycopeptides with O-glycans on the N-terminal of S or T (Supporting Information Table S2). Using this strategy, few multiple-glycosite glycopeptides are yielded with more than one O-glycan resident. Comparatively, assignment of O-glycopeptides produced by in-solution OpeRATOR or trypsin digestion alone may be ambiguous. Some peptides may contain more than one O-glycosite, and therefore, each glycan location and precise composition may not be resolved by MS spectral fragment ion patterns (Supporting Information Table S7).

Application of O-GIG on mucin-type glycoprotein and complex biological samples show that the method is powerful for annotating O-glycosylation. We identified many previously unidentified O-glycosites, on the order of 8-fold more sites than previously reported. NetOGlyc 4.0 was used to estimate O-glycosite glycosylation potential. Overall, O-GIG confirms those predicted sites with 0.5 probability or greater and discovers others having lower probability scores. Modeling on ZIKV glycoprotein demonstrates the feasibility to use O-GIG to fully characterize O-glycosylation on the ZIKA virus and reveal whether those glycosites are nearby Zika antigenic sites; it also provides a novel means for tracing O-glycosylation changes as it and other viruses propagate through the human population. The latter work is ongoing. It should be noted that O-GIG presented here may also be useful for studying protein N-glycosylation. In the O-GIG protocol (Figure 1c), N-glycans are released for quantitative analysis, while asparagine (N) simultaneously changes to aspartic acid (D). In the data search, we can use deamidation for targeting those N-glycosites for identification. The N-glycosites have been identified here using O-GIG (Supporting Information Table S8; Fetuin, MBS, Zika E and NS1), showing that the O-GIG chemoenzymatic method can be used for analysis of both N-linked and O-linked glycosylation. It is remarkable that O-GIG can identify new O-glycosites and N-glycosites that have not been reported. This will expand the horizons for studying O-glycosylation for different diseases, facilitating discovery of biomarkers and vaccines for diagnostics and immunization.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b01834.

Mucin-type O-glycoprotein with highly O-glycan decorated domain; H2N-AAPRLI...SPPDAA¹²⁵S¹²⁶[HexHexNAc]AAPLR...ACRTGDCOOH by OpeRATOR; OpeRATOR digestion on glycosylated and nonglycosylated proteins; tandem mass spectrometry of O-glycopeptides digested from bovine fetuin (DOCX)

Fetuin O-glycosites (XLSX)

O-GIG fetuin O-glycosite list (XLSX)

MBS O-glycosites (XLSX)

ZIKV glycoprotein sequence (XLSX)

ZIKV-E O-glycosylation (XLSX)

ZIKV-NS1 O-glycosylation (XLSX)

Fetuin O-glycopeptides by trypsin-HILIC (XLSX)

N-glycosites by O-GIG (XLSX)

AUTHOR INFORMATION

Corresponding Authors

*Address: G614, Bldg. 75, 10903 New Hampshire Ave, Silver Spring, MD 20993. Tel. (240) 402-8401. E-mail: shuang.yang@fda.hhs.gov or jake.yang@gmail.com.

*Address: G637, Bldg. 52/72, 10903 New Hampshire Ave, Silver Spring, MD 20993. Tel. (240) 402-9445. E-mail: john.cipollo@fda.hhs.gov.

ORCID

Shuang Yang: 0000-0001-7958-0594

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This technology was developed in Laboratory of Bacterial Polysaccharides at the Food and Drug Administration.

REFERENCES

- (1) Fukuda, M. *Biochim. Biophys. Acta, Gen. Subj.* **2002**, 1573, 394–405.
- (2) Hammond, C.; Braakman, I.; Helenius, A. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, 91, 913–7.
- (3) Go, E. P.; Chang, Q.; Liao, H. X.; Sutherland, L. L.; Alam, S. M.; Haynes, B. F.; Desaire, H. *J. Proteome Res.* **2009**, 8, 4231–42.
- (4) Moremen, K. W.; Tiemeyer, M.; Nairn, A. V. *Nat. Rev. Mol. Cell Biol.* **2012**, 13, 448–462.
- (5) Varki, A. *Glycobiology* **2017**, 27, 3–49.
- (6) Helenius, A.; Aebi, M. *Annu. Rev. Biochem.* **2004**, 73, 1019–1049.
- (7) Brockhausen, I. *EMBO Rep.* **2006**, 7, 599–604.
- (8) Arnold, J. N.; Wormald, M. R.; Sim, R. B.; Rudd, P. M.; Dwek, R. A. *Annu. Rev. Immunol.* **2007**, 25, 21–50.
- (9) Takahashi, N. *Biochem. Biophys. Res. Commun.* **1977**, 76, 1194–1201.
- (10) Zhou, H.; Ranish, J. A.; Watts, J. D.; Aebersold, R. *Nat. Biotechnol.* **2002**, 20, 512–515.
- (11) Yang, S.; Hu, Y.; Sokoll, L.; Zhang, H. *Nat. Protoc.* **2017**, 12, 1229–1244.
- (12) Yang, S.; Jankowska, E.; Kosikova, M.; Xie, H.; Cipollo, J. *Anal. Chem.* **2017**, 89, 9508–9517.
- (13) Fujita, K.; Oura, F.; Nagamine, N.; Katayama, T.; Hiratake, J.; Sakata, K.; Kumagai, H.; Yamamoto, K. *J. Biol. Chem.* **2005**, 280, 37415–37422.
- (14) Koutsoulis, D.; Landry, D.; Guthrie, E. P. *Glycobiology* **2008**, 18, 799–805.
- (15) Iwase, H.; Hotta, K. Release of O-linked glycoprotein glycans by endo- α -N-acetylgalactosaminidase. In *Glycoprotein Analysis in Biomedicine*, 1st ed.; Hounsell, E., Ed.; Humana Press: Clifton, NJ, 1993; Vol. 1, pp 151–159.
- (16) Patel, T.; Bruce, J.; Merry, A.; Bigge, C.; Wormald, M.; Parekh, R.; Jaques, A. *Biochemistry* **1993**, 32, 679–693.
- (17) Song, X.; Ju, H.; Lasanajak, Y.; Kudelka, M. R.; Smith, D. F.; Cummings, R. D. *Nat. Methods* **2016**, 13, 528–534.
- (18) O'Shannessy, D.; Hoffman, W. *Biotechnol. Appl. Biochem.* **1987**, 9, 488–496.
- (19) Kaji, H.; Saito, H.; Yamauchi, Y.; Shinkawa, T.; Taoka, M.; Hirabayashi, J.; Kasai, K.-i.; Takahashi, N.; Isobe, T. *Nat. Biotechnol.* **2003**, 21, 667–672.
- (20) Zhu, Z.; Desaire, H. *Annu. Rev. Anal. Chem.* **2015**, 8, 463–83.
- (21) An, Y.; McCullers, J. A.; Alymova, I.; Parsons, L. M.; Cipollo, J. F. *J. Proteome Res.* **2015**, 14, 3957–3969.
- (22) An, Y.; Cipollo, J. F. *Anal. Biochem.* **2011**, 415, 67–80.
- (23) Woo, C. M.; Iavarone, A. T.; Spiclarich, D. R.; Palaniappan, K. K.; Bertozzi, C. R. *Nat. Methods* **2015**, 12, 561–567.
- (24) Christiansen, M. N.; Kolarich, D.; Nevalainen, H.; Packer, N. H.; Jensen, P. H. *Anal. Chem.* **2010**, 82, 3500–3509.
- (25) Vakhrushev, S. Y.; Steentoft, C.; Vester-Christensen, M. B.; Bennett, E. P.; Clausen, H.; Levery, S. B. *Mol. Cell. Proteomics* **2013**, 12, 932–944.
- (26) Mancera-Arteu, M.; Giménez, E.; Benavente, F.; Barbosa, J.; Sanz-Nebot, V. *J. Proteome Res.* **2017**, 16, 4166–4176.
- (27) Hansen, J. E.; Lund, O.; Tolstrup, N.; Gooley, A. A.; Williams, K. L.; Brunak, S. *Glycoconjugate J.* **1998**, 15, 115–130.
- (28) Levery, S. B.; Steentoft, C.; Halim, A.; Narimatsu, Y.; Clausen, H.; Vakhrushev, S. Y. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, 1850, 33–42.
- (29) Rademaker, G. J.; Haverkamp, J.; Thomas-Oates, J. *Org. Mass Spectrom.* **1993**, 28, 1536–1541.
- (30) Rademaker, G. J.; Pergantis, S. A.; Blok-Tip, L.; Langridge, J. I.; Kleen, A.; Thomas-Oates, J. E. *Anal. Biochem.* **1998**, 257, 149–160.
- (31) Hanisch, F.-G.; Jovanovic, M.; Peter-Katalinic, J. *Anal. Biochem.* **2001**, 290, 47–59.
- (32) Zheng, Y.; Guo, Z.; Cai, Z. *Talanta* **2009**, 78, 358–363.
- (33) Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Mol. Cell. Proteomics* **2002**, 1, 791–804.
- (34) Zauner, G.; Kozak, R. P.; Gardner, R. A.; Fernandes, D. L.; Deelder, A. M.; Wuhrer, M. *Biol. Chem.* **2012**, 393, 687–708.
- (35) Yang, S.; Mishra, S.; Chen, L.; Zhou, J.-y.; Chan, D. W.; Chatterjee, S.; Zhang, H. *Anal. Chem.* **2015**, 87, 9671–9678.
- (36) Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. *Nat. Protoc.* **2009**, 4, 1–13.
- (37) Steentoft, C.; Vakhrushev, S. Y.; Joshi, H. J.; Kong, Y.; Vester-Christensen, M. B.; Katrine, T.; Schjoldager, B.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N. B.; et al. *EMBO J.* **2013**, 32, 1478–1488.
- (38) Dekker, J.; Rossen, J. W.; Büller, H. A.; Einerhand, A. W. *Trends Biochem. Sci.* **2002**, 27, 126–131.
- (39) Nath, S.; Mukherjee, P. *Trends Mol. Med.* **2014**, 20, 332–342.
- (40) Brockhausen, I.; Stanley, P. O-GalNAc glycans. In *Essentials of Glycobiology*, 3rd ed.; Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: New York, 2017.
- (41) Yang, S.; Li, Y.; Shah, P.; Zhang, H. *Anal. Chem.* **2013**, 85, 5555–5561.
- (42) Yang, S.; Zhang, L.; Thomas, S.; Hu, Y.; Li, S.; Cipollo, J.; Zhang, H. *Anal. Chem.* **2017**, 89, 6330–6335.
- (43) Yang, S.; Höti, N.; Yang, W.; Liu, Y.; Chen, L.; Li, S.; Zhang, H. *Clin. Proteomics* **2017**, 14, 1–11.

Deciphering protein *O*-glycosylation: solid-phase chemoenzymatic cleavage and enrichment

Running title: Solid-phase chemoenzymatic digestion on *O*-glycosylation

Shuang Yang^{1*}, Philip Onigman², Wells W. Wu³, Jonathan Sjogren², Helen Nyhlen⁴, Rong-Fong Shen³, John Cipollo^{1*}

¹ Laboratory of Bacterial Polysaccharides, Division of Bacterial, Parasitic and Allergenic Products, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD 20993, USA

² Genovis Inc., Cambridge, MA 02142, USA

³ Facility for Biotechnology Resources, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD 20993, USA

⁴ Genovis AB, Box 790, SE-220 07 Lund, Sweden

Abstract

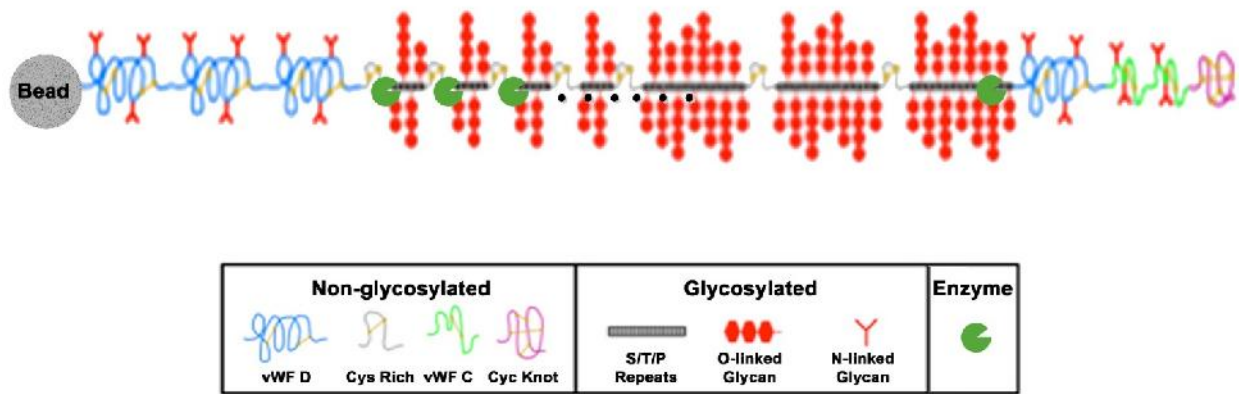
Glycosylation plays a critical role in the biosynthetic-secretory pathway in the endoplasmic reticulum (ER) and Golgi apparatus. Over 50% of mammalian cellular proteins are typically glycosylated and this modification is involved in a wide range of biological functions such as barrier formation against intestinal microbes and serving as signaling molecules for selectins and galectins in the innate immune system. *N*-linked glycosylation analysis has been greatly facilitated owing to a range of specific enzymes available for their release. However, system-wide analysis on *O*-linked glycosylation remains a challenge due to the lack of equivalent enzymes and the inherent structural heterogeneity of *O*-glycans. Although *O*-glycosidase can catalyze the removal of Core 1 and Core 3 *O*-linked disaccharides from glycoproteins, analysis of other types of *O*-glycans remains difficult, particularly when residing on glycopeptides. Here we describe a novel chemoenzymatic approach driven by a newly available *O*-protease and solid phase platform. This method renders the assignment of *O*-glycosylated peptides, *N*-glycan

profile, sialyl *O*-glycopeptides linkage, and mapping of heterogenous *O*-glycosylation. For the first time, we can analyze intact *O*-glycopeptides generated by *O*-protease and enriched using a solid-phase platform. We establish the method on standard glycoproteins, confirming known *O*-glycosites with high accuracy and confidence, and reveal up to eight-fold more glycosites than previously reported with concomitant increased heterogeneity. This technique is further applied for analysis of Zika virus recombinant glycoproteins, revealing their dominant *O*-glycosites and setting a basis set of *O*-glycosylation tracts in these important viral antigens. Our approach can serve as a benchmark for the investigation of protein *O*-glycosylation in diseases and other biomedical contexts. This method should become an indispensable tool for investigations where *O*-glycosylation is central.

Table of Contents

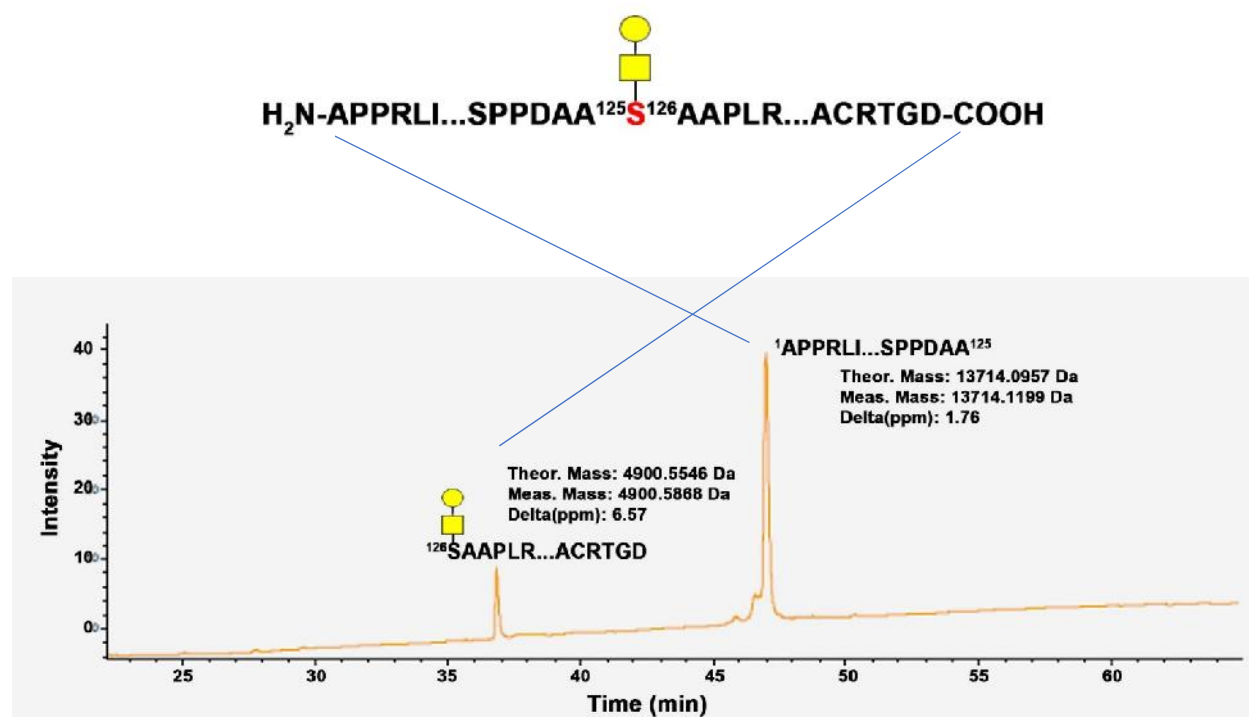
Supplementary Figure S1-----	S-3
Supplementary Figure S2-----	S-4
Supplementary Figure S3-----	S-5
Supplementary Figure S4-----	S-6

Supplementary Figure S1. Mucin-type O-glycoprotein with highly O-glycan decorated domain.



Supplementary Figure S2. Digest of O-glycopeptide H₂N-

APPRLI...SPPDAA¹²⁵S¹²⁶[HexHexNAc]AAPLR...ACRTGD-COOH by OperATOR. After operator digestion, the peptide was cleaved on the peptide bond N-terminal between AA125 and AA126.



Supplementary Figure S3. OperATOR digestion on glycosylated and non-glycosylated proteins. Fetuin and IgA are both N- and O-linked glycoproteins, while IgG is N-linked glycoprotein and Albumin has no glycosites. OperATOR only digests O-glycosylated glycoproteins of Fetuin and IgA.

