Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides

Robert J. Chalkley[a], Agnes Thalhammer[b], Ralf Schoepfer[a,b], and A. L. Burlingame[a,1]

[a]Department of Pharmaceutical Chemistry, University of California, 600 16th Street, Genentech Hall, Suite N472, San Francisco, CA 94158; and [b]Laboratory for Molecular Pharmacology, and Department of Neuroscience, Physiology, and Pharmacology, University College London, Gower Street, London WC1E 6BT, United Kingdom

Edited by James A. Wells, University of California, San Francisco, CA, and approved April 15, 2009 (received for review January 12, 2009)

Protein O-GlcNAcylation occurs in all animals and plants and is implicated in modulation of a wide range of cytosolic and nuclear protein functions, including gene silencing, nutrient and stress sensing, phosphorylation signaling, and diseases such as diabetes and Alzheimer's. The limiting factor impeding rapid progress in deciphering the biological functions of protein O-GlcNAcylation has been the inability to easily identify exact residues of modification. We describe a robust, high-sensitivity strategy able to assign O-GlcNAcylation sites of native modified peptides using electron transfer dissociation mass spectrometry. We have studied the murine postsynaptic density pseudoorganelle and report the assignment of 58 modification sites from a single experiment–significantly increasing the number of sites known in the literature. Components of several repressor complexes, such as NCoR1, polyhomeotic-like protein3, and EMSY, are modified. In addition, 28 O-GlcNAc sites were found on the protein Bassoon, effectively matching the number of phosphorylation sites reported previously on this protein. This finding suggests that on certain proteins, O-GlcNAcylation may be as extensive and important as phosphorylation in regulating protein function. Three of the newly discovered O-GlcNAc sites on Bassoon have previously been reported as phosphorylation sites, highlighting the interplay of the modifications. Surprisingly, several peptides with GlcNAc modifications on asparagines within the N-X-S/T consensus sequence were also observed from membrane protein extracellular domains. This powerful strategy fulfills a long-standing need in the biological community by facilitating modification site identifications that will accelerate understanding of the biological significance of this elusive regulatory posttranslational modification.

modification site assignment | posttranslational modification | proteomics | Bassoon

Regulation of protein activity through reversible posttranslational modification (PTM) is the primary method the cell uses to rapidly respond to endogenous and environmental stimuli (1). Therefore, characterizing the locations and dynamics of these different modifications is vital for understanding signal transduction and regulation, and many other protein functions.

Modification of serine and threonine residues of nuclear and cytoplasmic proteins with a single O-linked *N*-acetylglucosamine (GlcNAc) residue is increasingly being shown to be important in response to changes in nutritional state and cellular stress (2–5). O-GlcNAc modification can regulate targeting of proteins for proteasomal degradation and is probably important in regulating gene expression judging from the high occurrence on transcription factors (6). Gaining site-specific occupancies of proteins in pathways and networks should have great clinical importance because of the involvement of O-GlcNAcylation in diseases such as diabetes (4) and Alzheimer's (2). The role of this modification in the nervous system in particular is gaining increasing interest as not only are the enzymes that transfer and remove O-GlcNAc, the O-GlcNAc transferase (OGT), and O-GlcNAcase, respectively, enriched in synaptic contacts (7), but also proteins crucial in neuronal signaling and synaptic plasticity have been found O-GlcNAcylated (8). The importance of O-GlcNAcylation is highlighted by the brain-specific OGT knockout mouse model that displays developmental defects that result in neonatal death (9). Most interestingly, in analogy to phosphorylation, synaptic activity leads to a dynamic modulation of O-GlcNAc levels (10) and is also involved in long-term potentiation and memory (11).

Phosphorylation is the most heavily studied regulatory PTM, and proteomic approaches for global enrichment, detection, and characterization of this modification have provided the basis of our current knowledge of the complexity and redundancy in cell signaling networks (12–16). O-GlcNAc may be as widespread and important as phosphorylation, but currently the tools required to identify sites and thereby decipher a mechanistic understanding of the modification and its interplay with other PTMs are lacking to the extent that this deficiency has recently been described as the "most important impediment to research on O-GlcNAc" (2).

Because of the difficulties in identifying sites of modification, there are currently <80 exact residues of modification known on all proteins (17), the majority of which were determined by radio labeling, then Edman degradation. Chemical derivatization methods allow efficient enrichment of GlcNAc-modified peptides (18, 19), but have led to only a few site assignments in each study thus far.

Identification of sites of modification even using mass spectrometry has proven difficult. The predominant mode of peptide fragmentation used in mass spectrometers is collision-induced dissociation (CID), which is a vibrational activation fragmentation process that breaks the weakest bonds in the structure. The O-linkage between sugar modification and serine or threonine is considerably more labile than the peptide backbone, so under CID conditions the sugar residue is generally eliminated before peptide fragmentation, returning no information on the site of modification (20, 21). Recently, 2 different radical-based fragmentation approaches have become available: electron capture dissociation (ECD) (22) and electron transfer dissociation (ETD) (23). We have previously shown that under ECD conditions O-GlcNAc modification sites can be identified on native peptides (24), and ETD has also been used to identify such sites (5, 10, 25).

As high levels of O-GlcNAc modification have been reported in nerve terminals (7), postsynaptic density preparations (PSD) are a rich source for O-GlcNAc-modified peptides, and several large-scale studies have used these preparations to discover new O-GlcNAc modifications (24, 26). Here, we have used our lectin-based

Author contributions: R.J.C., R.S., and A.L.B. designed research; R.J.C. and A.T. performed research; R.J.C. analyzed data; and R.J.C., A.T., R.S., and A.L.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: alb@cgl.ucsf.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0900288106/DCSupplemental.

enrichment lectin weak affinity chromatography (LWAC) strategy (24) with online liquid chromatography and ETD analysis using an LTQ Orbitrap mass spectrometer to identify a large number of modification sites. This protocol has allowed identification of >50 previously undescribed sites of O-GlcNAc modification, including sites on several transcription repressor complex components and 28 sites of modification in the protein Bassoon, showing this strategy to be comfortably the most successful to date for assigning O-GlcNAc modification sites.

Results

Enrichment of GlcNAcylated Peptides. LWAC using a 3-m-long column did not produce a distinct UV-visible peak that eluted after the bulk of the unmodified peptides (see *SI Appendix*), so fractions from the tail of the main UV peak and subsequent fractions (13 fractions in total) were all analyzed by liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) employing both CID and ETD fragmentation.

Results from CID Data. Under CID conditions the O-glycosidic linkage between sugar and modified residue is expected to fragment before peptide backbone cleavage. The prominent loss of the sugar residue and formation of a GlcNAc oxonium ion can be used as markers to identify O-GlcNAc-modified peptides (20, 24). However, this prominent loss means that fragments usually do not contain the modification, and information about the site of modification is lost. Hence, the CID data of O-GlcNAc-modified peptides were analyzed allowing for either retention or loss of the sugar moiety from fragment ions.

Results from CID analysis permitted identification of 47 GlcNAc-modified peptides, but usually these spectra did not contain sufficient information to allow pinpointing of the site of modification. In fact, in only one case did the CID spectrum contain sufficient information to identify a peptide and site that could not be determined from the corresponding ETD spectrum. However, using the CID spectrum as a reference to identify the peptide, there were 3 occasions in which the ETD spectrum could then be used to determine the site of modification.

Results from ETD Data. The O-glycosidic linkage is not labile under ETD conditions. Hence, unlike in CID, there is no characteristic of ETD spectra of GlcNAc-modified peptides that allows immediate recognition of its modification state. However, fragments retaining the modification are observed, allowing assignment of the residue of O-GlcNAcylation. Hence, ETD data were analyzed assuming the modification is retained on fragment ions.

Interpretation of the ETD dataset revealed the identity of just >1,000 peptides from ≈300 unique proteins. Of these identified spectra, ≈35% were of doubly charged precursors, 60% of triply charged precursor ions, and the remaining 5% were 4+ or 5+ ions.

Eighty-seven of these ≈1,000 peptides were found to be modified by a single N-acetylhexosamine residue. The high-quality fragmentation data produced by ETD measurement in the linear ion trap allowed assignment of at least one site of GlcNAc modification in 67 (77%) of the modified peptide spectra by manual analysis, with some peptides containing multiple O-GlcNAc modifications. Indeed, for 6 of the peptides in which it was not possible to identify the exact site(s) of modification, the ETD data indicated that the spectra appeared to be a mixture of 2 positional isomers of the same peptide with O-GlcNAc modifications on different residues. In a few ETD spectra a loss of O-GlcNAc was observed from the charge-reduced species, but there was never any evidence of loss of sugar moiety from any other fragment ions.

An example of an O-GlcNAc-modified peptide is given in Fig. 1. This image shows the spectrum of a peptide from the scaffolding protein actin-binding LIM protein 1. In this spectrum, fragments formed by cleavage of most of the peptide backbone bonds were observed, and the mass difference of 291 Da between the $z_{\cdot 11}$ and

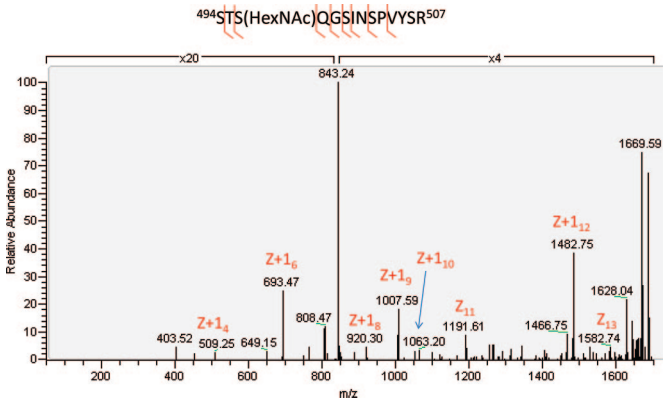

Fig. 1. ETD spectrum of an *m/z* 843.402 2+ precursor identifies serine 496 as a site of O-GlcNAc modification of actin-binding LIM protein 1. Serine 496 is also known to be a site of phosphorylation.

$(z + 1)_{12}$ ions definitively identifies serine 496 as a new site of O-GlcNAc modification.

Fig. 2 shows the ETD fragmentation spectrum of a peptide from Disks large-associated protein 1. This peptide bears 2 O-GlcNAc modifications and a phosphoryl moiety. Both c and z· ion series show extensive sequence coverage of the C-terminal half of this peptide and the mass differences between z·4 to z·5 and z·5 to z·6 (or in the other direction c7 to c8 and c6 to c7) identify the two O-GlcNAc modification sites as threonines 525 and 526. Unfortunately, fragments from the N-terminal part of the peptide were not observed, so it is not possible to determine which of the 3 serine residues was phosphorylated.

The complete list of sites of O-GlcNAc-modified residues determined in this study is provided in Table 1, and the corresponding annotated spectra supporting the site identifications are presented in *SI Appendix*.

N-linked GlcNAc Modification. Upon manual inspection, within the list of 67 GlcNAc-modified peptides there were 9 spectra, corresponding to 8 different sites, where the modification site was actually an asparagine residue located on an extracellular domain of a membrane protein. N-linked GlcNAc is stable in CID mass spectrometry (27), so these sites could also be identified from the CID spectra. An example of an ETD spectrum of a N-linked GlcNAc-modified peptide is given in Fig. 3. This peptide is from the N-terminal extracellular domain of the transmembrane Gamma-

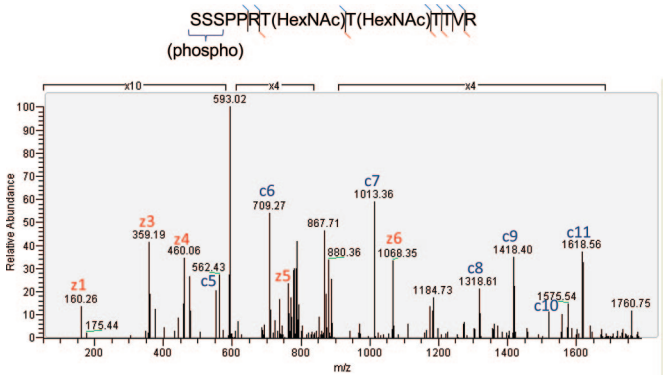

Fig. 2. ETD spectrum of an *m/z* 592.605 3+ precursor identifies a peptide from Disks large-associated protein 1 with 2 O-GlcNAc modifications and a phosphorylation. The sites of O-GlcNAc modification can be identified as threonines 525 and 526. The phosphorylation is on one of the serine residues.

BIOCHEMISTRY

Table 1. Sites of O-GlcNAc modification identified in this study

Accession no.	Protein	O-GlcNAc modification site/s
O88737	Protein Bassoon	1354, 1395, 1418, 1445, 1517, 1537, 1657, 1666, 1707, 1772, 1962, 2027, 2029, 2058, 2067, 2068, 2070, 2091, 2141, 2188, 2295, 2317, 2318, 2694, 2700, 2703, 2941, 2945
Q9QYX6	Protein piccolo	2634, 2639, 2656, 2930, 2948, 3873
Q69ZX8	Actin-binding LIM protein 3	383, 419, 423, 534, 546, 547
Q8K4G5	Actin-binding LIM protein 1	496, 499
P08553	Neurofilament medium polypeptide	37, 46, 430
P08551	Neurofilament light peptide	48, 414
O70511	270 kDa Ankyrin G isoform	1520
P97836	Disks large-associated protein 1	525, 526
Q8BMB0	Protein EMSY	499
O35927	Catenin delta-2	447, 453
Q9WV69	Dematin	285
Q60974	Nuclear receptor corepressor 1	1496
Q8CHP6	Polyhomeotic-like protein 3	238
Q3UHF7	Human immunodeficiency virus type-1 enhancer-binding protein 2	1271
Q9EQZ7	Regulating synaptic membrane exocytosis protein 2	1528

aminobutyric acid type B receptor. The mass difference between z·8 and z·9 ions corresponds to the glycosylated asparagine residue.

The sites of N-linked GlcNAc modification detected in this study are presented in Table 2, and the spectra supporting these site assignments are in *SI Appendix*. All of the proteins detected here are membrane proteins, mostly transmembrane receptors, and all sites are located in extracellular regions, where one would expect conventional N-glycosylation. All sites are located in consensus sequences for N-linked glycosylation (N-X-S/T), so are predicted sites of glycosylation.

There were also a number of complex N-linked glycan-modified peptides detected after LWAC enrichment. These showed characteristic formation of sugar fragments under CID conditions, but minimal peptide backbone cleavage, preventing peptide identification. Under ETD conditions, these components produced essentially no fragmentation products because almost all were over *m/z* 1,000 and therefore not charge-dense enough for efficient fragmentation (see *Discussion* for a more detailed explanation).

Discussion

The groundwork for this study was laid by our previous analysis of postsynaptic density after LWAC enrichment, where ECD or modified peptide derivatization were used to facilitate site identification (24). In this initial study, 18 sites of modification were determined using the combination of 3 different methods for site assignment, corresponding to a significant amount of work. In the present study we identified 58 sites from a single analysis of one PSD preparation. These ETD findings represent a dramatic increase in the number of modification sites that have been determined in a single experiment. This improvement is largely because of the increased sensitivity of ETD over ECD, allowing characterization of a large number of modified peptides and sites on a chromatographic time scale. The mass precision of the orbitrap for measurement of precursor mass is also important, as it significantly reduces the number of possible peptides needed to be considered, which is especially important when looking for posttranslational modifications, where every peptide has to be considered with the modification on any possible residue.

We do not suggest that these findings are comprehensive. The 58 sites identified in the current study include only 6 (of 18) of the same sites reported previously (24). The samples for the 2 studies were different preparations, so not identical, which could have led to biological differences in the modifications present. Some of these differences are likely because of analytical variability, as the mass spectrometer is selecting precursors for fragmentation in a data-dependent manner, and we are only identifying those that were fragmented. Hence, even if the same sample is analyzed twice in the mass spectrometer, slightly different peptides are typically identified in each run. To reduce this effect, an additional dimension of separation could be used so that each fraction is simpler, allowing

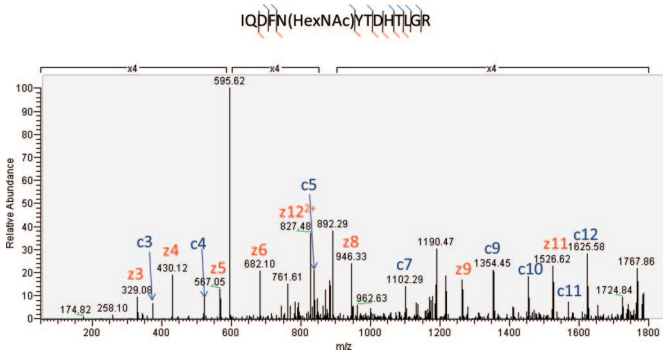

Fig. 3. ETD spectrum of an *m/z* 594.946 3+ identifies a peptide from Gamma-aminobutyric acid type B receptor subunit 2 with a single GlcNAc residue attached to asparagine 388.

Table 2. Sites of N-linked GlcNAcylation identified in this study

Accession no.	Protein	Site of N-glycosylation
Q00960	Glutamate [NMDA] receptor subunit epsilon-2 precursor	688
Q9R1V6	ADAM 22 precursor	517
Q80T41	Gamma-aminobutyric acid type B receptor subunit 2 precursor	388
P35438	Glutamate [NMDA] receptor subunit zeta-1 precursor	368
P55066	Neurocan core protein precursor	121
Q63912	Oligodendrocyte-myelin glycoprotein precursor	234
P13638	Sodium/potassium-transporting ATPase subunit beta-2	238
P01831	Thy-1 membrane glycoprotein precursor	42

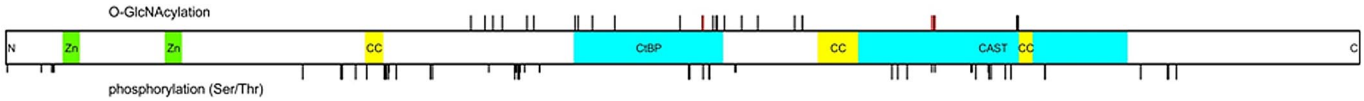

Fig. 4. Posttranslational modifications on Bassoon. The position of the 28 O-GlcNAc modification sites on protein Bassoon (3940 aa residues) from this study (Table 1) are indicated in relation to phosphorylation sites and its structural domains (39). Phosphorylated serine/threonine residues found on PSD-associated Bassoon are indicated by long lines (36, 37); short lines indicate additional sites described in preparations of synaptsomes (42, 43, 47). O-GlcNAc sites (at 2029, 2694, and 2703), which have also been reported as phosphorylation sites on Bassoon, are indicated in red. Zn, zinc finger domain; CC, predicted coil–coil domains; CtBP, area of interaction with the CtBP (C-terminal binding protein); CAST, area of interaction with ELKS/CAST/ERC proteins; N, N terminus; C, C terminus.

deeper analysis of the sample. Nevertheless, the relatively low overlap in site assignments suggests that there are still many more modifications present in this sample.

The efficiency of the ETD process for peptide fragmentation is reported to be related to charge density, with fragmentation of higher charge and lower *m*/*z* components providing more informative data (28). Our findings support this observation, in that two-thirds of all peptides identified and 64 of the 87 GlcNAc-modified peptides in the ETD results were triply charged or higher. From a tryptic digest one would expect the vast majority of peptides to be doubly charged, so these findings do suggest a bias toward characterizing higher charge-state peptides. However, combining CID data with ETD of the same peptides has the potential to mitigate this level of bias. CID analysis of tryptic peptides is most efficient with doubly charged peptides. Hence, if the CID data can be used for identifying the peptide, then a lower-quality ETD spectrum may still be good enough to provide site identification information, and this situation was the case for 3 of the sites identified in this study. In this study we did not adjust our data acquisition strategy to maximize peptide identifications from the CID data, because the emphasis was on identifying modification sites. The predominant fragmentation of O-GlcNAc-modified peptides in CID is the loss of sugar residue from the intact peptide, and this loss can lead to spectra with low information content in terms of identifying the peptide sequence (21). However, the selective observation of this loss can be used to trigger a further level of fragmentation analysis (referred to as MS3 analysis), that can provide better information for peptide identification (24). However, as the MS3 data does not assist in modification site identification, which was the emphasis of this work, and the MS3 acquisition would take extra time, meaning fewer precursors will be analyzed, we chose to maximize our chances of identifying modification sites over modified peptide sequences.

The use of a different digestion enzyme could also provide complementary information. Endoprotease Lys C has been used for ETD studies (29). It produces longer peptides, often with higher charge states. However, whether this strategy leads to better ETD fragmentation data is open to debate because the extra peptide length counteracting the increased charge state leads to, on the whole, peptides with a similar charge density to tryptic peptides.

N-linked GlcNAc Modification. The discovery of a number of N-linked GlcNAc-modified peptides is a surprising result, because a single N-linked GlcNAc is not be expected to be present on a protein. A possible explanation is that these are artifacts formed during sample preparation, probably after cell lysis. In the cytosol there is an enzyme, endo-β-N-acetylglucosaminidase (ENGase) that cleaves sugars to leave a single GlcNAc residue (30). This enzyme is predicted to act on free oligosaccharides, breaking them down on the way to lysosomal degradation. However, to the best of our knowledge, it has not been shown to have activity against oligosaccharides attached to a protein. Thus, the precise origin of these components must await further study.

The discovery of a number of N-linked GlcNAc modification sites in this study highlights that caution must be applied when reporting O-GlcNAc modification (one wonders whether antibodies that recognize O-linked GlcNAc would be able to distinguish this form from N linked), and emphasizes the importance of not only detecting the modification but being able to identify modified peptides and exact amino acid sites.

Functions of the Identified O-GlcNAc-Modified Proteins. Many of the modified proteins identified here, such as Bassoon, Piccolo, actin-binding LIM proteins 1 and 3, Ankyrin G, Disks large-associated protein 1, and the neurofilament proteins, are structural/scaffolding proteins closely associated with the plasma membrane on one side or other of the synaptic cleft. Several of these proteins were identified as O-GlcNAc modified in our previous study (24), although this is the first time the actin-binding LIM proteins and Disks large-associated protein 1 have been detected modified.

The detection of O-GlcNAcylation on proteins at the plasma membrane is intriguing in light of the reported translocation of the O-GlcNAc transferase (OGT), responsible for addition of the modification, to the plasma membrane in response to insulin stimulation (4). These findings hint that a similar localization may occur in neural cells and may be a common approach for allowing O-GlcNAcylation to regulate signal transduction initiated at the plasma membrane.

The identification of an O-GlcNAc modification site on nuclear receptor corepressor 1 (NCoR) is particularly interesting. This protein binds along with histone deacetylases in corepressor complexes to repress gene transcription of selected genes (31). OGT has been shown to copurify with NCoR in these complexes (32). It is proposed that the binding of OGT with the corepressor complex allows targeting of the enzyme to promoter regions, where it modifies transcriptional machinery such as RNA polymerase II and transcription factors. We believe this finding is the first evidence to show that OGT modifies members of the corepressor complex.

There are other proteins identified as modified in this study that are also involved in gene transcriptional repression. Polyhomeotic-like protein 3 is part of the polycomb PRC1 complex involved in long-term gene silencing (33), and the protein is believed to be a tumor suppressor protein (34). Another tumor suppressor protein, particularly associated with breast and ovarian cancer, is EMSY, a transcriptional repressor of the BRCA2 gene (35).

Obviously, more studies will be required to elucidate the exact role that O-GlcNAc is playing on these proteins, but the identification of exact sites of modification is a significant step toward facilitating this process.

High Abundance of O-GlcNAc Modification on Selected Proteins. Approximately half of the O-GlcNAc sites identified in this study were located on the protein Bassoon, and 6 sites were found on the related protein Piccolo (Table 1). The density of O-GlcNAc modification sites on Bassoon is comparable with its phosphorylation reported on serine/threonine residues (Fig. 4). In contrast, not a single O-GlcNAc site was found on some other highly phosphorylated PSD proteins, such as the microtubule-associated proteins (MAP), where 14, 28, and 11 phosphorylation sites have been identified on MAP 1A, 1B, and 2 in PSD, respectively (36, 37). The narrower scope of proteins subjected to O-GlcNAcylation may well reflect a less-evolved O-GlcNAcylation system, for which only 1 transferase is present in animals and 2 in plants (2), in contrast to the 540 kinases in mouse (38). Both Bassoon and the O-

Fig. 5. Sequence surrounding sites of O-GlcNAc modification reported in this study. This representation was created using WebLogo (48).

GlcNAcylation system are more recent additions of evolution. Bassoon is present in vertebrates, but no ortholog is known in worms or flies (39). The O-GlcNAcylation system apparently evolved with multicellular organisms (2).

Bassoon, together with the related protein Piccolo, are major components of the cytomatrix of the presynaptic active zone, which plays a critical role in the precise temporal and spatial control of the vesicular release of neurotransmitters (39). O-GlcNAcylation will increase the hydrophilicity of the protein or its modified segment, and modify its ability to interact with other proteins. Several O-GlcNAc sites on Bassoon are in regions involved in interaction with the proteins CtBP and CAST (Fig. 4), which are required for the correct formation of the presynaptic active zone (39, 40), suggesting a possible functional role for the modification.

O-GlcNAcylation and Phosphorylation. The interaction and, in a few cases, competition for site occupancy between O-GlcNAcylation and phosphorylation as regulatory modifications has created a lot of interest, and there are many examples where the modifications are clearly affecting each other (41). Of the 58 O-GlcNAc modification sites reported in this study, 4 have also been reported as phosphorylation sites: Ser-2029 (37, 42), Ser-2694 (43), and Thr-2703 (44) in Bassoon, and Ser-496 in actin-binding LIM 1 protein (45). In addition, we observed a peptide from Disks large-associated protein 1 with O-GlcNAc and phosphate modifications at different sites on the same peptide (Fig. 2). Hence, this data should further stoke interest in the relationship between these modifications.

Modification Site Sequence Specificity. There is not a strict consensus sequence for O-GlcNAc modification sites, but preference for a proline 2 or 3 residues N-terminal has been noted (2, 24). This preference is confirmed in the sequences surrounding the sites of modification identified in this study, a representation of which is shown in Fig. 5. A proline is frequently observed in the -3 position relative to the site of modification, whereas in most other positions within 5 residues of the modification site serines or threonines are most common.

Benefits over an Alternative Strategy. Another strategy reported for enrichment of GlcNAc-modified peptides is a chemoenzymatic tagging strategy that allows attachment of a biotin tag to the sugar residue (19). Though this tag allows efficient enrichment of modified peptides, it has a deleterious effect on the ability to identify modified peptides and sites. In CID, the tag fragments prominently, which can be useful for confirming the presence of the tag in the fragmented component, but fragmentation of the tag rather than the peptide makes peptide identification difficult. ETD analysis of these tagged peptides has also been used, but it allowed identification of only 3 sites of modification in the published study (10). This situation is probably because the tag is also problematic for subsequent ETD fragmentation analysis. As previously mentioned, ETD fragmentation is only efficient on charge-dense components. The chemically introduced tag is neutral and has a mass of $>$800 Da, which means that ETD analysis of the majority of tagged peptides will provide inadequate information for robust modification site assignment. In our strategy we do not introduce a tag, so peptides obtained are more charge dense and amenable to successful ETD-based fragmentation and site assignment.

The biological importance of protein and cellular regulation through O-GlcNAc posttranslational modification is becoming clearer with the development of new techniques to study the modification. We conclude that ETD of native O-GlcNAc-modified peptides is the most effective approach for identification of GlcNAc-modified peptides and sites. In particular, the combination of the precursor mass accuracy provided by an orbitrap and the fragmentation sensitivity of an ion trap should allow identification of large numbers of O-GlcNAc-modified peptides and sites, furthering the understanding of this regulatory modification in homeostasis and disease.

Methods

Sample Preparation. Postsynaptic density preparations from murine brains (ICR mice, aged 8 weeks) were obtained by a sucrose density centrifugation protocol as described in ref. 36. The inhibitor O-(2-acetamido-2-deoxy-Dglucopyranosylidene) amino-N-phenylcarbamate (PUGNAc, 20 μM) (Carbogen) was added until synaptosomes had been purified, but omitted for the last steps to dilute the PUGNAc concentration to levels that were not interfering with LWAC. Ten whole brains yielded 11 mg of PSD material. Two milligrams of material was used for this study.

The PSD pellet was resuspended in 6 M guanidine HCl; proteins were reduced using 2 mM Tris(2-carboxyethyl)phosphine for 1 h at 55 °C, then alkylated using 10 mM iodoacetamide before digestion for 40 h using 40 μg of modified trypsin (Promega; part no. 9PIV5113). After digestion, the sample was acidified using formic acid and then desalted using a C18 Sep-Pak (Waters) before drying down by vacuum centrifugation.

LWAC. A WGA column of approximately 3 m in length was created by packing WGA-agarose resin (Vector Laboratories) into Teflon tubing (1/16 inch (1.6 mm) outer diameter, 1/25 inch (1.0 mm) inner (Upchurch Scientific). Peptides were resuspended in 200 μL LWAC buffer (25 mM Tris [pH 7.8], 300 mM NaCl, 5 mM CaCl, 1 mM $MgCl_2$). GlcNAc peptide enrichment was performed using the WGA-agarose column and LWAC buffer at a flow rate of 100 μL/min. Fractions were manually collected at 1-min intervals during the elution of the main UV-visible peak and subsequent tail of the peak. A total of 13 fractions were analyzed by mass spectrometry to identify GlcNAc-modified peptides. At the end of the run, 200 μL of LWAC buffer containing 20 mM GlcNAc (Sigma) was injected to elute any complex glycans.

LC-MS/MS. Sample fractions were desalted using C18 ZipTips (Millipore), then analyzed by LC-MS/MS. Chromatography was performed using a Nanoacuity HPLC (Waters) at a flow rate of 300 nL/min. The column was a BEH130 C18 75 μM ID $\times$ 150 mm (Waters), and a 90-min gradient was used. Solvent A was water/0.1% formic acid, and solvent B was acetonitrile/0.1% formic acid; peptides were eluted by a gradient from 2% to 28% solvent B $>$70 min followed by a short wash at 50% solvent B, before returning to starting conditions. Peptide components eluted over a period of $\approx$60 min during these runs. Mass spectrometry was performed using an LTQ Orbitrap with the ability to perform ETD analysis (Thermo). After a precursor scan of intact peptides was measured in the orbitrap by scanning from *m/z* 350–1,500, the 3 most intense multiply charged precursors were selected for both CID and ETD analysis in the linear ion trap. Activation times were 30 msec and 100 msec for CID and ETD fragmentation, respectively. Automatic gain control (AGC) targets were 100,000 ions for orbitrap scans and 10,000 for MS/MS scans, and the AGC for the fluoranthene ions used for ETD was 100,000. Supplemental activation of the charge-reduced species was used in the ETD analysis to

improve fragmentation. Dynamic exclusion for 60 sec was used to prevent repeated analysis of the same components.

Data Analysis. Fragmentation data were converted to peaklists using an in-house script, then CID and ETD data were searched separately using Protein Prospector version 5.1.3 (46) against a database that consisted of the Swiss-Prot protein database downloaded on April 24, 2008, to which a randomized version had been concatenated. Only rodent entries were searched, meaning a total of 44,512 entries were queried. A second search was performed against rodent entries in a concatenated version of UniprotKB (Swiss-Prot and TrEMBL) downloaded on June 10, 2008 (150,724 entries searched). ETD peptide results were reported using a peptide false discovery rate level of 0.5% according to concatenated database search results. Where multiple protein entries matched the same peptide, the mouse version was preferentially reported.

For ETD data, a precursor mass tolerance of 15 ppm and a fragment mass error tolerance of 1.2 Da were allowed. The wide mass tolerance for fragment ions was to allow matching of $z + 1$ ions to allow for incorrect monoisotopic peak detection (which in our experience is more frequent in ETD data than CID data because of the generally higher mass and *m/z* of fragment ions). Cysteine residues were assumed to be carbamidomethylated. Other considered modifications were protein N-terminal acetylation, peptide N-terminal glutamine conversion to pyroglutamate, methionine oxidation, HexNAc modification of serine or threonine residues, or phosphorylation of serine or threonine residues. For CID data a precursor mass error tolerance of 15 ppm and a fragment mass tolerance of 0.6 Da were allowed. The same modifications as for ETD were considered, but in addition, a modification observed as a neutral loss of 203.08 Da (so all fragments are assumed to not have GlcNAc attached) was considered. Assignments of all modified peptides were checked manually; in every case, the CID results were consistent with the ETD assignment, and annotated versions of all spectra allowing modification site assignments are supplied in *SI Appendix*.

ACKNOWLEDGMENTS. We thank Katalin Medzihradszky and Jonathan Trinidad for useful discussion. This work was supported by National Institutes of Health Grant RR001614, National Center for Research Resources Grant SIG RR019934, and the Biotechnology and Biological Sciences Research Council of the United Kingdom.

1. Walsh CT (2006) *Posttranslational Modification of Proteins*: *Expanding Nature's Inventory* (Roberts and Co., Englewood, CO).
2. Hart GW, Housley MP, Slawson C (2007) Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 446:1017–1022.
3. Zachara NE, Hart GW (2006) Cell signaling, the essential role of O-GlcNAc! *Biochim Biophys Acta* 1761:599–617.
4. Yang X, et al. (2008) Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* 451:964–969.
5. Housley MP, et al. (2008) O-GlcNAc regulates FoxO activation in response to glucose. *J Biol Chem* 283:16283–16292.
6. Love DC, Hanover JA (2005) The hexosamine signaling pathway: Deciphering the ''O-GlcNAc code''. *Sci STKE* 2005:re13.
7. Cole RN, Hart GW (2001) Cytosolic O-glycosylation is abundant in nerve terminals. *J Neurochem* 79:1080–1089.
8. Murrey HE, Hsieh-Wilson LC (2008) The chemical neurobiology of carbohydrates. *Chem Rev* 108:1708–1731.
9. O'Donnell N, Zachara NE, Hart GW, Marth JD (2004) Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol Cell Biol* 24:1680–1690.
10. Khidekel N, et al. (2007) Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol* 3:339–348.
11. Tallent MK, et al. (2009) In vivo modulation of O-GlcNAc levels regulates hippocampal synaptic plasticity through interplay with phosphorylation. *J Biol Chem* 284:174–181.
12. Olsen JV, et al. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127:635–648.
13. Beausoleil SA, et al. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci USA* 101:12130–12135.
14. Collins MO, Yu L, Choudhary JS (2007) Analysis of protein phosphorylation on a proteome-scale. *Proteomics* 7:2751–2768.
15. Thingholm TE, Jorgensen TJ, Jensen ON, Larsen MR (2006) Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nat Protoc* 1:1929–1935.
16. Thingholm TE, Jensen ON, Robinson PJ, Larsen MR (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol Cell Proteomics* 7(4):661–671.
17. Copeland RJ, Bullen JW, Hart GW (2008) Cross-talk between GlcNAcylation and phosphorylation: Roles in insulin resistance and glucose toxicity. *Am J Physiol Endocrinol Metab* 295:E17–E28.
18. Wells L, et al. (2002) Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* 1:791–804.
19. Khidekel N, et al. (2003) A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J Am Chem Soc* 125:16162–16163.
20. Greis KD, et al. (1996) Selective detection and site-analysis of O-GlcNAc-modified glycopeptides by beta-elimination and tandem electrospray mass spectrometry. *Anal Biochem* 234:38–49.
21. Chalkley RJ, Burlingame AL (2001) Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. *J Am Soc Mass Spectrom* 12:1106–1113.
22. Zubarev RA (2004) Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol* 15:12–16.
23. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA* 101:9528–9533.
24. Vosseller K, et al. (2006) O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 5:923–934.
25. Housley MP, et al. (2009) A PGC-1alpha:O-GlcNAc transferase complex regulates FoxO transcription factor activity in response to glucose. *J Biol Chem* 284:5148–5157.
26. Khidekel N, Ficarro SB, Peters EC, Hsieh-Wilson LC (2004) Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci USA* 101:13132–13137.
27. Medzihradszky KF (2005) Characterization of protein N-glycosylation. *Methods Enzymol* 405:116–138.
28. Swaney DL, McAlister GC, Coon JJ (2008) Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* 5(11):959–964.
29. Chi A, et al. (2007) Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci USA* 104:2193–2198.
30. Suzuki T, Funakoshi Y (2006) Free N-linked oligosaccharide chains: Formation and degradation. *Glycoconj J* 23:291–302.
31. Alland L, et al. (1997) Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. *Nature* 387:49–55.
32. Bowe DB, et al. (2006) O-GlcNAc integrates the proteasome and transcriptome to regulate nuclear hormone receptors. *Mol Cell Biol* 26:8539–8550.
33. Shao Z, et al. (1999) Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98:37–46.
34. Deshpande AM, et al. (2007) PHC3, a component of the hPRC-H complex, associates with E2F6 during G0 and is lost in osteosarcoma tumors. *Oncogene* 26:1714–1722.
35. Hughes-Davies L, et al. (2003) EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* 115:523–535.
36. Trinidad JC, Specht CG, Thalhammer A, Schoepfer R, Burlingame AL (2006) Comprehensive identification of phosphorylation sites in postsynaptic density preparations. *Mol Cell Proteomics* 5:914–922.
37. Trinidad JC, et al. (2008) Quantitative analysis of synaptic phosphorylation and protein expression. *Mol Cell Proteomics* 7:684–696.
38. Caenepeel S, Charydczak G, Sudarsanam S, Hunter T, Manning G (2004) The mouse kinome: Discovery and comparative genomics of all mouse protein kinases. *Proc Natl Acad Sci USA* 101:11707–11712.
39. Schoch S, Gundelfinger ED (2006) Molecular organization of the presynaptic active zone. *Cell Tissue Res* 326:379–391.
40. Takao-Rikitsu E, et al. (2004) Physical and functional interaction of the active zone proteins, CAST, RIM1, and Bassoon, in neurotransmitter release. *J Cell Biol* 164:301–311.
41. Wang Z, Gucek M, Hart GW (2008) Cross-talk between GlcNAcylation and phosphorylation: Site-specific phosphorylation dynamics in response to globally elevated O-GlcNAc. *Proc Natl Acad Sci USA* 105:13793–13798.
42. Collins MO, et al. (2005) Proteomic analysis of in vivo phosphorylated synaptic proteins. *J Biol Chem* 280:5972–5982.
43. Dai J, et al. (2007) Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. *J Proteome Res* 6:250–262.
44. Armstrong SA, Barry DA, Leggett RW, Mueller CR (1997) Casein kinase II-mediated phosphorylation of the C terminus of Sp1 decreases its DNA binding activity. *J Biol Chem* 272:13489–13495.
45. Villen J, Beausoleil SA, Gerber SA, Gygi SP (2007) Large-scale phosphorylation analysis of mouse liver. *Proc Natl Acad Sci USA* 104:1488–1493.
46. Chalkley RJ, Baker PR, Medzihradszky KF, Lynn AJ, Burlingame AL (2008) In-depth analysis of tandem mass spectrometry data from disparate instrument types. *Mol Cell Proteomics* 7:2386–2398.
47. Munton RP, et al. (2007) Qualitative and quantitative analyses of protein phosphorylation in naive and stimulated mouse synaptosomal preparations. *Mol Cell Proteomics* 6:283–293.
48. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14:1188–1190.