

## Methods for Enrichment and Assignment of N-Acetylglucosamine Modification Sites

Jason C. Maynard and Robert J. Chalkley\*

Department of Pharmaceutical Chemistry, University of California San Francisco

\*Corresponding Author:

600 16th Street, Genentech Hall Room N474A, San Francisco, CA 94143.

E-mail: [chalkley@cgl.ucsf.edu](mailto:chalkley@cgl.ucsf.edu)

ORCID ID: 0000-0002-9757-7302

Running title: O-GlcNAc Analysis

### Abbreviations

BEMAD	Beta-elimination followed by Michael Addition of DTT
CID	Collision Induced Dissociation
DTT	Dithiothreitol
ETD	Electron Transfer Dissociation
EThcD	Electron Transfer Dissociation combined with Collision Induced Dissociation
GalNAc	N-Acetylgalactosamine
HexNAc	N-Acetylhexosamine
O-GlcNAc	O-linked N-Acetylglucosamine

### **ABSTRACT**

O-GlcNAcylation, the addition of a single N-acetylglucosamine residue to serine and threonine residues of cytoplasmic, nuclear, or mitochondrial proteins, is a widespread regulatory post-translational modification. It is involved in response to nutritional status and stress and its dysregulation is associated with diseases ranging from Alzheimer's to diabetes. While the modification was first detected over thirty-five years ago, research into the function of O-GlcNAcylation has accelerated dramatically in the last ten years due to the development of new enrichment and mass spectrometry techniques that facilitate its analysis. This article summarizes methods for O-GlcNAc enrichment, key mass spectrometry instrumentation advancements, particularly those that allow modification site localization, and software tools that allow analysis of data from O-GlcNAc modified peptides.

## **O-GLCNACYLATION IS A WIDESPREAD REGULATORY MODIFICATION**

The cell uses several post-translational modifications to transiently regulate a protein's activity. One of the most important and widespread modifications is the addition of a single beta N-acetylglucosamine (GlcNAc) sugar residue to serine and threonine residues of nuclear, cytoplasmic, and mitochondrial proteins. This is a regulatory modification whose addition and removal are each catalyzed by a single enzyme, uridine diphospho-*N*-acetylglucosamine:polypeptide  $\beta$ -*N*-acetylglucosaminyltransferase (OGT) and O-GlcNAcase (OGA), respectively. The donor molecule is UDP-GlcNAc, which is the product of the hexosamine synthetic pathway and is thought to be a sensor for glucose levels (1). As a result, O-GlcNAcylation is broadly a response to nutritional status and stress (2). It regulates gene transcription, protein translation, and has a complex interplay with phosphorylation (3), in some cases competing for the same site of modification. O-GlcNAcylation is found in practically all multicellular organisms. In humans it is dysregulated in neuropathological disorders such as Alzheimer's disease, metabolic disorders such as diabetes, and is elevated in most cancers. For a more comprehensive description of O-GlcNAc's many roles see this recent review article (4).

## **EARLY METHODS FOR O-GLCNAC DETECTION**

The O-GlcNAc modification was serendipitously discovered over thirty-five years ago through glycan radiolabeling experiments of lymphocytes (5). The authors expected to label terminal N-acetylglucosamine residues of cell surface glycans using a galactosyltransferase but also found considerable labeling in the nucleus and cytoplasm. This radiolabeling approach (mostly using tritiated galactose) was the dominant method of detection for the first twenty years of studying the O-GlcNAc modification. Early attempts were made to use mass spectrometry to detect modified peptides; however, researchers were unable to determine modification sites in these peptides using the then available tandem mass spectrometry instrumentation. As alternatives, radiolabeling and Edman sequencing were

used to determine modification sites (6, 7). Due to instrument developments discussed later in this review, mass spectrometry has now become the dominant method for O-GlcNAc site determination.

This article will only focus on methods that allow modification site determination, but in many cases there are related approaches that can be used to visualize modified proteins through blotting or fluorescent labeling(8–11).

### **MODIFICATION ENRICHMENT IS REQUIRED**

Most tandem mass spectrometry data is acquired using a process known as data-dependent acquisition, where the instrument automatically selects the most intense ions for fragmentation analysis. Modified peptides are generally not among the most abundant species in a sample because modifications are typically sub-stoichiometric, so unmodified peptides will dominate. In a large-scale re-analysis of published proteomic datasets to look for GlcNAc modified peptides only 126 O-GlcNAc modified peptides were found among just under fourteen million spectra analyzed (12). Hence, for comprehensive analysis of O-GlcNAcylation an enrichment step is required to move modified peptides to among the most abundant in the sample such that during data dependent acquisition they are automatically selected for fragmentation analysis.

### **LECTIN ENRICHMENT**

Nature produces many sugar-binding proteins that are referred to collectively as lectins. One well studied lectin is wheat germ agglutinin (WGA), which is used by the plant in its defense system. WGA's highest affinity is to terminal GlcNAc residues and it also has lower affinity to sialic acids (13). WGA forms dimers with two binding sites within each subunit. These act in a co-operative manner, such that WGA can bind tightly to glycans with multiple terminal GlcNAcs. Several researchers have tried to use WGA to create a method for O-GlcNAcylated peptide enrichment but found that the affinity was too low to perform bind-

and-elute enrichments. However, effective chromatographic methods based on retardation of modified peptides when passed through a column of WGA, referred to as lectin weak affinity chromatography (LWAC), have been developed. The first of these used WGA-agarose resin, but it was found that very long columns, up to 12 meters, were desirable to get useful separation (14). Two significant improvements have since been made. The first was using POROS resin, which provides a higher density of WGA per unit length, allowing shorter columns to be used. The second was the realization that the separation efficiency was hampered simply by the vast excess of unmodified peptides, meaning the tail of the unmodified peptides co-eluted with the modified peptides. By collecting this tail and then reloading for a second, or even third round of LWAC highly enriched glycopeptide fractions (> 70 % modified peptides) can be produced (15). WGA-based LWAC chromatography enriches nearly all types of N- and O-glycosylation; not just O-GlcNAc. In a whole cell lysate the O-GlcNAc modified peptides will be a subset of the total glycopeptides in the sample. Preparing a cytosolic, or nuclear preparation before performing WGA-LWAC will produce more highly enriched O-GlcNAc fractions, but even then, it is common to get contamination from ER, golgi and lysosomal proteins. Hence, it is important to confirm that peptides identified as being modified by a single HexNAc are O-GlcNAc and not N-GlcNAc or O-GalNAc modified.

Succinylated WGA (sWGA) does not bind sialic acid residues, so should have higher specificity for O-GlcNAc over some other glycoforms. However, our attempt to use this for LWAC was not effective. This is probably because sWGA has lower affinity than WGA for GlcNAc (16) such that the interaction is not strong enough for chromatographic enrichment.

### **ENZYMATIC LABELING**

The original galactosyltransferase labeling approach has been adapted as a method for modified peptide enrichment. The galactosyltransferase was mutated to allow addition of a galactose containing an unnatural ketone group that can be used to attach an affinity tag such as biotin (17). This tag was used

for measuring changes in O-GlcNAcylation in the brain following stimulation (18). However, the large tag hampers the ability to identify modified peptides as one gets extensive tag fragmentation rather than peptide backbone. Hence, a photocleavable cleavable linker was added, such that after enrichment the biotin moiety can be removed before mass spectrometric analysis (19). This method was used to study cross-talk between O-GlcNAcylation and phosphorylation during cytokinesis (20). A photocleavable tag has also been developed that incorporates isotopic labeling that allows quantitative comparison of two samples (21), and this was used to discover differences in O-GlcNAcylation in drug-resistant vs drug-sensitive HepG2 cells. More recently a variation of this protocol was developed that attached an affinity tag linked via a disulfide bond that could be reduced for tag cleavage and removal after enrichment (22). As described earlier, the original method for detecting O-GlcNAcylation was through the enzymatic labeling of terminal GlcNAc residues with galactose. This terminal galactose can be enriched using the lectin Ricinus Communis Agglutinin I (RCA I) such that labeled O-GlcNAc modified proteins can be enriched (23). This approach has mostly been used for enrichment of modified proteins, but has been demonstrated for modified peptide enrichment (24).

### **METABOLIC LABELING**

Rather than introducing a handle for tagging by enzymatic addition of a second sugar residue it is possible to introduce a functional group directly. The O-GlcNAc transferase is able to add an azide analog of GlcNAc referred to as GlcNAz, so one can study new O-GlcNAcylation through metabolic incorporation of this azide sugar, then enrichment and mass spectrometric analysis (25). An alkyne equivalent can also be employed in the same way (26).

An elaborate strategy for glycopeptide analysis is Isotope-targeted glycoproteomics (IsoTaG). This approach uses metabolic labeling to introduce the affinity handle, but then adds an isotope encoded tag containing two bromines (27). Bromine has two stable isotopes of masses 79 Da and 81 Da at roughly

equal abundance. As a result, peptides labeled by this tag produce a unique isotope pattern that allows them to be recognized in the mass spectrometer as modified. This can be useful if one does not have an efficient enrichment strategy such that only a small percentage of peptides are modified, as it allows one to produce a list of peptide masses that are modified that can then be targeted for fragmentation analysis in a subsequent run. Using this strategy for O-GlcNAc analysis just over two thousand modified peptides corresponding to several hundred modification sites were identified among tryptic and chymotryptic digests of T-cells (28).

A limitation of metabolic labeling is that one is competing with endogenous UDP-GlcNAc for incorporation, so stoichiometry of modification is never going to be high, leading to a significant sensitivity hit. In the large-scale IsoTaG study (28) enrichment was performed at the protein-level such that only one site in the protein needed to be GlcNAz-modified for enrichment. In the subsequent analysis many of the modification sites were identified as GlcNAc rather than GlcNAz-modified, exemplifying the incorporation issue. Protein-level enrichment leads to a high percentage of unmodified peptides being present, which is why the isotope pattern introduced using IsoTaG is useful to find modified peptides.

### **IMMUNOPRECIPITATION**

One limitation of all the enrichment methods described thus far is the lack of specificity for O-GlcNAc over other types of glycosylations containing GlcNAc. A few antibodies have been raised against O-GlcNAcylated proteins that have shown some affinity to O-GlcNAcylation in general. The best known of these are the RL2 antibody that was raised against nuclear pore complex proteins (which are heavily O-GlcNAcylated) (29) and CTD110.6 (10), which binds the heavily O-GlcNAcylated C-terminal domain of RNA Polymerase II. A set of monoclonal antibodies were produced using a designed three component immunogen including a synthetic O-GlcNAc-modified peptide for protein-level enrichment (11) and in a

subsequently study were used to identify 83 O-GlcNAc modification sites (30). However, it is clear that each of these monoclonal antibodies only recognize a subset of modified proteins.

To produce a specific pan-O-GlcNAc enrichment approach the Van Aalten group have attempted to make use of an enzymatically dead mutant of OGA(31). This version of the enzyme still binds O-GlcNAc but is unable to release the moiety. Using a GST-tagged version of their mutant OGA enzyme they could visualize modified proteins on a 1D gel with an anti-GST antibody, but it has yet to be demonstrated that this mutant OGA could be used for immunoprecipitating modified proteins or peptides.

### **BEMAD**

As discussed below, the highly labile nature of the glycosidic linkage in O-GlcNAc modified peptides makes modification site assignment extremely difficult in collision induced dissociation data. One approach to bypass the high lability of the glycosidic linkage is to chemically derivatize the modified site, replacing it with a more stable moiety which could also be used as an affinity handle. Strong base catalyzed beta-elimination of O-linked glycopeptides is a commonly used approach for releasing glycans for subsequent study. This reaction produces an unsaturated carbonyl on the formerly modified serine or threonine, which can then be modified by a nucleophilic attack reaction. This chemistry has been used to create an approach referred to as BEMAD (Beta Elimination followed by Michael Addition of dithiothreitol (DTT)). The DTT derivatization introduces a stable modification with a free thiol that can be used for enrichment using thiol chromatography. This allows for the identification of formerly O-GlcNAcylated sites (32). A complication of this approach is that the same beta elimination reaction occurs (albeit at a slower rate) for other O-linked modifications such as phosphorylation, alkylated cysteines, and it can even occur at low levels on unmodified serines and threonines. Hence, it is important to have additional information to confirm that the former modification was O-GlcNAc. One demonstrated approach to address this is to split a sample and treat half with a glycosidase. BEMAD with normal dithiothreitol (DTT) (d0) or



deuterated DTT (d6) is performed on each half. Formerly glycosylated peptides will produce a single peak, whereas formerly phosphorylated or unmodified serines will produce pairs of peaks at equal intensity (33). BEMAD has also been used for modification site localization after using a chemoenzymatic glycopeptide enrichment protocol (12).

Table 1 presents a summary of the discussed enrichment methods and their different strengths, weaknesses and biases.

### MASS SPECTROMETRY FOR O-GLCNAc ANALYSIS

Most methods for peptide fragmentation in the mass spectrometer involve introducing internal energy by colliding the peptides with neutral gas molecules. These collision-induced dissociation (CID) methods cause the weakest bonds in the molecule to break. Unfortunately, the weakest bonds in a glycopeptide are glycosidic linkages. When an O-GlcNAcylated peptide is subjected to CID the primary fragmentation is cleavage of the bond between the GlcNAc and the side-chain of the attached amino acid. This leads to fragment ions that correspond to the oxonium ion of the GlcNAc at  $m/z$  204.087 and the unmodified peptide (34). Fragmentation in an ion trap only puts energy into the precursor ion, so an ion trap fragmentation spectrum essentially only contains these ions. In a quadrupole collision cell such as in a quadrupole-TOF or quadrupole-Orbitrap instrument fragment ions can undergo further collisions, so fragments that are the product of peptide backbone cleavages are also observed (b and y ions, see Figure 1), allowing the peptide sequence to be determined, but as these are mostly deglycosylated fragment ions there is generally not enough information to determine the site of glycosylation.

One approach to try to circumvent this problem was to derivatize the modification site. Using a base-catalyzed beta-elimination reaction the formerly modified site is observed with a mass that is a water smaller than an unmodified residue. However, as mentioned above, this reaction can occur for other O-

linked modifications, so measuring the peptide before and after the elimination reaction was used to determine the modification sites (34).

A series of improvements in mass spectrometry instrumentation has since transformed the ability to identify O-GlcNAcylation sites directly. The first of these improvements was the commercial availability of quadrupole-TOF instruments. Compared to triple quadrupole or ion trap instruments, quadrupole-TOF instruments vastly improved signal-to-noise, such that low intensity O-glycosylated fragment ions could be detected for the first time (35). This allowed for the first identification of O-GlcNAc modification sites directly using tandem mass spectrometry (36).

The next instrument breakthroughs were the development and commercial availability of new fragmentation methods. Rather than breaking the weakest bonds in the molecule and producing b and y ions, these new methods break molecular bonds at sites of electron capture to produce c and z ions (see Figure 1). The first of these methods was electron capture dissociation (37) which allowed detection of 12 sites of O-GlcNAcylation from a mouse post-synaptic density preparation (14). It was then shown that electron transfer from an anion was more efficient than using a beam of electrons (38). Electron transfer dissociation (ETD) has since become the method of choice for O-GlcNAc peptide identification and site localization. The first demonstration of using ETD for O-GlcNAc analysis identified seven modification sites in neurons (18). By combining ETD analysis with lectin enrichment, 58 sites of O-GlcNAcylation were identified from mouse post-synaptic density (39). This roughly doubled the number of known modification sites at the time. Shortly thereafter, this list of sites was dwarfed by a more extensive study of O-GlcNAcylation and phosphorylation that reported over 1750 O-GlcNAc sites (15). Since then, a handful of studies have been published identifying hundreds of sites at a time (20, 28, 40–42). Some of these studies have made use of a combination of HCD and ETD data, where the HCD data is used to identify a precursor as being HexNAc modified based on the formation of the HexNAc oxonium ion at  $m/z$  204.087; then matched ETD data is used for identification of the peptide and modification site (30, 39). Acquisition

software can be told to only perform ETD fragmentation if a given precursor forms the HexNAc oxonium ion (or fragments of this ion), which is useful if the sample contains many unmodified peptides; for example if enrichment was performed at the protein rather than peptide level (30). As discussed below, newer studies of this type will likely make use of a combination of ETD and CID fragmentation referred to as EThcD (43). Some peptides fragment better by one or other type of activation, so by combining fragmentation types a higher percentage of precursors provide an informative spectrum, although the fragments formed by ETD are generally necessary for modification site localization.

### SOFTWARE FOR O-GLCNAC ANALYSIS

The diagnostic loss of the GlcNAc modification in CID can be used to identify spectra of modified peptides. The  $m/z$  204.087 GlcNAc oxonium ion is specific to glycopeptides. Unfortunately, this ion is produced from other types of glycosylation in addition to O-GlcNAcylation: N-linked glycosylation produces the same GlcNAc-derived fragment and extracellular O-GalNAc linked glycosylation produces the same mass ion. However, there are methods to differentiate O-GlcNAcylated peptide spectra from these other types of glycosylation. Extended glycans produce many other glycan oxonium ions, most notably an ion corresponding to HexNAcHex at mass 366.139 Da, so the presence of larger glycan ions can be used to exclude spectra from O-GlcNAc assignment. As stated, the GalNAc oxonium ion has the same mass as the GlcNAc oxonium ion. However, the two isomers both further fragment into a series of ions ( $m/z$  126.055, 138.055, 144.065, 168.066, and 186.076) and the relative intensity of these fragment ions can be used to differentiate between GlcNAc and GalNAc (44, 45) (and see Figure 2). Specifically, when O-GlcNAc further fragments the  $m/z$  138 ion is always significantly more intense than the  $m/z$  144 ion, whereas O-GalNAc produces these fragments at similar intensity. EThcD fragmentation spectra can both identify modification sites and differentiate between HexNAc isomers.

The O Score software was written to identify spectra as potential glycopeptides based on the presence of the  $m/z$  204.087 ion and its fragments (46). More powerful software along the same theme is MS-Filter in

Protein Prospector (47). When presented with a peak list from a mass spectrometry run MS-Filter can be used to create new peak list files that either contain or do not contain peaks corresponding to the ion(s) of interest. With multiple rounds of filtering, peak list files can be generated that include all spectra that contain the  $m/z$  204.087 peak while removing those that also contain the  $m/z$  366.139 peak (and potentially other glycan oxonium ions) to filter out N- and extended O-GalNAc linked glycopeptide spectra.

Most database search engines can be adapted to analyze O-GlcNAcylation. The primary complication is that in CID-type fragmentation the software needs account for the precursor ion to be shifted by the mass of the modification (+203.080 Da), but assume all fragments are unmodified, whereas in ETD it should assume the modification remains on the fragment ions. As such several search engines have been used for O-GlcNAc peptide identification including Protein Prospector (39), Mascot (12), OMSSA (22) and Byonic has been used for analysis of IsoTaG data (28). An important aspect of any software for O-GlcNAc analysis is that it evaluates modification site localization. In the case of Protein Prospector this is in-built (48); for some other software it may be necessary to run the results through separate tools to evaluate which sites can be reliably identified (49). EThcD data analysis of O-GlcNAc modified peptides is more complicated, as fragments from ETD activation are expected to retain modifications, but those from HCD activation will not. In-house analysis of EThcD data using Protein Prospector allowing for both loss and retention of the modification on fragments has indicated that modification site localization becomes complicated: if you observe a fragment modified due to ETD, but an equivalent fragment unmodified due to CID this can create ambiguity. In our experience assuming all fragments are modified in EThcD data produces more reliable modified peptide identifications and site localizations, even if it means a few fragment ions are not explained.

### **O-GLCNACYLATION SITE PREDICTION**

The human genome encodes over five hundred kinases for addition of a phosphate group to proteins. The fact that there is only a single O-GlcNAc transferase, OGT (albeit with three splice variants) that is responsible for thousands of modifications begs the question as to how it can have sequence specificity. Analyses of identified modification sites have shown a strong enrichment for a proline two or three amino acids prior to the modification site, and serines and threonines preferentially in all other surrounding positions. It is also common to find several residues modified in close proximity, so multiply-modified peptides are frequently detected. The proline residue introduces a kink in the protein chain which may make regions of the protein more accessible to an enzyme for modification, but OGT probably does not formally have any further sequence preference. OGT is often thought of as corresponding to a catalytic subunit that is part of a larger protein complex, the other components of which are responsible for targeting to specific proteins for modification.

A database of known O-GlcNAcylation sites, dbOGAP was published about ten years ago (50), but this has since disappeared. dbOGAP attempted to predict O-GlcNAcylation sites, and other tools have also been developed for modification site prediction (51), but these all have high error rates, so should be used with caution.

### **HOW A RECENT GLOBAL O-GLCNAcylation STUDY PERFORMS**

Some of the largest O-GlcNAc datasets created thus far have been from our group using lectin weak affinity chromatography with wheat germ agglutinin (15, 40). These studies employed ETciD for fragmentation analysis. As an example of the amount of information that can be achieved in global O-GlcNAcylation studies using the newer EThcD technique we present an experiment where O-GlcNAcylation was studied in human monocyte cells.

## O-GlcNAc Analysis

We started with 20 mg of THP1 whole cell lysate to enrich glycopeptides. In our experience, the amount of starting material required for LWAC varies from sample to sample, though in general the higher the amount of material the more O-GlcNAcylated peptides are recovered. We typically try to start with at least 10 mg of protein from a whole cell lysate, although we have had excellent results with as little as 5 mg starting material (unpublished results). The tryptic peptides from the whole cell lysate were split into 10 injections for glycopeptide enrichment using LWAC. The separate injections are to try to minimize overloading of the column, which reduces the resolution of separation of the glycosylated peptides from the unmodified majority. The tails of each of these runs were collected together using an in-line C18 column and eluted in one fraction. After this first enrichment step it is typical that about 5% of peptides are glycosylated. Two subsequent rounds of LWAC enrichment were then performed. The resulting enriched glycopeptide sample was fractionated offline by high pH reverse phase chromatography to generate 50 fractions. Pairs of these fractions were combined (fraction 1 + fraction 26, 2+27, etc.) to generate 25 fractions that were run on an Orbitrap Fusion Lumos using EThcD fragmentation. Based on the percent of obtained spectra that contained a HexNAc oxonium ion at  $m/z$  204.087, 41% of precursors were glycosylated. This is at the low end for enrichment using LWAC. We typically achieve 40-70% modified peptides depending on the sample.

We used MS-Filter to generate a set of HexNAc oxonium ion containing peak lists. These were searched against the human SwissProt database allowing for nearly 300 glycan compositions on N, S, and T. This resulted in the identification of over 8000 unique glycopeptides, the vast majority of which contained extended N- and O-linked sugars. These extended glycan IDs are useful to help distinguish N-GlcNAc and O-GalNAc modifications from O-GlcNAc modifications. We also used known subcellular localization if available and the intensity ratio of the HexNAc oxonium ion fragments discussed above to confirm HexNAc modifications as O-GlcNAc. After manually verifying the O-GlcNAc modified peptides, we identified just

over 1800 unique O-GlcNAc modified peptides from 420 proteins. We were also able to identify over 700 sites of O-GlcNAcylation from this dataset by thresholding at a 5% false localization rate threshold per spectrum using the SLIP scoring in Protein Prospector (48), which typically leads to around 1% incorrect site assignments at the dataset level. These O-GlcNAcylated peptides, and the unique modification sites localized are summarized in Supplementary Tables 1 and 2, and annotated spectra can be viewed through a web browser in MS-Viewer submission gi6ztunb9r (52).

Global analysis of O-GlcNAcylated peptides reveals information about what proteins and pathways in a given system are regulated by O-GlcNAc modification. Various analyses, such as interaction networks in STRING (53) or functional analyses with PANTHER (54) or DAVID (55), can be performed on these lists of proteins to infer the role of O-GlcNAcylation in the system being studied. Supplemental Figure 2 and Supplemental Table 4 show functional interaction network analysis of the O-GlcNAc modified proteins in our THP1 dataset created using stringApp and Cytoscape. Clear functional networks emerge including mRNA splicing, clathrin mediated endocytosis, mitotic cell cycle, and protein transport in the nucleus and endoplasmic reticulum (ER). The proteins found to be involved in ER protein transport are cytosolic-membrane associated proteins, not located within the ER (where one should not find any O-GlcNAc modification). High confidence site localization is of great importance to researchers attempting to study mechanistic insights into particular O-GlcNAcylated proteins.

### **CONCLUSIONS**

The ability to analyze O-GlcNAcylation of proteins on a large-scale, as well as to reliably identify exact sites of modification has been transformed by the development of enrichment strategies and new fragmentation techniques in the mass spectrometer. Several enrichment methods have been developed, but there is no one method that can selectively enrich all O-GlcNAc modified peptides

without also enriching other types of glycosylation. Nevertheless, due to the fast speed of modern mass spectrometers it is now possible to routinely identify many hundreds of O-GlcNAc modified peptides in a single study, although modification site localization within peptides is still not a formality and must be assessed. ETD (or hybrid fragmentation such as EThcD) is generally necessary for modification site localization, but even using these fragmentation techniques some spectra provide ambiguous results as to the exact site of modification. Hence, use of software that can estimate false-localization rates is essential (49). Differentiating between O-GlcNAc and other single HexNAc modifications is also required. Using site localization software can help recognize N-linked GlcNAc. In addition, the presence of the consensus N-glycosylation motif (N-!P-S/T, where !P is anything other than a proline) is a strong clue (although O-GlcNAc has been found on serines and threonines in this motif). To differentiate O-GlcNAc from O-GalNAc the relative intensity of their fragment ions can be used. Knowing the protein localization also allows one to assign O-GlcNAc over other types of glycosylation. Availability of a database of previously identified O-GlcNAc proteins and sites to replace the previously useful dbOGAP would be a useful resource for the field. In the future all these types of information for glycosylation type differentiation could be assessed automatically using software. Nevertheless, it is now possible to perform global quantitative analyses of O-GlcNAc modification in response to stress or stimulation, so an improved understanding of the many signaling pathways regulated by O-GlcNAcylation is starting to be uncovered.

### **ACKNOWLEDGEMENTS**

We thank Juan Osés-Prieto and Shreya Chand for supplying THP1 tryptic peptides. This work was financially supported by the Dr Miriam & Sheldon G. Adelson Medical Research Foundation, the Howard Hughes Medical Institute and the UCSF Program for Breakthrough Biomedical Research.



## DATA AVAILABILITY

Raw data has been uploaded to the MassIVE repository with the identifier MSV000085653.

Annotated spectra, peak lists, and the table of results for the annotated spectra can be viewed and downloaded from MS-Viewer with the keyword gi6ztunb9r (52).

## EXPERIMENTAL PROCEDURES

### Sample Preparation.

THP1 cells were grown in RPMI 1640 supplemented with 5% fetal calf serum, 1X penicillin/streptomycin, 1X glutamine, and 1X fungizone. Cells were pelleted, washed with PBS twice, and then sonicated in 50 mM ammonium bicarbonate containing 8M urea, 4X Phosphatase Inhibitor Cocktails II and III (Sigma-Aldrich), and 40 mM PUGNAc (Tocris Bioscience). Protein concentrations were estimated with bicinchoninic acid (BCA) protein assay (ThermoFisher Scientific). The protein lysate was reduced for 1 hour at room temperature with 5 mM DTT and subsequently alkylated using 10mM iodoacetamide for 45 min at room temperature in the dark. Lysates were diluted to 2M urea using 50 mM ammonium bicarbonate, pH 8.0, and digested overnight at room temperature with sequencing grade trypsin (ThermoFisher Scientific) at an enzyme to substrate ratio of 1:50 (w/w). Following digestion, samples were acidified using formic acid (Sigma-Aldrich), desalted using a 35 cc C18 Sep-Pak SPE cartridge (Waters), and dried to completeness using a SpeedVac concentrator (Thermo).

**Lectin Weak Affinity Chromatography.** Glycopeptides were enriched as described previously (15, 45). Briefly, 20 mg of desalted tryptic peptides were resuspended in 1000  $\mu$ L LWAC buffer (100 mM Tris pH 7.5, 150 mM NaCl, 2 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>, 5% acetonitrile) and 100  $\mu$ L were run over a 2.0 x 250mm POROS-WGA column at 100  $\mu$ L/min under isocratic conditions with LWAC buffer and eluted with a 100  $\mu$ L injection of 40 mM GlcNAc. Glycopeptides were collected inline on a C18 column (Phenomenex, Torrance,

CA). Enriched glycopeptides from 10 initial rounds of LWAC were eluted with 50% acetonitrile, 0.1% FA in a single 500  $\mu$ L fraction, dried, and LWAC enrichment was repeated for a total of 3 enrichment steps.

**Offline Fractionation.** Glycopeptides were separated on a 4.0  $\times$  150 mm Gemini 5 $\mu$  110A C18 column (Phenomenex, Torrance, CA). Peptides were loaded onto the column in 20 mM NH<sub>4</sub>OCH<sub>3</sub>, pH 10 and subjected to a gradient from 2% to 9% over 2 mL then from 9% to 50% 20mM NH<sub>4</sub>OCH<sub>3</sub>, pH10 in 90% acetonitrile over 20 mL collecting 50 fractions.

**Mass Spectrometry Analysis.** Glycopeptides were analyzed on an Orbitrap Fusion Lumos (Thermo Scientific, San Jose, CA) equipped with a NanoAcquity UPLC (Waters, Milford, MA). Peptides were fractionated on a 50cm  $\times$  75  $\mu$ m ID 2 $\mu$ m C18 EASY-Spray column using a linear gradient from 3.5-30% solvent B over 185 min. Precursor ions were measured from 375 to 1500 m/z in the Orbitrap analyzer (resolution: 120,000; AGC: 4.0e5). Each precursor ion (charged 2-7+) was isolated in the quadrupole (selection window: 1.6 m/z; dynamic exclusion window: 30 s; MIPS Peptide filter enabled) and underwent EThcD fragmentation (Maximum Injection Time: 250 ms, Supplemental Activation Collision Energy: 25%) measured in the Orbitrap (resolution: 30,000; AGC; 5.0e4). The scan cycle was 3 seconds.

Peak lists for EThcD were extracted using Proteome Discoverer 2.2. EThcD peak lists were filtered with MS-Filter and only spectra containing a 204.0867 m/z peak corresponding to the HexNAc oxonium ion were used for database searching. EThcD data was searched against human entries in the SwissProt protein database downloaded on July 31, 2019 concatenated with a randomized sequence for each entry (a total of 40862 sequences searched) using Protein Prospector (v6.1.10). Cleavage specificity was set as tryptic, allowing for 2 missed cleavages. Carbamidomethylation of Cys was set as a constant modification. The required mass accuracy was 10 ppm for precursor ions and 30 ppm for fragment ions. Variable

## O-GlcNAc Analysis

modifications included methionine oxidation, protein N-terminus methionine removal and/or acetylation, pyroglutamate formation from peptide N-terminal glutamine and a total of 295 different N- or O-linked glycosylation compositions, which are listed in Supplementary Table S3. For N-linked glycosylations they were required to occur in the motif N-!P-S/T, where !P is any amino acid other than proline. One N-linked and up to four O-linked modifications per peptide were permitted. Unambiguous PTMs were determined using a minimum SLIP score of six, which corresponds to a 5% local false localization rate (48). Modified peptides were identified with a peptide false discovery rate of 1%. O-GlcNAc and O-GalNAc modifications were differentiated based on known protein subcellular localization and HexNAc oxonium ion fragment ratios (44).

**O-GlcNAc Network Analysis.** Network analysis was performed in Cytoscape v3.7.1 with stringApp v1.4.2. The network analysis used a confidence score cutoff of 0.9 and the functional enrichment analysis used an FDR value cutoff of 0.05.

## REFERENCES

1. Zachara, N. E., and G. W. Hart. 2006. Cell signaling, the essential role of O-GlcNAc! *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*. **1761**: 599–617.
2. Hardivillé, S., and G. W. Hart. 2014. Nutrient Regulation of Signaling, Transcription, and Cell Physiology by O-GlcNAcylation. *Cell Metab*. **20**: 208–213.
3. Hart, G. W., C. Slawson, G. Ramirez-Correa, and O. Lagerlof. 2011. Cross Talk Between O-GlcNAcylation and Phosphorylation: Roles in Signaling, Transcription, and Chronic Disease. *Annu Rev Biochem*. **80**: 825–858.
4. Hart, G. W. 2019. Nutrient regulation of signaling and transcription. *J. Biol. Chem*. **294**: 2211–2231.
5. Torres, C. R., and G. W. Hart. 1984. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J. Biol. Chem*. **259**: 3308–3317.
6. Reason, A. J., H. R. Morris, M. Panico, R. Marais, R. H. Treisman, R. S. Haltiwanger, G. W. Hart, W. G. Kelly, and A. Dell. 1992. Localization of O-GlcNAc modification on the serum response transcription factor. *J. Biol. Chem*. **267**: 16911–16921.
7. Dong, D. L.-Y., Z.-S. Xu, G. W. Hart, and D. W. Cleveland. 1996. Cytoplasmic O-GlcNAc Modification of the Head Domain and the KSP Repeat Motif of the Neurofilament Protein Neurofilament-H. *J. Biol. Chem*. **271**: 20845–20852.
8. Pedowitz, N. J., B. W. Zaro, and M. R. Pratt. 2020. Metabolic Labeling for the Visualization and Identification of Potentially O-GlcNAc-Modified Proteins. *Current Protocols in Chemical Biology*. **12**: e81.
9. Darabedian, N., J. W. Thompson, K. N. Chuh, L. C. Hsieh-Wilson, and M. R. Pratt. 2018. Optimization of Chemoenzymatic Mass Tagging by Strain-Promoted Cycloaddition (SPAAC) for the Determination of O-GlcNAc Stoichiometry by Western Blotting. *Biochemistry*. **57**: 5769–5774.
10. Comer, F. I., K. Vosseller, L. Wells, M. A. Accavitti, and G. W. Hart. 2001. Characterization of a Mouse Monoclonal Antibody Specific for O-Linked N-Acetylglucosamine. *Analytical Biochemistry*. **293**: 169–177.
11. Teo, C. F., S. Ingale, M. A. Wolfert, G. A. Elsayed, L. G. Nöt, J. C. Chatham, L. Wells, and G.-J. Boons. 2010. Glycopeptide specific monoclonal antibodies suggest new roles for O-GlcNAc. *Nat Chem Biol*. **6**: 338–343.
12. Hahne, H., N. Sobotzki, T. Nyberg, D. Helm, V. S. Borodkin, D. M. F. van Aalten, B. Agnew, and B. Kuster. 2013. Proteome Wide Purification and Identification of O-GlcNAc-Modified Proteins Using Click Chemistry and Mass Spectrometry. *J. Proteome Res*. **12**: 927–936.
13. Bhavanandan, V. P., and A. W. Katlic. 1979. The interaction of wheat germ agglutinin with sialoglycoproteins. The role of sialic acid. *J. Biol. Chem*. **254**: 4000–4008.
14. Vosseller, K., J. C. Trinidad, R. J. Chalkley, C. G. Specht, A. Thalhammer, A. J. Lynn, J. O. Snedecor, S. Guan, K. F. Medzihradzsky, D. A. Maltby, R. Schoepfer, and A. L. Burlingame. 2006. O-linked N-

- acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol. Cell Proteomics*. **5**: 923–934.
15. Trinidad, J. C., D. T. Barkan, B. F. Gullledge, A. Thalhammer, A. Sali, R. Schoepfer, and A. L. Burlingame. 2012. Global Identification and Characterization of Both O-GlcNAcylation and Phosphorylation at the Murine Synapse. *Molecular & Cellular Proteomics*. **11**: 215–229.
16. Zachara, N. E., K. Vosseller, and G. W. Hart. 2011. Detection and Analysis of Proteins Modified by O-Linked N-Acetylglucosamine. *Curr Protoc Protein Sci*. **CHAPTER**: Unit12.8.
17. Khidekel, N., S. Arndt, N. Lamarre-Vincent, A. Lippert, K. G. Poulin-Kerstien, B. Ramakrishnan, P. K. Qasba, and L. C. Hsieh-Wilson. 2003. A Chemoenzymatic Approach toward the Rapid and Sensitive Detection of O-GlcNAc Posttranslational Modifications. *J. Am. Chem. Soc.* **125**: 16162–16163.
18. Khidekel, N., S. B. Ficarro, P. M. Clark, M. C. Bryan, D. L. Swaney, J. E. Rexach, Y. E. Sun, J. J. Coon, E. C. Peters, and L. C. Hsieh-Wilson. 2007. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nature Chemical Biology*. **3**: 339–348.
19. Wang, Z., N. D. Udeshi, M. O'Malley, J. Shabanowitz, D. F. Hunt, and G. W. Hart. 2010. Enrichment and Site Mapping of O-Linked N-Acetylglucosamine by a Combination of Chemical/Enzymatic Tagging, Photochemical Cleavage, and Electron Transfer Dissociation Mass Spectrometry. *Molecular & Cellular Proteomics*. **9**: 153–160.
20. Wang, Z., N. D. Udeshi, C. Slawson, P. D. Compton, K. Sakabe, W. D. Cheung, J. Shabanowitz, D. F. Hunt, and G. W. Hart. 2010. Extensive Crosstalk Between O-GlcNAcylation and Phosphorylation Regulates Cytokinesis. *Sci. Signal*. **3**: ra2–ra2.
21. Li, J., Z. Li, X. Duan, K. Qin, L. Dang, S. Sun, L. Cai, L. C. Hsieh-Wilson, L. Wu, and W. Yi. 2019. An Isotope-Coded Photocleavable Probe for Quantitative Profiling of Protein O-GlcNAcylation. *ACS Chem. Biol.* **14**: 4–10.
22. Ma, J., W.-H. Wang, Z. Li, J. Shabanowitz, D. F. Hunt, and G. W. Hart. 2019. O-GlcNAc Site Mapping by Using a Combination of Chemoenzymatic Labeling, Copper-Free Click Chemistry, Reductive Cleavage, and Electron-Transfer Dissociation Mass Spectrometry. *Anal Chem*. **91**: 2620–2625.
23. Roquemore, E. P., T.-Y. Chou, and G. W. Hart. 1994. *In Methods in Enzymology Guide to Techniques in Glycobiology*. pp. 443–460. , Academic Press. [online]  
<http://www.sciencedirect.com/science/article/pii/0076687994300283> (Accessed August 19, 2020).
24. Hayes, B. K., K. D. Greis, and G. W. Hart. 1995. Specific Isolation of O-Linked N-Acetylglucosamine Glycopeptides from Complex Mixtures. *Analytical Biochemistry*. **228**: 115–122.
25. Nandi, A., R. Sprung, D. K. Barma, Y. Zhao, S. C. Kim, J. R. Falck, and Y. Zhao. 2006. Global Identification of O-GlcNAc-Modified Proteins. *Anal. Chem*. **78**: 452–458.
26. Zaro, B. W., Y.-Y. Yang, H. C. Hang, and M. R. Pratt. 2011. Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1. *PNAS*. **108**: 8146–8151.

27. Woo, C. M., A. T. Iavarone, D. R. Spiciarich, K. K. Palaniappan, and C. R. Bertozzi. 2015. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nature Methods*. **12**: 561–567.
28. Woo, C. M., P. J. Lund, A. C. Huang, M. M. Davis, C. R. Bertozzi, and S. J. Pitteri. 2018. Mapping and Quantification of Over 2000 O-linked Glycopeptides in Activated Human T Cells with Isotope-Targeted Glycoproteomics (Isotag). *Molecular & Cellular Proteomics*. **17**: 764–775.
29. Snow, C. M., A. Senior, and L. Gerace. 1987. Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. *J Cell Biol*. **104**: 1143–1156.
30. Zhao, P., R. Viner, C. F. Teo, G.-J. Boons, D. Horn, and L. Wells. 2011. Combining High-Energy C-Trap Dissociation and Electron Transfer Dissociation for Protein O-GlcNAc Modification Site Assignment. *J. Proteome Res*. **10**: 4088–4104.
31. Mariappa, D., N. Selvan, V. S. Borodkin, J. Alonso, A. T. Ferenbach, C. Shepherd, I. H. Navratilova, and D. M. F. van Aalten. 2015. A mutant O-GlcNAcase as a probe to reveal global dynamics of protein O-GlcNAcylation during *Drosophila* embryonic development. *Biochem J*. **470**: 255–262.
32. Wells, L., K. Vosseller, R. N. Cole, J. M. Cronshaw, M. J. Matunis, and G. W. Hart. 2002. Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell Proteomics*. **1**: 791–804.
33. Vosseller, K., K. C. Hansen, R. J. Chalkley, J. C. Trinidad, L. Wells, G. W. Hart, and A. L. Burlingame. 2005. Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics*. **5**: 388–398.
34. Greis, K. D., B. K. Hayes, F. I. Comer, M. Kirk, S. Barnes, T. L. Lowary, and G. W. Hart. 1996. Selective Detection and Site-Analysis of O-GlcNAc-Modified Glycopeptides by  $\beta$ -Elimination and Tandem Electrospray Mass Spectrometry. *Analytical Biochemistry*. **234**: 38–49.
35. Alving, K., H. Paulsen, and J. Peter-Katalinic. 1999. Characterization of O-glycosylation sites in MUC2 glycopeptides by nanoelectrospray QTOF mass spectrometry. *Journal of Mass Spectrometry*. **34**: 395–407.
36. Chalkley, R. J., and A. L. Burlingame. 2001. Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. *J. Am. Soc. Mass Spectrom*. **12**: 1106–1113.
37. Zubarev, R. A., D. M. Horn, E. K. Fridriksson, N. L. Kelleher, N. A. Kruger, M. A. Lewis, B. K. Carpenter, and F. W. McLafferty. 2000. Electron Capture Dissociation for Structural Characterization of Multiply Charged Protein Cations. *Anal. Chem*. **72**: 563–573.
38. Syka, J. E. P., J. J. Coon, M. J. Schroeder, J. Shabanowitz, and D. F. Hunt. 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *PNAS*. **101**: 9528–9533.
39. Chalkley, R. J., A. Thalhammer, R. Schoepfer, and A. L. Burlingame. 2009. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *PNAS*. **106**: 8894–8899.

40. Xu, S.-L., R. J. Chalkley, J. C. Maynard, W. Wang, W. Ni, X. Jiang, K. Shin, L. Cheng, D. Savage, A. F. R. Hühmer, A. L. Burlingame, and Z.-Y. Wang. 2017. Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* **114**: E1536–E1543.
41. Alfaro, J. F., C.-X. Gong, M. E. Monroe, J. T. Aldrich, T. R. W. Clauss, S. O. Purvine, Z. Wang, D. G. Camp, J. Shabanowitz, P. Stanley, G. W. Hart, D. F. Hunt, F. Yang, and R. D. Smith. 2012. Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets. *PNAS.* **109**: 7280–7285.
42. Maynard, J. C., H. Fujihira, G. E. Dolgonos, T. Suzuki, and A. L. Burlingame. 2020. Cytosolic N-GlcNAc proteins are formed by the action of endo- $\beta$ -N-acetylglucosaminidase. *Biochemical and Biophysical Research Communications.* **530**: 719–724.
43. Frese, C. K., A. F. M. Altelaar, H. van den Toorn, D. Nolting, J. Griep-Raming, A. J. R. Heck, and S. Mohammed. 2012. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **84**: 9668–9673.
44. Halim, A., U. Westerlind, C. Pett, M. Schorlemer, U. Rüetschi, G. Brinkmalm, C. Sihlbom, J. Lengqvist, G. Larson, and J. Nilsson. 2014. Assignment of Saccharide Identities through Analysis of Oxonium Ion Fragmentation Profiles in LC–MS/MS of Glycopeptides. *J. Proteome Res.* **13**: 6024–6032.
45. Kim, S., J. C. Maynard, Y. Sasaki, A. Strickland, D. L. Sherman, P. J. Brophy, A. L. Burlingame, and J. Milbrandt. 2016. Schwann Cell O-GlcNAc Glycosylation Is Required for Myelin Maintenance and Axon Integrity. *J. Neurosci.* **36**: 9633–9646.
46. Hahne, H., and B. Kuster. 2011. A novel two-stage tandem mass spectrometry approach and scoring scheme for the identification of O-GlcNAc modified peptides. *J. Am. Soc. Mass Spectrom.* **22**: 931–942.
47. Medzihradzky, K. F., K. Kaasik, and R. J. Chalkley. 2015. Characterizing sialic acid variants at the glycopeptide level. *Anal. Chem.* **87**: 3064–3071.
48. Baker, P. R., J. C. Trinidad, and R. J. Chalkley. 2011. Modification site localization scoring integrated into a search engine. *Mol. Cell Proteomics.* **10**: M111.008078.
49. Chalkley, R. J., and K. R. Clauser. 2012. Modification site localization scoring: strategies and performance. *Mol. Cell Proteomics.* **11**: 3–14.
50. Wang, J., M. Torii, H. Liu, G. W. Hart, and Z.-Z. Hu. 2011. dbOGAP - An Integrated Bioinformatics Resource for Protein O-GlcNAcylation. *BMC Bioinformatics.* **12**: 91.
51. Jia, C., Y. Zuo, and Q. Zou. 2018. O-GlcNAcPRED-II: an integrated classification algorithm for identifying O-GlcNAcylation sites based on fuzzy undersampling and a K-means PCA oversampling technique. *Bioinformatics.* **34**: 2029–2036.
52. Baker, P. R., and R. J. Chalkley. 2014. MS-viewer: a web-based spectral viewer for proteomics results. *Mol. Cell Proteomics.* **13**: 1392–1396.
53. Szklarczyk, D., A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, and C. von Mering. 2019. STRING v11: protein–protein

association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**: D607–D613.

54. Mi, H., A. Muruganujan, X. Huang, D. Ebert, C. Mills, X. Guo, and P. D. Thomas. 2019. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols.* **14**: 703–721.

55. Huang, D. W., B. T. Sherman, and R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols.* **4**: 44–57.

56. Boyce, M., I. S. Carrico, A. S. Ganguli, S.-H. Yu, M. J. Hangauer, S. C. Hubbard, J. J. Kohler, and C. R. Bertozzi. 2011. Metabolic cross-talk allows labeling of O-linked  $\beta$ -N-acetylglucosamine-modified proteins via the N-acetylgalactosamine salvage pathway. *Proc Natl Acad Sci U S A.* **108**: 3141–3146.



## O-GlcNAc Analysis

Enrichment	Binding Affinity	Specificity	Enriches all O-GlcNAc	Comments	References
<b>WGA Lectin</b>	Weak	Enriches most glycopeptides	Yes	Probably the highest sensitivity of the methods due to not requiring a labeling step before enrichment	(14, 15, 39, 40, 42, 45)
<b>BEMAD</b>	Medium	Enriches most O-linked modifications	Yes	As original modification is replaced, it is important to have independent verification that it was O-GlcNAc	(32, 33)
<b>Enzymatic Tagging</b>	Strong	Enriches terminal GlcNAc-containing glycopeptides	Yes		(17, 19, 20, 22, 41)
<b>Metabolic Tagging</b>	Strong	Enriches all types of GlcNAc-containing glycopeptides	Subset of newly-modified	Lower sensitivity due to competing UDP-GlcNAc levels leading to sub-stoichiometric incorporation.	(25, 28, 56)
<b>Immunoprecipitation</b>	Weak	Enriches subset of O-GlcNAc	Each enriches a subset	Will only enrich some modified peptides	(10, 11, 30)

Table 1: Summary of O-GlcNAc enrichment methods.

## FIGURE LEGENDS

Figure 1. Major bond cleavages in CID and ETD of glycopeptides. The primary cleavage in CID fragmentation of HexNAc modified peptides is of the glycosidic bond resulting in a 204.087 Da oxonium ion.

Figure 2. O-GlcNAc and O-GalNAc modified peptides can be differentiated in EThcD spectra. A) Annotated EThcD spectrum of O-GlcNAc modified peptide from Host Cell Factor 1. The fragments  $c_2$  and  $z+1_{14}$  localize the HexNAc moiety to serine 1150. Ions with a glycan loss are represented with an \*. B) Low mass region of the spectrum in A. The intensity of  $m/z$  138 is much higher than that of  $m/z$  144, indicating the modification is O-GlcNAc. C) Low mass region of spectrum shown in Supplemental Figure 1. Ions at  $m/z$  138 and  $m/z$  144 are of similar intensity, suggesting the peptide is O-GalNAc modified.

# O-GlcNAc Analysis

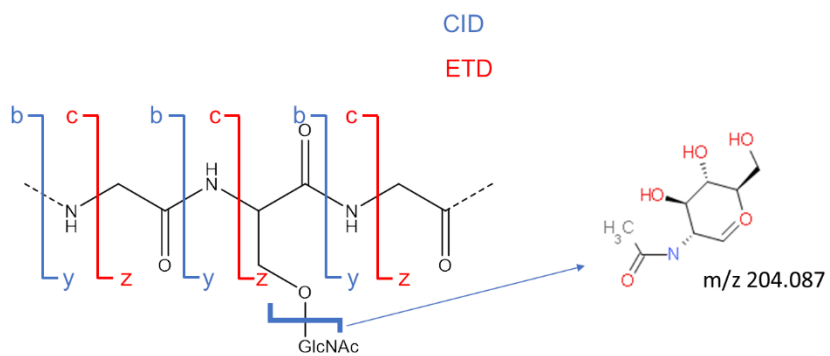


Figure 1

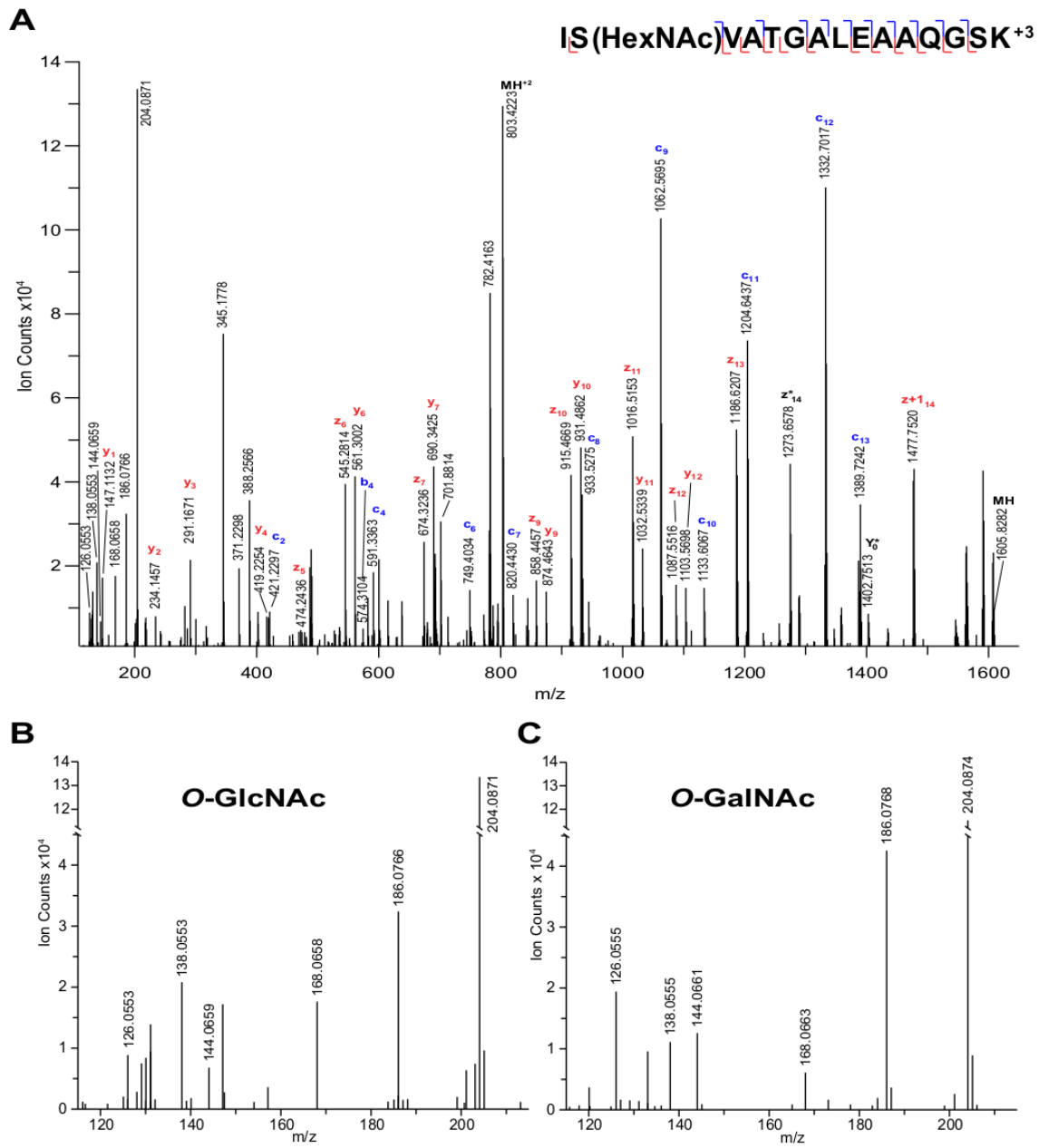


Figure 2