Running Title: Synaptic GlcNAcylation and Phosphorylation

Global Identification and Characterization of Both O-GlcNAcylation and Phosphorylation at the Murine Synapse.

Jonathan C. Trinidad[1], David T. Barkan[1,2,3,4], Brittany F. Gulledge[5], Agnes Thalhammer[6], Andrej Sali[1,2,4], Ralf Schoepfer[6], Alma L. Burlingame[1]

[1]Department of Pharmaceutical Chemistry; [2]Department of Bioengineering and Therapeutic Sciences; [3]Graduate Group in Bioinformatics; [4]California Institute of Quantitative Biosciences; [5]Cardiovascular Research Institute and Cellular and Molecular Pharmacology, UCSF, San Francisco, CA 94158 and the [6]Laboratory for Molecular Pharmacology, Department of Neuroscience, Physiology and Pharmacology, UCL, Gower Street, London WC1E 6BT, UK; Tel +44-20-76797242, Fax +44-20-76797245

Corresponding Authors: AL Burlingame; Ralf Schoepfer

Abbreviations: Define Secondary Structure of Proteins (DSSP); electron capture dissociation (ECD); electron transfer dissociation (ETD); exponentially modified protein abundance index (emPAI);false discovery rate (FDR); higher-energy C-trap dissociation (HCD); lectin weak affinity chromatography (LWAC); O-glycoprotein 2-acetamido-2-deoxy-β-D-glucopyranosidase, referred to as O-GlcNAcase (OGA); O-linked N-acetylglucosamine (O-GlcNAc); O-GlcNAc-transferase (OGT); post-translational modification (PTM); tetratricopeptide repeat (TPR); wheat germ agglutinin (WGA)

Summary

O-linked N-acetylglucosamine (O-GlcNAc) is a dynamic, reversible monosaccharide modifier of serine and threonine residues on intracellular protein domains. Crosstalk between O-GlcNAcylation- and phosphorylation has been hypothesized. Here, we identified over 1,750 and 16,500 sites of O-GlcNAcylation and phosphorylation from murine synaptosomes, respectively. In total, 135 (7%) of all O-GlcNAcylation sites were also found to be sites of phosphorylation, although we estimate that O-GlcNAcylated serine and threonine residues are phosphorylated at similar rates to non-O-GlcNAcylated serine and threonine residues.

While many proteins were extensively phosphorylated and minimally O-GlcNAcylated, proteins found to be extensively O-GlcNAcylated were almost always phosphorylated to a similar or greater extent, indicating the O-GlcNAcylation system is specifically targeting a subset of the proteome that is also phosphorylated. Both PTMs usually occur on disordered regions of protein structure, within which, the location of O-GlcNAcylation and phosphorylation is virtually random with respect to each other, suggesting that negative crosstalk at the structural level is not a common phenomenon. As a class, protein kinases are found to be more extensively O-GlcNAcylated than proteins in general, indicating the potential for crosstalk of phosphorylation with O-GlcNAcylation via regulation of enzymatic activity.

Introduction

O-GlcNAcylation, the addition of a single sugar (β-N-acetylglucosamine) to serine and threonine residues on intracellular domains of proteins, is a regulated, reversible post-translational modification. The O-GlcNAcylation state of proteins is responsive to numerous cellular stimuli, including nutrient levels and stress. The addition of this post-translational modification is catalyzed by a single enzyme known as uridine diphosphate-N-acetylglucosamine:peptide β-N-acetylglucosaminyltransferase, referred to as O-GlcNAc-transferase (OGT), and it is removed by a single enzyme known as O-glycoprotein 2-acetamido-2-deoxy-β-D-glucopyranosidase, referred to as O-GlcNAcase (OGA) (1). Both of these enzymes are highly expressed in the brain. The physiological roles of protein O-GlcNAcylation may be particularly important in the central nervous system (2) (3). OGT is present in dendrites and axon terminals; it is also associated with microtubules (4). Neuron-specific deletion of OGT results in neonatal lethality due in part to abnormal neuronal development and motor deficits (5).

Because O-GlcNAcylation modifies serine and threonine side chains, there is the potential for interplay between the function(s) of this moiety and those of phosphorylation. Over 1000 proteins have been identified as O-GlcNAc modified. While the majority of these are also phosphorylated (6), the implications of this are unclear given that the majority of all cellular proteins are probably phosphorylated. In most cases, the exact site of O-GlcNAc modification within a protein is still unknown. Traditional biochemical analysis has revealed numerous proteins that have been shown to be both phosphorylated and O-GlcNAcylated, including c-Myc, nitric oxide synthase, RNA polymerase II, synapsin I, tau, and amyloid precursor protein (reviewed in (6). In cell culture, modulation of the global levels of phosphorylation is accompanied by changes in O-GlcNAcylation levels of many proteins, and vice versa (7) (8). However, the specific sites involved have not been reported (6). Furthermore, pharmacological inhibition of a kinase causes a complex response, with an increase in O-GlcNAcylation of some proteins and a decrease in others (9). A mechanistic explanation of results from these experiments is lacking.

Cross talk has been proposed at the cellular level based on experiments inhibiting or promoting either O-GlcNAc or phosphorylation levels directly and examining what effect these treatments have on levels of the other PTM. For example, cells have been treated with lithium to inhibit GSK-3 (9), resulting in identification of 10 and 19 proteins with increased and decreased levels of O-GlcNAcylation, respectively. Inhibition of the O-GlcNAcase in NIH-3T3 cells resulted in elevated O-GlcNAcylation levels, which in turn was correlated with increases or decreases in over 50% of the measured phosphorylation sites (7). Over-expression of the O-GlcNAc transferase resulted in decreased phosphorylation in 17% of phosphorylation sites and increased phosphorylation at 7% of sites (10). Interpretation of these studies is complicated by the fact that the stimulation approaches were non-physiological and resulted in widespread changes in cellular physiology. For example, if over-expression of OGT altered activity of a kinase, all phosphorylation changes, however far downstream, might be interpreted as evidence for crosstalk between the PTM systems. In general, these studies have not established unambiguously that individual protein molecules showing an increase in O-GlcNAcylation also had a change in phosphorylation (or vice versa).

Driven by advances in affinity chromatography and the development of several generations of even more powerful tandem mass spectrometers (11) (12), our knowledge of the complexity and extent of cellular phosphorylation is still growing dramatically. In contrast, our knowledge of O-GlcNAcylation is progressing slowly, due to less robust enrichment methodologies and lack of suitable, broadly applicable and sensitive mass spectrometric methodologies.

Early tandem mass spectrometric studies required relatively large amounts of individual proteins and were successful in the assignment of sites of O-GlcNAc modifications in α-crystallin (13) as well as of both O-GlcNAc and phosphorylation sites in isolated serum response factor (14). These studies on O-GlcNAcylation employed traditional collision-induced dissociation (CID); however, this common internal energy deposition technique suffered from the instability of the sugar group upon vibronic energy deposition in the gas phase (15). Therefore, while O-GlcNAcylated peptides could be identified readily by the shift in their molecular weights by 203 Da, the

assignment of the exact site of modification within a peptide required care in adjustment of minimal collision energy. Later, using the widely available CID MS technique, derivatization strategies such as beta-elimination/Michael addition were pursued to alleviate the site assignment problem (16) (17). Recent work has addressed the need for O-GlcNAc-specific peptide enrichment based on both lectin chromatography[a] (18) (15) (19) and derivatization of O-GlcNAcylated peptides with biotinylated reagents (20) (10). The recent electron capture and electron transfer dissociation (ECD and ETD, respectively) techniques allow dissociation of peptide backbone linkages without causing elimination of O-GlcNAc-moieties from O-GlcNAcylated peptides. Thus, the formation of O-GlcNAc-containing sequence ion series permits unambiguous site localization without the need for derivatization (15).

In the present work, we established a workflow that permits the combined detection and determination of O-GlcNAcylation and phosphorylation sites on proteins in the *same* biological sample. Using this approach, we extensively characterized mouse synaptosome preparations. We identified 6621 proteins, including 1,750 sites of O-GlcNAcylation and 16,500 sites of phosphorylation. We estimate that 19% and 63% of synaptosome proteins are O-GlcNAcylated and phosphorylated, respectively. These results permit the first statistically robust analyses regarding crosstalk between O-GlcNAcylation and phosphorylation at the structural level. In addition, kinases are more frequently targets of OGT than proteins in general, demonstrating crosstalk at the catalytic level.

---

[a] New insight on GlcNAcylation sites from IRS2 and proteins from osteoblast cell lysates were also recently reported (L Ball, M Schilling, A Nagel, L Waller, S Comte-Walters, Characterization of O-GlcNAc peptides by electron transfer dissociation MS/MS" Abstracts 9th Uppsala Conf on Electron Capture and Transfer Dissociation, Feb 2012, Charleston, SC USA)

Experimental Procedures

Preparation of mouse synaptic membranes

Synaptic membrane samples were purified at 4ºC, as described previously (21) in the presence of the O-GlcNAcase inhibitor PUGNAc (Toronto Research Chemicals, North York, ON, Canada) and a cocktail of phosphatase inhibitors throughout the preparation. Briefly, brains from adult mice (strain C57BL/6J) were dissected; the cerebellum was removed and the brains immediately frozen in liquid nitrogen. Material from several animals was combined prior to the biochemical purification. The brain tissue was homogenized in a sucrose buffer containing phosphatase inhibitors (1 mM Na_3VO_4, 1 mM NaF, 1 mM Na_2MoO_4, 4 mM sodium tartrate, 100 nM fenvalerate, 250 nM okadaic acid) and 20 μM PUGNAc, and cleared by centrifugation. 10 ml of buffer was used per gram of brain. The membranous fraction was layered on a sucrose density and fractionated by centrifugation. Synaptic membranes were collected at the 1.0-1.2 M interface and harvested by centrifugation.

Digestion of synaptosome samples

30 mg of synaptosome was resuspended in 1 ml buffer containing 50 mM ammonium bicarbonate, 6 M guanidine hydrochloride 6x Phosphatase Inhibitor Cocktails I and II (Roche), and 20 μM PUGNAc (Tocris). The mixture was incubated for one hour at 57ºC with 2 mM Tris(2-carboxyethyl)phosphine hydrochloride to reduce cysteine side chains, these side chains were then alkylated with 4.2 mM iodoacetamide in the dark for 45 min at 21ºC. The mixture was diluted six fold with ammonium bicarbonate to a final ammonium bicarbonate concentration of 100 mM and 1:50 (w/w) modified trypsin (Promega, Madison, WI, USA) was added. The pH was adjusted to 8.0 and the mixture was digested for 12 hours at 37ºC. The digests were desalted using a C_{18} Sep Pak cartridge (Waters, Milford, MA, USA) and lyophilized to dryness using a SpeedVac concentrator (Thermo Electron, San Jose, CA, USA).

Preparation of the lectin weak affinity chromatography column

300 μg of POROS Al resin (Applied Biosystems) was reacted with 25 mg of WGA per the manufacturer's instructions. Briefly, 10 mM bicine, pH 7.5 was used as the

reaction buffer and 5 mg/ml sodium cyanoborohydride was added along with 200 μl 2M sodium sulfate. The mixture was rotated at 21°C for 24 hours. The resin was spun down and washed with 10 ml bicine, then quenched with 10 ml 200 mM Tris/acetate buffer, pH 7.5 and 200 μl sodium cyanoborohydride (100 mg/ml). The resin was then packed into a 2 x 250 mm stainless steel column.

Enrichment of GlcNAcylated peptides using a WGA column

Peptides were resuspended in 50 μl buffer A (100 mM Tris pH 7.5, 150 mM NaCl, 2 mM MgCl$_2$, 2 mM CaCl$_2$, 5% acetonitrile). Peptides were run over the column at 125 μl/min. GlcNAcylated peptides eluted as an unresolved smear on the right side of the flow through tail peak. After 1.3 ml, an additional 100 μl of 20 mM GlcNAc in buffer A was injected to elute any remaining peptides. A GlcNAc-enriched fraction was collected between approximately 1.3 and 3.7 ml. To decrease the chance of overloading the column each 10 mg portion was split into two 5 mg samples and run separately and the GlcNAc enriched fractions were combined subsequently. For subsequent rounds of enrichment, the pooled fractions were run together in a similar fashion as before. In subsequent rounds, the GlcNAc-enriched fraction was also collected from 1.3 ml to 3.7 ml.

Enrichment of phosphorylated peptides using titanium dioxide

Peptides were resuspended in 250 μl buffer B1 (1% TFA, 20% acetonitrile). The samples were run at 80 μl/min in buffer B1 over an analytical guard column with a 62 μl packing volume (Upchurch Scientific, Oak Harbor, WA USA) packed with 5 μm titanium dioxide beads (GL Sciences, Tokyo Japan)(22). The column was rinsed with H$_2$O, then eluted with 3 x 250 μl saturated KH$_2$PO$_4$ followed by 3 x 250 μl 5% phosphoric acid. A switching valve was used to direct these elutions onto a C$_{18}$ macrotrap peptide column (Michrom Bioresources, Auburn, CA, USA). The peptides were washed with H$_2$O then eluted with 50% acetonitrile, and this solution was lyophilized to dryness using a SpeedVac concentrator.

High pH Reverse Phase Chromatography

High pH RP chromatography was performed using an ÄKTA Purifier (GE Healthcare, Piscataway, NJ, USA) equipped with a 1 x 100 mm Gemini 3µ C18 column (Phenomenex, Torrance, CA). Individual GlcNAc-enriched or phospho-enriched fractions loaded onto the column in 1% buffer A (20 mM NH$_4$FA, pH 10). Buffer B consisted of buffer A with 50% acetonitrile. The gradient went from 1% B to 21% B over 1.1 ml, to 62% B over 5.4 ml, and then directly to 100% B. The flow rate was 80 µl/min. 20 fractions were collected and dried down using a SpeedVac concentrator. 1 mg of the GlcNAc- and phospho- depleted flow through material was separated by high pH reverse phase to collect 60 fractions.

Mass Spectrometry Analysis

All peptides were analyzed on an LTQ Orbitrap Velos equipped with a nano-Acquity UPLC. GlcNAc-enriched fractions were analyzed using electron transfer dissociation (ETD). Phospho-enriched fractions were analyzed using collision activated dissociation (CID). Non-modified peptides were analyzed using HCD. Peptides were eluted using a 90 minutes gradient. MSMS peaklists were extracted using the PAVA program (23). Data was searched against the UniProt Mus Musculus database (downloaded January 11, 2011; 72932 entries) using Protein Prospector (version 5.10.0). To this database, a randomized version was concatenated to allow determination of false discovery rates. The cleavage specificity was set to "trypsin", allowing for one missed cleavage. Carbamidomethylation of cysteine residues was set as a fixed modification. Acetylation of protein amino termini, oxidation of methionine residues, pyrolization of amino terminal glutamine residues, and loss of protein terminal methionine residues were set as variable modifications. For the main GlcNAc search, HexNAc modification of serine, threonine and asparagines was set as variable modifications. For the phosphorylation search, phosphorylation of serine, threonine and tyrosine was set as variable modifications. Data was searched initially with a 20 ppm tolerance of the parent ion, 0.6 Da tolerance of MS/MS measured in the ion trap (CID and ETD) and 20 ppm tolerance for HCD MS/MS. The precursor mass tolerance was then recalibrated on a file by file basis based upon the mass accuracy of high scoring peptides. Final precursor mass tolerances were between 10 and 13 ppm.

Subsequent searches were carried out to find those peptides simultaneously modified by both PTMs. The search parameters were as above, this time allowing for phosphorylation of serine and threonine residues and HexNAc modification of serine/threonine residues (ETD data), or HexNAc modification with neutral loss (CID data). Peptides simultaneously bearing both modifications were accepted if the corresponding peptide had been found with at least one of the PTMs and the peptide expectation value was below 0.01.

Searches were also conducted allowing the addition of 283.04 Da, corresponding to addition of protein glycosyl phosphorylation on a single residue (24). Peptides with both modifications, (or potential glycosyl phosphorylation) were accepted if the corresponding peptide had been found bearing at least one of the PTMs and the peptide expectation value was below 0.01. PTMs were considered positively localized to a single residue if they possessed a SLIP score greater than or equal to six (corresponding to a local false localization rate of less than 5%)(25).

For the resulting output, the corresponding UniGene name, gene, and entry numbers were appended (http://www.ncbi.nlm.nih.gov/unigene). UniProt entries were grouped by their corresponding UniGene genes and redundant peptides within a gene group were removed.

For the non-modified peptide identifications, a peptide expectation value threshold ≤ 0.01 was used. A protein was considered positively identified if the most confident peptide for that protein had an expectation value ≤ 1e-7. This resulted in the identification of 6621 UniGene entries and 60,421 unique peptides. At this threshold, the decoy database contained 6 entries and 8 unique peptides (protein FDR = 0.097%, peptide FDR = 0.013%).

O-GlcNAcylation and O-GalNAcylation both increase the mass of the modified peptide by the same amount (203.08 Da), and therefore these two PTMs are indistinguishable in the mass spectrometer. While O-GlcNAcylation occurs almost exclusively on intracellular protein regions, the extracelluar domain of Notch is O-GlcNAcylated (26). Peptides were assigned as ambiguous between O-GalNAcylated or O-GlcNAcylated based upon their annotation in UniProt (downloaded April 2011) as

located in extracelluar or luminal regions, without further corrections (27). In the case of transmembrane proteins, UniProt topological information was used when available to determine protein extracellular and cytosolic regions. Proteins and protein regions not annotated as extracellular or luminal were assumed to be cytosolic.

Calculations of expected *versus* observed frequencies.

The expected *versus* observed crosstalk between the two types of PTMs was determined in three different contexts. In all cases, the analysis was restricted to peptides where the site of modification was assigned with a false localization probability of less than five percent (25). (1): For crosstalk at a single residue, we restricted the entire analysis to the portion of each protein predicted to be in disordered regions in order to maximize the probability that all the serine and threonine residues we were considering were accessible for modification. Of the 135 observed alternatively modified residues, 96 were on disordered regions. For each protein, we then counted the number of times a residue was observed to be both O-GlcNAcylated and phosphorylated in different experiments. For that protein, we then calculated the number of residues expected to be alternatively modified by chance as $n * r_g * r_p$, where n represented the number of serine and threonine residues on that protein and r_g and r_p were the rates of O-GlcNAcylation and phosphorylation, respectively for the same protein (calculated as the number of each modification over the total number of serine and threonine residues). The total expected number of alternatively modified residues was determined by summing across all proteins and this was compared to the observed value using χ^2 evaluation. (2) For crosstalk within sequence proximity level, we compared the observed versus expected values for the number of times an O-GlcNAcylation event was observed at a distance of *n* residues from a phosphorylation, for different values of *n* along the protein sequence. Thus, for each O-GlcNAcylation, we counted the number of phosphorylations at distance *n* to create a distribution of observed distances. Expected distances were calculated as $n * r_p$ limiting the serine and threonine residues to those also at distance *n.* Values of *n* were binned in intervals of five to create a larger sample size. Expected values were compared to observed values at each bin interval using χ^2. (3) For crosstalk at the 3D proximity level, we compared

the expected and observed values for the number of times an O-GlcNAcylation was observed within n Å of a phosphorylation, for different values of n. Calculations proceeded as in (2). Analysis was limited to those modifications falling in a solved structure or good quality comparative model of the protein. In all alternative modification analysis, we limited the serine and threonine residues to those falling in disordered regions only.

Structural Analysis of PTMs

For proteins with an experimentally determined atomic structure or a 'good quality' (defined as a model being evaluated with a GA341 score of > 0.8 (28) comparative model in ModBase (29)), secondary structure assignments for each residue were computed by DSSP (30). For proteins without structural information, secondary structure was predicted from sequence using PSIPRED (31). For all proteins, disorder was predicted from sequence using the DISOPRED algorithm with the default five percent false positive threshold parameter. The median length of disordered regions was 28 residues, in line with expected values (32). To assess the distribution of PTMs with respect to protein kinase domains, ATP binding sites, and proton acceptor sites, the corresponding amino acid ranges and positions were obtained from Uniprot (uniprot.org). If this information was not annotated for a given entry (as was the case for some TrEMBL entries), the homologous Swiss-Prot entry was used.

Protein-Protein Docking

Known OGT substrate proteins with an experimentally determined atomic structure or comparative models in ModBase (29) with greater than 85% sequence identity (to limit to models likely to be of near-native quality) were docked to Chain A of the open OGT structure (PDB ID 3PE4) using PatchDock (http://bioinfo3d.cs.tau.ac.il/PatchDock/) (33). 26 O-GlcNAcylation sites from 20 substrate proteins (B1ATI9-T34, O88935-T262, P01942-S53, P05064-S354, P12787-S100, P14094-S160, P18760-S8, P60710-S199;-S365, P62874-S136, P68134-S241, P70365-T642, P97315-S192, Q3UHK5-S45;-S559;-S650, Q80X68-T454, Q8K4S1-S2139, Q8VC88-S159, Q9CWF2-T285, Q9CZ13-T217, Q9D1Q6-T367;-S380;-S381, Q9JKV1-S211;-S213) were docked to a truncated version, residues 467-1031, of Chain A of the open OGT structure, which excludes the

TPR domain. Two distance constraints were applied to each of the substrate proteins. Atom 1462 in the 3PE4 structure, the nitrogen atom on the catalytic H498 of OGT, was required to be within 13.5 Å of the oxygen atom on the serine or threonine to be O-GlcNAcylated. Atom 11140 in the 3PE4 structure, an oxygen atom on the α-phosphate of UDP, was required to be within 13.0 Å of the nitrogen atom on the serine or threonine residue to be O-GlcNAcylated. A clustering RMSD of 4.0 Å and the default complex type were used. Using the link provided on the PatchDock results page, the best 100 solutions were further automatically refined using FireDock (34). The following substrates had O-GlcNAcylation sites that were not successfully docked: A2AA49; O08599; P08249; P14094; P48962; P56480; P60710; P62874; P68134; P68369; Q03265; Q80X68; Q91VR2; Q9CWF2; Q9CWS0; Q9JKV1)

Results

Abundance of O-GlcNAcylation and phosphorylation

We developed a workflow to sequentially enrich O-GlcNAcylated and phosphorylated peptides from tryptic digests of mouse synaptosomes, which also allowed for analysis of the protein content from the PTM-depleted sample (Figure 1A). O-GlcNAcylated peptides were isolated using three rounds of lectin weak affinity chromatography (LWAC) (Figure 1B), yielding a final pool containing approximately 30% O-GlcNAcylated peptides. Phosphorylated peptides were isolated using an automated TiO_2-based enrichment step (Figure S1). Then these two PTM-enriched fractions as well as the final unbound fraction (containing non-modified peptides) were fractionated using high pH reverse phase chromatography (Figure 1C). All fractions were analyzed on an LTQ-Orbitrap Velos mass spectrometer using electron transfer dissociation (ETD) for O-GlcNAc peptides and collisional dissociation (CID or HCD) for phospho- and other peptides. Interpretation of these mass spectral analyses resulted in the identification of 2,434 distinct O-GlcNAcylated- and 23,206 phosphorylated peptides, respectively. These assignments correspond to over 1,750 unique sites of O-GlcNAcylation and 16,500 unique sites of phosphorylation. Additional analysis of the PTM-depleted peptide fraction identified 60,422 peptides from 6,621 proteins, all at global false discovery rates (FDRs) of less than 1% (Tables S1-3). As we have

previously reported, our enrichment technique using the lectin wheat germ agglutinin (WGA) also enriches for N-GlcNAcylated peptides (15). In our current analysis, we found 546 N-HexNAcylated peptides (Table S4). While WGA has been reported to be selective for GlcNAcylated peptides and proteins (35), we have identified 181 peptides in the WGA-enriched fractions that are potentially O-GalNAcylated based upon sub-cellular localization of the modification (Table S5).

Efficiencies of PTM-detection

A major factor affecting whether or not a given peptide is detected in a proteomic study is its relative abundance (36). To estimate how efficiently we identified sites of O-GlcNAcylation and phosphorylation within the synaptosome preparation, we took advantage of our concurrent in-depth analysis of proteins from the same sample. The 6,621 identified proteins were divided into bins based upon their relative abundance as determined by calculating an exponentially modified protein abundance index (emPAI) value for each protein (37). We then calculated the percentage of proteins in each bin that were either O-GlcNAcylated or phosphorylated. For the most abundant proteins, we identified 19% and 63% of them to be O-GlcNAcylated and phosphorylated, respectively (Figure 2A-2B). Proteins present at lower abundance were substantially less likely to be identified as O-GlcNAcylated (an average of 9.8% for the 12 lowest bins). For phosphorylation, this decrease was more modest. For 52% of the proteins in the 12 lowest bins, at least one site of phosphorylation was identified. Proteins in the most abundant bin had an average of 0.51 and 5.9 sites of O-GlcNAcylation and phosphorylation, respectively (Figures 2C-2D). The average number of sites identified per protein dropped off significantly with decreased protein abundance for both PTMs.

Overall, this finding suggests that while we were able to identify large numbers of both PTMs, we are not identifying all PTM-modified peptides present in the sample, particularly not those originating from lower abundance proteins. This is most likely due to the fact that when a given analysis does not identify all the components in a mixture, there is a strong bias towards acquiring MS/MS on higher abundance components. Based upon the average modifications per protein for the most abundant/thoroughly characterized proteins, we now can postulate the existence of at least 3,400 O-

GlcNAcylation sites and 39,000 phosphorylation sites for the 6,621 proteins identified in our synaptosome preparation. Using the same rationale, we estimated that we identified approximately 51% and 42% of the O-GlcNAcylation and phosphorylation sites in our sample, respectively.

Mass spectrometric characterization of PTM-modified peptides

Multiple sites of O-GlcNAcylation were found close in protein primary sequence. We observed 130 instances where the same peptide sequence was found singly O-GlcNAcylated at different residues. Figures 3A and 3B show MS/MS spectra of two O-GlcNAc positional site isomers of the peptide sequence, TAVKPTPIILTDQGMDLTSLAVEAR, from the protein bassoon. We observed 439 instances of peptides containing multiple sites of O-GlcNAcylation. An example of such a peptide, SVTDTALPGQSSGPFYSPR, modified at serine 1 and threonine 3, is shown in Figure S2A. The presence of a relatively bulky O-GlcNAc moiety carboxy terminal to an arginine or lysine residue may decrease tryptic cleavage efficiency, but these data indicated such cleavage events proximal to an O-GlcNAcylation site are not strictly prevented.

Overall, we observed 135 peptide pairs where one version was O-GlcNAcylated and the other phosphorylated at the same residue. Figures 3C and 3D show MS/MS spectra of two alternatively modified versions of the peptide AAVVTSPPPTTAPHK from the protein α-adducin, either phosphorylated or O-GlcNAcylated at serine 6. In addition, we found 66 peptide sequences that were simultaneously modified by both phosphorylation and O-GlcNAcylation (Table S6). One such example is the peptide RASQp(SS)LESSTGPSYgSR, shown in Figure S2C. Figure S2B shows an example of an N-GlcNAcylated peptide with the sequence LNGTDPIVAADSKR from the Prolow-density lipoprotein receptor-related protein 1, modified at aspargine 2.

PTM sequence motifs

Previous analyses, based on a significantly smaller scale O-GlcNAcylated peptide dataset, suggested a PVXS/T motif for substrates of OGT (18). While this motif is found in a subset of modified peptides in this study, the majority of our O-GlcNAcylation sites fit this motif poorly. In fact, less than 20% of our modified peptides

display this motif. Figure 3E shows a weblogo representation of the amino acid residues surrounding the modified serine/threonine residue. There is a moderate preference for a proline residue either two or three residue positions amino-terminal to the site of modification (the -2 and -3 positions). There is also a slight preference for valine residues at either -1 or -3 positions. Overall, O-GlcNAc appears to be targeted towards regions on substrates rich in serine/threonine residues, as evidenced by an increased frequency our detection of these residues within the five residue region around the modification site that we examined. Such a preference for serine/threonine rich stretches may help explain our finding 439 multiply O-GlcNAc-modified peptides. In summary, these observations suggest that OGT does not recognize the exact primary sequence on substrates.

To investigate motifs within our phosphorylation dataset, we used Motif-X to search for over-represented patterns (38). We find that a total of 56 distinct sub-motifs show statistically significant overrepresentation (Table S7). To examine the motifs, we grouped amino acid residue types by chemical property (e.g. small hydrophobic, charged/polar side chains) as shown in Figure 3F-3I. When grouped by chemical property, the most prevalent amino acid residue types present around the site of O-GlcNAcylation are small/non-polar residues, indicating existence of a hydrophobic residue at the -3 position. Overall, phosphorylation sites in our dataset have a similar preference for small/non-polar residues. In addition, due to the prevalence of proline-directed kinases in the mammalian kinome, there was an increased probability of having a hydrophobic residue at the +1 position. Finally, we examined those serine/threonine residues showing co- modification by both PTMs. This subset had a motif most similar to that of the overall O-GlcNAcylation motif. We compared these motifs to the population of serine/threonine residues not found to be PTM-modified. Hydrophobic residues are most prevalent at all positions immediately surrounding these serine/threonine residues (Figure 3I).

O-GlcNAcylation and phosphorylation on PTM-modifying enzymes

We identified one site of O-GlcNAcylation on OGT, on the carboxy-terminal tail. However we did not identify any phosphorylation on OGT despite the protein being

present at relatively abundant levels (28 unique peptides identified) in our synaptosome preparation. Six unique phosphorylation sites on OGT from mitotically active cells have been previously identified (39). On the O-GlcNAcase, we identified two sites of phosphorylation and no sites of O-GlcNAcylation. We identified 280 and 87 proteins in Gene Ontology annotated with the protein kinase (GO:0004672) and phosphatase activities (GO:0004721), respectively. Kinases were more frequently phosphorylated than proteins in the full dataset (66% versus 48%, $p < 3.8 \times 10^{-11}$, hypergeometric distribution). Kinases were also more frequently O-GlcNAcylated than proteins in the full dataset (16% versus 10%, $p < 3.6 \times 10^{-4}$, hypergeometric distribution). Protein phosphatases however, were not found to be PTM-modified at rates different from the overall dataset (52% phosphorylated and 8% O-GlcNAcylated). This evidence supports the notion that O-GlcNAcylation crosstalks with phosphorylation via OGT's regulation of a substantial subset of kinases.

To examine in more detail the manner in which OGT activity may modify kinase behavior, we looked at the localization of O-GlcNAcylation sites on the 46 modified kinases with existing structural information in UniProt. In particular, we asked how closely sites of O-GlcNAcylation mapped to the proton acceptor and ATP binding sites as well as the protein kinase catalytic domain as a whole. Of the 131 O-GlcNAcylated peptides mapping to these 46 proteins, in only two instances was a site of O-GlcNAcylation located within the approximately 250 residue kinase catalytic domain (1.5%). In contrast, 134 of the 480 phosphorylated peptides (28%) mapped to the protein kinase domain of kinases. Therefore, while protein kinase domains themselves are subject to extensive modification by phosphorylation, they are minimally O-GlcNAcylated.

With respect to the two kinase domains that were modified, in the first instance (CaMKII α) the O-GlcNAcylation site was more than 100 residues from either the proton or ATP acceptor site. In the second instance (CaMKII β), the O-GlcNAcylation site was 41 residues from the proton acceptor site. While it is possible that O-GlcNAc may be modifying kinase regulatory domains, it does not appear to be directly affecting kinase catalytic ability or the ability of the catalytic domain to interact with substrate proteins. For several CaMKII isoforms, we observed O-GlcNAcylation of threonine 306/307 (in α

and δ, respectively). Upon auto-phosphorylation, this residue inhibits calmodulin binding to CaMKII (40). The phosphorylation state of the adjacent threonine 305 and threonine 306 residues in CaMKII α modulates LTP (41). We detected 11 sites of phosphorylation on CaMKII α, but did not identify phosphopeptides corresponding to threonine 306 or 307. The PTM's of CaMKII α are mapped on its homology-based structural model in Figure S3.

Modeling OGT interactions with its substrates

A recent solved crystallographic structure of human OGT in complex with an O-GlcNAcylated peptide and a tetratricopeptide repeat (TPR) domain gave insight into possible factors mediating interaction specificity with the enzyme and its substrates (42). First, the contacts between the enzyme and peptide were made primarily through the peptide main-chain atoms, indicating that peptide side-chain identity does not play a strong role in substrate recognition (Figure 3E). The possibility exists that this specificity is determined through enzyme-substrate interactions distal to the active site. Second, the TPR domain appears to restrict substrate access to the active site, leading to speculation that the TPR domain swings out in a hinge motion prior to OGT-substrate complex formation. We examined the catalytic face of OGT with the TPR domain removed and observed a large basic patch that encompasses the catalytic site of OGT (Figure 4A). A complementary acidic patch is present on the proximal TRP domain that interacts with this basic patch (Figure 4B). The existence of these complementary electrostatic regions was not reported in the original crystal structure manuscript. To determine possible exosite contacts, and examine the possible role of the TPR domain, we applied computational docking methods to model the conformations of 46 O-GlcNAcylated sites identified in our study (from 29 O-GlcNAcylated proteins); the proteins we selected all had solved structures or high quality comparative models (Experimental Procedures). Two docking runs were performed; the first incorporated the TPR domain as observed in the solved structure of the OGT complex, and the second did not.

The docking protocol attempted to model the native conformation of a complex, assuming that complex formation does not significantly alter the three dimensional

structure of either component. Only one modified protein (cysteine and glycine-rich protein 1, P97315) was able to dock with OGT when the TPR domain was included. However, upon removal of the TPR domain, 26 O-GlcNAcylated sites docked successfully. This observation suggests that it is indeed necessary for the TPR domain to 'swing out' prior to complex formation. Furthermore, the inability of the remaining 20 substrates to dock indicates that the bona fide enzyme:substrate complex in these instances requires conformational changes of the substrate and/or OGT for O-GlcNAcylation to occur.

Using the docked conformations of these 26 substrates, we searched for interactions between OGT side-chains and residues structurally conserved across multiple substrates. While we observed no fully conserved interactions, we identified four charged patches conserved across subsets of substrates that interact with oppositely charged residues on the enzyme (Figure 4C-G).

Secondary structural elements involved in modification of substrate proteins

To gain insight into what secondary structure context may be needed for O-GlcNAcylation and phosphorylation, we determined the frequency with which they appeared on loops, α-helices, and β-sheets. The secondary structure states of all residues in a solved structure or a good quality comparative model were assigned by the Define Secondary Structure of Proteins (DSSP) program (30). For substrate proteins without known or modeled structures, the secondary structure states were predicted by the PSIPRED program (31). Relative to the distribution of these structural elements in general, both O-GlcNAcylation and phosphorylation moieties were enriched within loops and relatively less prevalent within sheets and helices (Figure 5A). For both PTMs, the site of modification occurred on loops for approximately 90% of the sample. We then calculated to what extent the two PTMs were found in ordered versus disordered regions of protein structure. Both phosphorylation and O-GlcNAcylation were approximately six-fold more likely to occur on disordered rather than ordered regions of protein structure (Figure 5A), in agreement with a previous observation for phosphorylation (43).

O-GlcNAcylation and phosphorylation on individual proteins

To investigate how these two PTMs might be coupled at the level of individual proteins, we examined the number of phosphorylation sites per protein as a function of the number of O-GlcNAcylation sites per protein (Figure 5B). Phosphorylated proteins were significantly more likely to be also O-GlcNAcylated than non-phosphorylated proteins (25.6% versus 4.8%, $p < 1.2 \times 10^{-18}$, hypergeometric distribution). There is a weak correlation between the frequencies of these two PTMs ($r^2 = 0.25$). Interestingly, the vast majority of proteins partitioned to the top left half (i.e. above the diagonal with a phospho:O-GlcNAc ratio of one). The number of O-GlcNAc sites per protein was approximately equal to the minimum number of phosphorylation sites per protein, particularly when the number of O-GlcNAc sites was greater than two. However, for many proteins we observed extensive phosphorylation and only a limited number of O-GlcNAcylation sites. In contrast, in only a single instance did we observe a heavily O-GlcNAc-modified protein that was not also heavily phosphorylated.

The only protein that appeared heavily O-GlcNAcylated (13 sites) with no observed phosphorylation was CCR4-NOT transcription complex subunit 1. CCR4-NOT was amongst the 20% most abundant proteins in our preparation as estimated by the emPAI value. As a transcription factor, this protein is likely to partition between the nucleus and cytoplasm (for a review see (44)). Regulation of gene transcription is a protein functional class known to be preferentially O-GlcNAcylated (45). Only one site of phosphorylation on CCR4-NOT has been reported (46). Since only a minor fraction of this protein was likely present in our synaptosome preparation, it is possible that analysis of a total cell lysate (rather than of a specific compartment) would reveal additional sites of CCR4-NOT phosphorylation.

Substrates of OGT in general appear to consist of a subset of kinase targets. The fact that the extent of potential protein O-GlcNAcylation tracks with the extent of potential protein phosphorylation is consistent with OGT and serine/threonine kinases using similar mechanisms to target substrates. For example, a portion of OGT has been reported to occur in complex with catalytic subunits of protein phosphatase 1 (47).

Sites that are modified by either PTM

As noted above, in 135 instances, we observed phosphorylation and O-GlcNAcylation of the same residue, representing 8% of the identified O-GlcNAcylation sites found in this study. This number of alternatively modified sites suggests coupling between these two PTM systems; however, given the extensive number of phosphorylation and O-GlcNAcylation sites we identified in this study, it is expected that both PTMs would map to the same amino acid residue at some frequency by chance alone. If these two PTM systems have evolved to crosstalk functionally, the observed frequency with which the same residue was found modified by both PTMs should substantially exceed the frequency predicted by chance alone. We assumed that all serine and threonine residues on disordered regions were accessible to enzymatic modification (which was not necessarily the case for serine and threonine residues in ordered protein regions). We therefore limited our analysis to disordered regions. Approximately 50% of all serine and threonine residues are in disordered regions, and 96 of our 135 observed alternatively modified sites mapped on disordered regions (Figure 5A). For a given protein, the number of alternatively modified sites expected by chance alone is modeled as the rates of phosphorylation and O-GlcNAcylation on disordered regions for that protein multiplied by the total number of disordered serine and threonine residues. Summing the expected alternatively modified residues across all proteins in our dataset resulted in 96.4 sites of alternative modification expected by chance alone. Therefore, while both PTMs are preferentially targeted to disordered regions of protein structure, within these disordered regions, we find no increased propensity for alternative modification.

Relationship between the two PTMs with respect to primary structure

Spatial proximity between sites of O-GlcNAcylation and phosphorylation has been posited as a mechanism for crosstalk (6), whereby addition of one PTM will impact on the odds of the other *via* electrostatic or steric factors. If an organism has evolved to utilize such a mechanism, sites of O-GlcNAcylation should display an increased propensity to be localized in spatial proximity to sites of phosphorylation. We tested this hypothesis using two different representations of spatial proximity. The first computes proximity as the number of co-occurrences within a given number of residues along the protein sequence ("sequence proximity") and the second computes it as the number of

co-occurrences within a given three-dimensional distance in solved substrate structures ("3D proximity"; next section). For each, we compared the number of observed co-occurrences within different distances to the number expected by chance alone.

To calculate the expected number of co-occurrences in sequence proximity for a given protein and residue distance cutoff, we multiplied the phosphorylation rate for that protein by the number of serine and threonine residues within the distance cutoff of each observed O-GlcNAcylation, and summed the results for all observed O-GlcNAcylations in that protein. We found that there was essentially no increase in the number of observed alternative modifications at close distances (*i.e.*, fewer than ten residues) in sequence proximity relative to that expected by chance alone (Figure 5E). We compared this distribution to the propensity of phosphorylation sites to cluster within a protein. We calculated the expected *versus* observed number of phosphate-phosphate co-occurrences at different distances in sequence proximity using the same approach as for phosphorylation – O-GlcNAcylation and detected a significant enrichment (Figure 5C) similar to that reported previously (48) (49) (50). We then examined the same enrichment of observed *versus* expected co-occurrences of pairs of O-GlcNAcylation modifications (Figure 5D) and detected a similar propensity to cluster. While there is a slight increase in the number of observed phosphorylation – O-GlcNAcylation co-occurrences relative to that expected by chance, it is significantly less than the robust increase in co-occurrences found within each type, indicating there is not strong evolutionary pressure for crosstalk through sequence proximity between O-GlcNAcylation and phosphorylation. In addition, our observation of 66 peptide sequences simultaneously bearing both PTMs (Table S6) demonstrates that electrostatic/steric factors do not strictly prohibit simultaneous alternative modification by both PTMs nearby in primary sequence distance.

Relationship between the two PTMs with respect to tertiary structure

While sequence proximity between PTM pairs is a proxy for the underlying 3D distance, we could not calculate this value for all pairs since most of our OGT substrates lack solved angstrom level structures. Nevertheless, of the 285 proteins with both types of PTMs, 52 had such a structure or comparative model available covering

both sites of modification. For these, we counted the number of observed phosphorylated residues within a given 3D distance of each O-GlcNAcylated residue, using shells of increasing radii. We compared this observed value to that expected by chance, using the phosphorylation rates and total number of serine and threonine residues at the same distance, as was done in the sequence proximity analysis. We found that there was no enrichment for co-occurrences relative to that expected by chance (Figure 5G). We compared this result to the enrichment for pairs of O-GlcNAc modifications in 3D proximity. Here, we found that observed pairs of these modifications are more likely to co-occur than expected by chance at less than 10Å (Figure 5F). This result supports the one determined from the sequence proximity analysis, indicating that there is little to no evolutionary pressure for crosstalk of these two types of PTMs within either type of proximity.

Discussion

O-GlcNAcylation is a widespread post-translational modification

Previous investigations of protein O-GlcNAcylation have been limited in scope and in particular have lacked analogous characterization of phosphorylation for modified proteins occurring in the same biological preparations. Our identification of 1,750 O-GlcNAcylation sites provides a 20-fold increase in the number of sites previously identified from any sample with endogenous levels of O-GlcNAcylation.

Our extensive O-GlcNAcylation coverage of both proteins modified and sites occupied, coupled with identification of over 16,500 phosphorylation sites, allowed us to systematically characterize O-GlcNAc distribution on synaptic proteins, model the interaction of target sites with OGT, and address potential crosstalk between these two post-translational modifications. Our increased coverage was mainly due to three factors: (a) the use of more sensitive mass spectrometry (an Orbitrap Velos equipped with ETD fragmentation), (b) high *p*H fractionation of the O-GlcNAc-enriched fractions prior to LC-MS/MS, and (c) improved efficiency of the lectin-enrichment step. The primary improvements in the lectin-enrichment step were the switch from an agarose-immobilized lectin to one immobilized on POROS resin (51) carried out in three rounds of enrichment at 4°C.

Comparison with chemoenzymatic strategies for O-GlcNAc analysis

An alternative approach for the enrichment of O-GlcNAc-modified peptides involves the chemoenzymatic addition of biotinylated GalNAc to O-GlcNAc using an engineered version of galactosyl transferase (20). This approach was used previously in combination with ETD to identify 141 sites of O-GlcNAcylation from mitotic spindle and midbodies (10). A paper was published while this current manuscript was in review extending the chemoenzymatic strategy to identify 458 sites (52). While our study identified several-fold more O-GlcNAc peptides, it is hard to compare the selectivity and sensitivity of the two approaches, because the biological samples studied and amounts employed were different and because the previous experiment was carried out in a cell line over-expressing OGT, thus elevating the level of modification above endogenous levels. A possible shortcoming of the chemoenzymatic approach is less than 100%

efficiency of the enzymatic incorporation of the tag and/or photocleavage from the resin. Despite the high coverage of O-GlcNAcylation at the spindle, only 3 of the 138 (2.1%) modified peptides reported previously (10) were doubly O-GlcNAcylated. In contrast, 439 of the 2,434 peptides (18%) in the current study were found to be multiply O-GlcNAcylated. The chemoenzymatic approach may bias against multiply modified peptides (e.g. by limited enzyme efficiencies). In contrast, our chromatographic approach may bias in favor of multiply modified peptides because such peptides may be more efficiently separated from the flow-through peak during lectin-enrichment.

Physiological Role of O-GlcNAcylation in the brain

O-GlcNAcylation plays a critical role in neuronal biology. Neuron-specific knock-out of OGT leads to early postnatal death, which suggests a role for this enzyme in essential pathways (5). Both OGT and OGA are enriched at synapses (2) (3). In addition, O-GlcNAc has been implicated in a diverse set of neuronal processes, such as axonal branching and LTP at CA3/CA1 hippocampal synapses (53) (54).

We examined potential biological functions of O-GlcNAc using gene ontology analysis (http://amigo.geneontology.org). For this analysis, we used a background consisting of proteins in our dataset not found to be O-GlcNAcylated and of a similar abundance distribution to the O-GlcNAcylated proteins. Consistent with O-GlcNAc modifications occurring on a large percentage of proteins, there were no GO categories in which O-GlcNAcylated proteins were significantly (greater than 50%) enriched. This finding suggests that modification by O-GlcNAc regulates a wide range of biological processes in synaptic regions of the brain.

The protein bassoon is extensively modified by both O-GlcNAcylation and phosphorylation (185 and 117 sites, respectively). Bassoon is a core presynaptic active zone protein, and participates in targeting of cargo to distal axons as a component of piccolo-bassoon transport vesicles. The binding of bassoon to dynein light chain is thought to regulate transport of these vesicles along microtubules (55). Bassoon contains three functional dynein light chain-binding motifs. We identified O-GlcNAcylation sites within two of these motifs, while none of them were found to be phosphorylated. O-GlcNAcylation of these motifs may disrupt interactions between

dynein and bassoon, indicating a potential role for O-GlcNAcylation in regulation of vesicular transport.

A fraction of CaMKII α is O-GlcNAcylated at threonine residue 306. Auto-phosphorylation at both threonine residue 306 and the adjacent threonine residue 305 reduced the sensitivity of CaMKII to stimulation by CaM (inhibitory auto-phosphorylation). Genetic mutations of these residues affected spatial learning and synaptic plasticity *in vivo*. Inhibitory auto-phosphorylation of CaMKII controls PSD association, plasticity, and learning (56). Our observation of O-GlcNAcylation at threonine residue 306, suggests that mutation of this threonine to alanine (phospho-null mutation) needs more careful interpretation since this mutation eliminates the possibility of either O-GlcNAcylation or phosphorylation. The strong phenotype shown by mice carrying the phospho-mimicking allele (T306D) confirmed the importance of phosphorylation at this residue, but leaves the role of O-GlcNAcylation unclear. Our failure to find phosphopeptides corresponding to threonine residues 305/306 despite identifying over 16,500 phosphorylation sites suggests that the phosphorylation stoichiometry of these residues is low. Proteomic analyses (such as ours) typically rely on data-dependent acquisition of MS/MS spectra, which is biased towards the identification of abundant peptides. We cannot therefore rule out the possibility that phosphorylation at threonine residues 305/306 was present at some low level and that we failed to identify these sites due to chance. This example highlights the future need to determine the absolute occupancy of alternative site-specific PTMs as an important step towards understanding their respective functional roles for all sites that are affected by both PTMs.

The microtubule associated protein tau is extensively phosphorylated (for a review on tau PTMs, see (57) and has been reported to be modified by O-GlcNAc (58), giving rise to the hypothesis that O-GlcNAc may be regulating levels of tau phosphorylation (59). We identified 112 unique tau peptides, including 67 phosphopeptides mapping to 46 sites of phosphorylation. However, we did not find any peptides corresponding to O-GlcNAcylation on tau. This finding calls into question whether endogenous synaptic tau in healthy young animals is sufficiently O-GlcNAcylated to significantly modulate its phosphorylation site occupancies.

O-GlcNAcylation sites are often found in local clusters

Our dataset allowed for the first time to address whether or not O-GlcNAcylation sites are randomly distributed or occur in local clusters on the primary protein sequence or within 3D distance. Our data strongly support the presence of local clusters of O-GlcNAcylation sites on the primary sequence with a distribution that is akin to but distinct from the one described for serine/threonine phosphorylation sites (48) (49) (50) While members of clusters must originate from the catalysis of the single OGT, there remains a remote possibility that different OGT co-proteins are involved in the O-GlcNAcylation of the adjacent sites; more likely explanations include a tendency of OGT to progressively modify a substrate in a concerted manner (perhaps via a tendency of OGT itself to directly bind O-GlcNAcylated substrates), and a promiscuous transfer of the high energy GlcNAc group to nearby nucleophilic serine and threonine residues on the docked substrate (60). Our lectin-based enrichment scheme may lead to an overrepresentation of multiply O-GlcNAcylated peptides, which in turn could lead to an overestimation of the extent to which O-GlcNAcylated sites are clustered. However, repeating the analysis in Figures 5D/F using only singly O-GlcNAcylated peptides resulted in a consistent observation of clustering with respect to both primary sequence distance and 3D spatial proximity. It appears that Ser/Thr kinases and OGT share a similar tendency to cluster their respective modifications. As a result, future investigations on the functional role of individual O-GlcNAcylation sites will need to account for the fact that O-GlcNAcylation may occur locally in a cluster and that O-GlcNAcylation at a distinct site may not be necessary and sufficient to cause a biological effect, nor may a O-GlcNAcylation-null mutation at a single site within a cluster be sufficient to test for a phenotype.

PTMs can potentially crosstalk at multiple levels

Crosstalk between the two types of PTMs can occur *via* three distinct yet potentially simultaneous mechanisms. A: Crosstalk at the level of substrate where addition of one PTM directly regulates the ability of an enzyme to add the second PTM (structural crosstalk). B: Crosstalk at the enzymatic level where one type of PTM modifies enzymes responsible for the addition/removal of the second PTM, thus

regulating their activity (catalytic crosstalk). C: Crosstalk where addition of one PTM changes the sub-cellular localization of the substrate and thereby regulates the ability of substrate to be modified by the second PTM (localization crosstalk)(61).

In this study, we identified 135 instances of individual serine and threonine residues alternatively modified by both O-GlcNAcylation and phosphorylation. Previously, fewer than 20 such instances have been reported in the literature (6). The scope of our analysis allowed us to demonstrate that the frequency of alternatively modified sites was essentially equal to the frequency expected by chance alone. Thus, our data suggests that there is no common evolutionary pressure to increase local structural crosstalk. Nevertheless, for these 135 instances addition of O-GlcNAc will compete with addition of phosphate, and vice versa. Biological relevance of this competition requires that the levels of modification be sufficiently high, in order to significantly alter the concentration of unmodified protein. While we did not measure absolute modification levels, recent reports have examined these values for both PTMs on a range of proteins (62) (63). An examination of O-GlcNAcylation occupancy at the protein level for seven proteins showed a range of 2 to 100%, although the occupancies at individual sites for multiply-modified proteins will likely be lower. Wu and colleagues calculated phosphorylation occupancies for over 5000 yeast phosphorylation sites. These values varied from 1 to 100%, with a median phosphorylation occupancy value of approximately 25%. Based upon these results, it would appear that basal occupancies for both PTMs are in a range where moderate increases in one PTM may or may not result in a significant decrease in the substrate availability of the other PTM.

In addition to demonstrating that kinases and OGT do not preferentially target the same residues, we also show that the distribution of these two post-translational modifications do not co-cluster on proteins. Therefore, it does not appear to be a general principle that addition of one type of PTM on a protein inhibits the addition of the other (via electrostatic or steric effects). Nevertheless, our results leave open the possibility of positive interactions between these two PTMs, whereby addition of one PTM increases the rate of addition of the other. For example, addition of phosphorylation may create a docking site used by OGT to GlcNAcylate a

phosphorylated protein. Such a mechanism may explain our observation in Figure 5B that the extent to which a protein may be GlcNAcylated is a function of the extent to which that protein is phosphorylated. Regulation of this sort has recently been proposed for GlcNAcylation of phosphorylated CREB (64). Importantly, if substrate phosphorylation creates a docking site (or conformational change) enabling recognition by an OGT complex, the resulting site of GlcNAcylation may not be in close proximity to the phosphorylation site. Our analysis relied on identification of peptides after tryptic cleavage. While we identified 66 peptides simultaneously modified by both PTMs, digestion of the proteins likely resulted in many cases were nearby O-GlcNAc and phosphorylation sites ended up on different tryptic peptides. As such, our current dataset does not readily allow us to determine the extent to which one type of PTM might potentially act as a primer for a second modification.

It has recently been reported that O-GlcNAc and phosphorylation levels are of similar abundance at spindles and midbodies (10). However, without controlling for differential detection efficiency of the two PTMs, it is difficult to make such claims with a high degree of confidence. When we attempted to account for this effect (Figure 2C-D), we observed 11-fold more sites of phosphorylation than O-GlcNAcylation in synaptosomes. While different subcellular compartments will undoubtedly have different ratios of these two PTMs at subsets of proteins, our results encompass measurements for over 6,000 proteins.

In summary, our large-scale analysis of O-GlcNAcylation and phosphorylation revealed three key findings. First, O-GlcNAcylation occurs infrequently on proteins that are not phosphorylated, indicating that targeting of OGT does not extend beyond the set of substrates that has evolved for kinases. Second, similar to phosphorylation by serine/threonine kinases, O-GlcNAcylation also occurs in a clustered pattern, likely indicating similar evolutionary selection of sites or enzymatic mechanism for these two PTMs and their transferring enzymes. Third, while both O-GlcNAcylation and phosphorylation occur in clusters, these clusters do not co-localize. The localization of O-GlcNAcylation sites is statistically independent from localization of phosphorylation. This is consistent OGT and kinases having independently evolved substrate site

specificity. It also indicates that at the proteome level, there was no evolutionary advantage to promote local spatial crosstalk between these two PTMs.

References

1. Hanover, J. A., Krause, M. W., and Love, D. C. (2010) *Biochim. Biophys. Acta* **1800**, 80–95
2. Cole, R. N., and Hart, G. W. (2001) *J. Neurochem.* **79**, 1080–1089
3. Gao, Y., Wells, L., Comer, F. I., Parker, G. J., and Hart, G. W. (2001) *J. Biol. Chem.* **276**, 9838–9845
4. Akimoto, Y., Comer, F. I., Cole, R. N., Kudo, A., Kawakami, H., Hirano, H., and Hart, G. W. (2003) *Brain Res.* **966**, 194–205
5. O'Donnell, N., Zachara, N. E., Hart, G. W., and Marth, J. D. (2004) *Mol. Cell. Biol.* **24**, 1680–1690
6. Hart, G. W., Slawson, C., Ramirez-Correa, G., and Lagerlof, O. (2011) *Annu. Rev. Biochem.* **80**, 825–858
7. Wang, Z., Gucek, M., and Hart, G. W. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13793–13798
8. Griffith, L. S., and Schmitz, B. (1999) *Eur. J. Biochem.* **262**, 824–831
9. Wang, Z., Pandey, A., and Hart, G. W. (2007) *Mol. Cell Proteomics* **6**, 1365–1379
10. Wang, Z., Udeshi, N. D., Slawson, C., Compton, P. D., Sakabe, K., Cheung, W. D., Shabanowitz, J., Hunt, D. F., and Hart, G. W. (2010) *Sci. Signal.* **3**, ra2
11. Choudhary, C., and Mann, M. (2010) *Nat. Rev. Mol. Cell Biol* **11**, 427–439
12. Thingholm, T. E., Jensen, O. N., and Larsen, M. R. (2009) *Proteomics* **9**, 1451–1468
13. Chalkley, R. J., and Burlingame, A. L. (2001) *J. Am. Soc. Mass Spectrom.* **12**, 1106–1113
14. Chalkley, R. J., and Burlingame, A. L. (2003) *Mol. Cell Proteomics* **2**, 182–190
15. Chalkley, R. J., Thalhammer, A., Schoepfer, R., and Burlingame, A. L. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8894–8899
16. Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J., and Hart, G. W. (2002) *Mol. Cell Proteomics* **1**, 791–804
17. Vosseller, K., Hansen, K. C., Chalkley, R. J., Trinidad, J. C., Wells, L., Hart, G. W., and Burlingame, A. L. (2005) *Proteomics* **5**, 388–398
18. Vosseller, K., Trinidad, J. C., Chalkley, R. J., Specht, C. G., Thalhammer, A., Lynn, A. J., Snedecor, J. O., Guan, S., Medzihradszky, K. F., Maltby, D. A., Schoepfer, R., and Burlingame, A. L. (2006) *Mol. Cell Proteomics* **5**, 923–934
19. Myers, S. A., Panning, B., and Burlingame, A. L. (2011) *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9490–9495
20. Khidekel, N., Ficarro, S. B., Clark, P. M., Bryan, M. C., Swaney, D. L., Rexach, J. E., Sun, Y. E., Coon, J. J., Peters, E. C., and Hsieh-Wilson, L. C. (2007) *Nat. Chem. Biol* **3**, 339–348
21. Trinidad, J. C., Specht, C. G., Thalhammer, A., Schoepfer, R., and Burlingame, A. L. (2006) *Mol. Cell Proteomics* **5**, 914–922
22. Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. R. (2004) *Anal. Chem.* **76**, 3935–3943
23. Guan, S., Price, J. C., Prusiner, S. B., Ghaemmaghami, S., and Burlingame, A. L. (2011) *Mol. Cell Proteomics* **10**, M111.010728
24. Graham, M. E., Thaysen-Andersen, M., Bache, N., Craft, G. E., Larsen, M. R., Packer, N. H., and Robinson, P. J. (2011) *J. Proteome Res.* **10**, 2725–2733
25. Baker, P. R., Trinidad, J. C., and Chalkley, R. J. (2011) *Mol. Cell Proteomics* **10**, M111.008078
26. Matsuura, A., Ito, M., Sakaidani, Y., Kondo, T., Murakami, K., Furukawa, K., Nadano, D., Matsuda, T., and Okajima, T. (2008) *J. Biol. Chem.* **283**, 35486–35495
27. Thalhammer, A., Trinidad, J. C., Burlingame, A. L., and Schoepfer, R. (2009) *J. Neurochem.* **109**, 297–302
28. Melo, F., Sánchez, R., and Sali, A. (2002) *Protein Sci.* **11**, 430–448

29. Pieper, U., Webb, B. M., Barkan, D. T., Schneidman-Duhovny, D., Schlessinger, A., Braberg, H., Yang, Z., Meng, E. C., Pettersen, E. F., Huang, C. C., Datta, R. S., Sampathkumar, P., Madhusudhan, M. S., Sjölander, K., Ferrin, T. E., Burley, S. K., and Sali, A. (2011) *Nucleic Acids Res.* **39**, D465–474

30. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577–2637

31. Jones, D. T. (1999) *J. Mol. Biol.* **292**, 195–202

32. Schlessinger, A., Schaefer, C., Vicedo, E., Schmidberger, M., Punta, M., and Rost, B. (2011) *Curr. Opin. Struct. Biol.* **21**, 412–418

33. Duhovny, D., Nussinov, R., and Wolfson, H. J. (2002) in *Algorithms in Bioinformatics* (Guigó, R., and Gusfield, D., eds.) pp. 185–200, Springer Berlin Heidelberg, Berlin, Heidelberg

34. Andrusier, N., Nussinov, R., and Wolfson, H. J. (2007) *Proteins: Structure, Function, and Bioinformatics* **69**, 139–159

35. Nagata, Y., and Burger, M. M. (1974) *J. Biol. Chem.* **249**, 3116–3122

36. Liu, H., Sadygov, R. G., and Yates, J. R., 3rd (2004) *Anal. Chem.* **76**, 4193–4201

37. Shinoda, K., Tomita, M., and Ishihama, Y. (2010) *Bioinformatics* **26**, 576 –577

38. Schwartz, D., and Gygi, S. P. (2005) *Nat. Biotechnol.* **23**, 1391–1398

39. Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S., and Mann, M. (2010) *Sci Signal* **3**, ra3

40. Hudmon, A., and Schulman, H. (2002) *Biochem J* **364**, 593–611

41. Lucchesi, W., Mizuno, K., and Giese, K. P. (2011) *Brain Res. Bull.* **85**, 2–8

42. Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P., and Walker, S. (2011) *Nature* **469**, 564–567

43. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., and Dunker, A. K. (2004) *Nucleic Acids Res.* **32**, 1037–1049

44. Collart, M. A. (2003) *Gene* **313**, 1–16

45. Jackson, S. P., and Tjian, R. (1988) *Cell* **55**, 125–133

46. Tang, L.-Y., Deng, N., Wang, L.-S., Dai, J., Wang, Z.-L., Jiang, X.-S., Li, S.-J., Li, L., Sheng, Q.-H., Wu, D.-Q., Li, L., and Zeng, R. (2007) *Mol. Cell Proteomics* **6**, 1952–1967

47. Wells, L., Kreppel, L. K., Comer, F. I., Wadzinski, B. E., and Hart, G. W. (2004) *J. Biol. Chem.* **279**, 38466–38470

48. Yachie, N., Saito, R., Sugahara, J., Tomita, M., and Ishihama, Y. (2009) *Mol. Cell Proteomics* **8**, 1061–1071

49. Schweiger, R., and Linial, M. (2010) *Biol. Direct* **5:6**

50. Moses, A. M., Hériché, J.-K., and Durbin, R. (2007) *Genome Biol.* **8**, R23

51. Afeyan, N. B., Gordon, N. F., Mazsaroff, I., Varady, L., Fulton, S. P., Yang, Y. B., and Regnier, F. E. (1990) *J. Chromatogr.* **519**, 1–29

52. Alfaro, J. F., Gong, C.-X., Monroe, M. E., Aldrich, J. T., Clauss, T. R. W., Purvine, S. O., Wang, Z., Camp, D. G., 2nd, Shabanowitz, J., Stanley, P., Hart, G. W., Hunt, D. F., Yang, F., and Smith, R. D. (2012) *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7280–7285

53. Francisco, H., Kollins, K., Varghis, N., Vocadlo, D., Vosseller, K., and Gallo, G. (2009) *Dev Neurobiol* **69**, 162–173

54. Tallent, M. K., Varghis, N., Skorobogatko, Y., Hernandez-Cuebas, L., Whelan, K., Vocadlo, D. J., and Vosseller, K. (2009) *J. Biol. Chem.* **284**, 174–181

55. Fejtova, A., Davydova, D., Bischof, F., Lazarevic, V., Altrock, W. D., Romorini, S., Schöne, C., Zuschratter, W., Kreutz, M. R., Garner, C. C., Ziv, N. E., and Gundelfinger, E. D. (2009) *J. Cell Biol.* **185**, 341–355

56. Elgersma, Y., Fedorov, N. B., Ikonen, S., Choi, E. S., Elgersma, M., Carvalho, O. M., Giese, K. P., and Silva, A. J. (2002) *Neuron* **36**, 493–505

57. Martin, L., Latypova, X., and Terro, F. (2011) *Neurochem. Int.* **58**, 458–471

58. Arnold, C. S., Johnson, G. V., Cole, R. N., Dong, D. L., Lee, M., and Hart, G. W. (1996) *J. Biol. Chem.* **271**, 28741–28744

59. Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W., and Gong, C.-X. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10804–10809

60. Zhao, P., Viner, R., Teo, C. F., Boons, G.-J., Horn, D., and Wells, L. (2011) *J. Proteome Res.* **10**, 4088–4104

61. Hunter, T. (2007) *Mol. Cell* **28**, 730–738

62. Rexach, J. E., Rogers, C. J., Yu, S.-H., Tao, J., Sun, Y. E., and Hsieh-Wilson, L. C. (2010) *Nat. Chem. Biol.* **6**, 645–651

63. Wu, R., Haas, W., Dephoure, N., Huttlin, E. L., Zhai, B., Sowa, M. E., and Gygi, S. P. (2011) *Nat. Methods* **8**, 677–683

64. Rexach, J. E., Clark, P. M., Mason, D. E., Neve, R. L., Peters, E. C., and Hsieh-Wilson, L. C. (2012) *Nat. Chem. Biol.* **8**, 253–261

65. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041

Figure Legends

Figure 1: **Combined analysis for O-GlcNAcylation, phosphorylation and proteins.**

(A) Workflow schematic for the serial analysis of O-GlcNAcylation, phosphorylation and protein content. Synaptosome preparations were digested with trypsin, and the resulting peptides were enriched for GlcNAcylation using lectin weak affinity chromatography. The GlcNAc-depleted peptides were next enriched for phosphorylated peptides using TiO_2 chromatography. The final GlcNAc/phospho depleted fractions as well as the individual PTM-enriched fractions were each fractionated using high pH reverse phase chromatography. Individual fractions were then analyzed on an Orbitrap Velos. O-GlcNAc-enriched peptides were analyzed by ETD, phosphorylated peptides by CID, and non-modified peptides by HCD.

(B) The UV trace during three sequential runs of lectin chromatography. In each case, an aliquot of GlcNAc (eluting at 2.5 ml) was used to release tightly binding peptides. Peptides were collected as a single fraction starting at 1.2 ml, desalted and re-run for a total of three rounds. The UV shift to later elution times corresponded with a subsequent enrichment in the percentage or peptides that were O-GlcNAc-modified.

(C) The UV trace of the high pH reverse phase gradient for the final O-GlcNAc-enriched fraction. Similar gradients were used to fractionate all peptides prior to analysis by LC-MS/MS. See also Figure S1.

Figure 2: **Prevalence of O-GlcNAcylation and phosphorylation**

(A/B) Comparison of the percentage of proteins found to be O-GlcNAcylated or phosphorylated as a function of relative protein abundance. The total number of unique non-phosphorylated and non-O-GlcNAcylated peptides per protein was used as an estimate of protein abundance. Only proteins identified with at least three peptides were used in analysis for this figure. The ability to detect that a protein is O-GlcNAcylated increased with protein abundance. 19% and 63% of the highest abundance proteins were O-GlcNAcylated and phosphorylated, respectively.

(C/D) Comparison of the average number of modification sites as a function of relative protein abundance. The number of identified PTM-sites per protein increased with protein abundance. An average of 0.5 and 5.9 sites of O-GlcNAcylation and phosphorylation per protein was observed for the most abundant proteins. Note: because the proteins bassoon and piccolo were O-GlcNAcylated to a much higher extent than other proteins, they excluded from the calculations in 2C, where they would occur in the third most abundant protein group. Including them would raise the average number of O-GlcNAcylations per protein in this group from 0.40 to 1.33.

Figure 3: **PTM-modified MS/MS spectra and motif analysis.**

Peptides from the O-GlcNAc enrichment were analyzed using ETD, while those from the phosphopeptides enrichment were analyzed using CID.

(A/B) The peptide TAVKPTPIILTDQGMDLTSLAVEAR O-GlcNAcylated at the 1st or 11th amino acid, respectively.

(C/D) An example of the peptide AAVVTSPPPTTAPHK phosphorylated or O-GlcNAcylated at the serine in the 6th position, respectively.

(E) The weblogo motif for an alignment of O-GlcNAcylation sites identified in this study. (F-I) show a weblogo motif where individual amino acids are grouped by chemical property. "S" designates small/non-polar (A, G, S, T); "A" designates acidic (D, E); "B" designates basic (H, K, R); "H" designates hydrophobic (C, R< I, L, M, P, V, W); and "P" designates polar (N, Q, Y). The chemical property weblogos for the O-GlcNAc motif, phospho motif, alternatively modified sites found modified by both PTMs and the background/negative distribution (those serine/threonine residues not found modified) are shown. See also Figure S2

Figure 4: Structural and docking analysis of OGT with a series of known substrates.

(A) The catalytic domain of OGT as visualized with the surface electrostatic potential visualized using the APBS Tools plugin for Pymol (65). The peptide binding grove exists within a basic patch on the protein surface (blue). Four residues on the catalytic face that made frequent contacts with docked substrates are colored green.

(B) The TPR domains positioned against the catalytic domain. In the left side, the TPR domains are illustrated green in ribbon format. On the right side, the TPR domain is rotated 180° and shown as the charge distribution along the surface. An acidic patch can be seen (red) which would complement the basic patch on the catalytic domain in the closed confirmation.

The protein-protein interface between OGT and each docked substrate was examined. Charged OGT surface residues were identified along with oppositely charged substrate residues in which their charged moiety was found within 10Å of and oriented towards the charged OGT residue.

(C) Four residues on OGT that were found to make frequent contacts with a range of substrates are shown. The positions of these four residues are shown in red on the OGT catalytic domain.

(D) aspartate 471 shown potentially interacting residues from six substrates (Q9D1Q6-Lys106; P18760-Lys22; Q61423-Arg28; Q8OX68-Lys393; Q9CQA3-Lys139; Q8K4S1-Lys2216)

(E) arginine 475 shown potentially interacting residues from eight substrates (P18760-Asp17; P62874-Glu172; Q3UHK5-Glu228; Q9CQA3-Asp50; Q9D1Q6-Asp324; Q9CZ13-Glu157; Q3UHK5-Asp640; Q8K4S1-Asp2194).

(F) histidine 558 shown potentially interacting residues from four substrates (Q80X68-Asp455; P97450-Asp88; Q9CZ13-Glu157; P62874-Glu138).

(G) arginine 491 shown potentially interacting residues from five substrates (P14094-Asp270; Q8K4S1-Glu2171; P12787-Asp60; Q3UHK5-Glu221; Q9D1Q6-Glu290). See also Figure S3.

Figure 5: **Relationship between O-GlcNAcylation and phosphorylation sites.**

(A) A structural analysis of O-GlcNAcylation and phosphorylation sites compared to a background list of non-modified serine and threonine residues. Both PTMs showed a

propensity to occur on loop structures. In addition, both were significantly enriched in disordered protein regions.

(B) A comparison of the number of identified O-GlcNAcylation sites versus phosphorylation sites per protein. The analysis was restricted to the 919 most abundant proteins (based upon emPAI index), where our ability to identify PTMs was highest. The size of each data-point is proportional to the square root of occurrences. Note: There were 2060 proteins at the 0-0 data-point. For clarity, this data-point was given a size of 196 rather than 2060.

(C) An analysis of phosphorylation distribution with respect to primary sequence. For proteins found multiply phosphorylated, the distance along the primary sequence between pairs of phosphorylation sites was calculated. Phosphorylation sites show a propensity to cluster within regions of protein primary sequence.

(D) An analysis of O-GlcNAcylation distribution with respect to primary sequence. For proteins found multiply O-GlcNAcylated, the distance along the primary sequence between pairs of O-GlcNAcylation sites was calculated. O-GlcNAcylation sites show a propensity to cluster within regions of protein primary sequence.

(E) An analysis of the relationship between O-GlcNAcylation and phosphorylation with respect to primary sequence. For proteins found to be modified by both types of PTMs, the distance along the primary sequence from sites of O-GlcNAcylation to sites of phosphorylation was calculated. This distribution was not significantly different from the overall distribution of serine and threonine residues.

(F) An analysis of the relationship between O-GlcNAcylation sites with respect to three dimensional distance for multiply O-GlcNAcylated proteins. For those proteins with solved structures or high quality comparative models, we took each site of O-GlcNAcylation and calculated the density of additional O-GlcNAcylation sites at various distances. O-GlcNAcylation sites showed a propensity to co-occur within 10 Å of other O-GlcNAcylation sites.

(G) An analysis of the relationship between O-GlcNAcylation and phosphorylation with respect to three dimensional distance. This indicates that within alternatively modified

proteins, phosphorylation was not more likely to be localized near O-GlcNAcylation than chance alone.
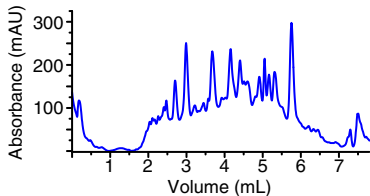
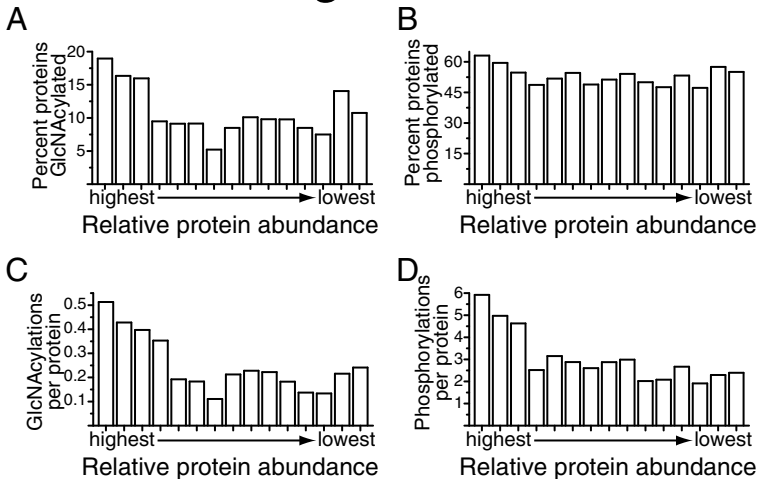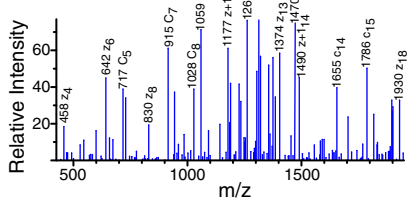**Acknowledgements**
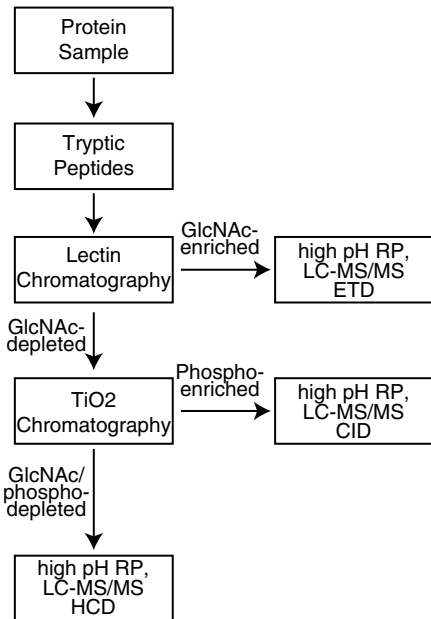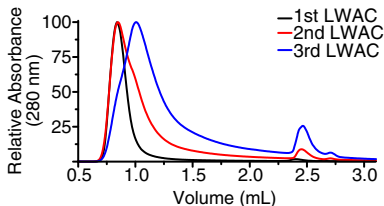
Figure 1

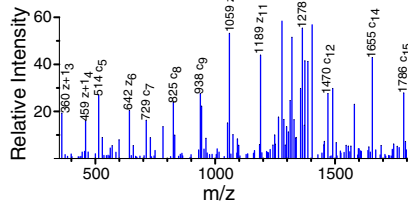Figure 2

Figure 3

A

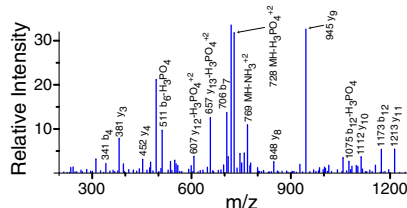gTAVKPTPIILTDQGMDLTSLAVEAR

B

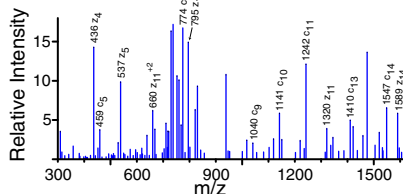TAVKPTPIILgTDQGMDLTSLAVEAR

C

AAVVTpSPPPTTAPHK

D

AAVVTgSPPPTTAPHK

E GlcNAc motif (amino acid)

F GlcNAc motif (chemical group)

G phospho motif (chemical group)

H reciprocal motif (chemical group)

I negative motif (chemical group)

S = small/non-polar
H = hydrophobic
B = basic
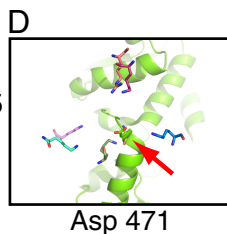A = acidic
P = polar

Figure 4

Figure 5