# Exploring the *O*-GlcNAc proteome: Direct identification of *O*-GlcNAc-modified proteins from the brain

Nelly Khidekel*, Scott B. Ficarro†, Eric C. Peters†, and Linda C. Hsieh-Wilson*‡

*Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125; and †Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121

The covalent modification of intracellular proteins by *O*-linked β-*N*-acetylglucosamine (*O*-GlcNAc) is emerging as a crucial regulatory posttranslational modification akin to phosphorylation. Numerous studies point to the significance of *O*-GlcNAc in cellular processes such as nutrient sensing, protein degradation, and gene expression. Despite its importance, the breadth and functional roles of *O*-GlcNAc are only beginning to be elucidated. Advances in our understanding will require the development of new strategies for the detection and study of *O*-GlcNAc-modified proteins *in vivo*. Herein we report the direct, high-throughput analysis of *O*-GlcNAc-glycosylated proteins from the mammalian brain. The proteins were identified by using a chemoenzymatic approach that exploits an engineered galactosyltransferase enzyme to selectively label *O*-GlcNAc proteins with a ketone-biotin tag. The tag permits enrichment of low-abundance *O*-GlcNAc species from complex mixtures and localization of the modification to short amino acid sequences. Using this approach, we discovered 25 *O*-GlcNAc-glycosylated proteins from the brain, including regulatory proteins associated with gene expression, neuronal signaling, and synaptic plasticity. The functional diversity represented by this set of proteins suggests an expanded role for *O*-GlcNAc in regulating neuronal function. Moreover, the chemoenzymatic strategy described here should prove valuable for identifying *O*-GlcNAc-modified proteins in various tissues and facilitate studies of the physiological significance of *O*-GlcNAc across the proteome.

Protein posttranslational modifications (PTMs) represent an important mechanism for the regulation of cellular physiology and function. The covalent addition of phosphate, acetate, carbohydrate, and other chemical groups extends the capabilities of proteins and provides a selective and temporal means of controlling protein function (1, 2). Despite the importance of PTMs, their extent and significance are only beginning to be understood. Our laboratory has been investigating O-linked β-*N*-acetylglucosamine (*O*-GlcNAc) glycosylation, the covalent attachment of β-*N*-acetylglucosamine to serine or threonine residues of proteins (3, 4). Unlike most carbohydrate modifications, *O*-GlcNAc is dynamic and intracellular and, as such, shares common features with protein phosphorylation (3). Nearly 80 proteins bearing the *O*-GlcNAc group have been identified to date, including transcription factors, cytoskeletal proteins, protein kinases, and nuclear pore proteins (4). Recent studies have elucidated diverse roles for the *O*-GlcNAc modification, ranging from nutrient sensing to the regulation of proteasomal degradation and gene silencing (3, 5). Moreover, perturbations in *O*-GlcNAc levels have been associated with disease states such as cancer, Alzheimer's, and diabetes (3, 4).

Several lines of evidence suggest an important role for *O*-GlcNAc in the brain. First, activation of protein kinase A or C pathways leads to reduced levels of *O*-GlcNAc in certain protein fractions from cerebellar neurons (6), suggesting an intriguing, dynamic interplay between the two modifications. Second, *O*-GlcNAc transferase (OGT), the enzyme that catalyzes the modification, is most abundant in the brain and pancreas (7). Interestingly, the activity of OGT appears to be modulated by

complex mechanisms, including differential splicing, interaction with regulatory partners, and regulation via PTMs (7). Finally, a critical role for *O*-GlcNAc in the brain is suggested by its presence on proteins important for neuronal function and pathogenesis such as cAMP-responsive element binding protein (CREB) (8) and β-amyloid precursor protein (APP) (4).

Despite strong evidence of its significance, the *O*-GlcNAc modification has been definitively linked to only a handful of proteins from the brain (9). Efforts to identify proteins have been challenged by the difficulty of detecting the modification *in vivo*. Like many PTMs, *O*-GlcNAc is often dynamic, substoichiometric, and prevalent on low-abundance regulatory proteins. The sugar is both enzymatically and chemically labile, being subject to reversal by cellular glycosidases and facile fragmentation during MS analysis. As with many protein kinases, the lack of a well defined consensus sequence for OGT has precluded the determination of *in vivo* modification sites based on primary sequence alone.

Several powerful methods have been reported for the identification of *O*-GlcNAc-modified proteins. Proteins have been tritium-labeled (10), enriched with lectins or antibodies (11, 12), or chemically tagged by metabolic labeling or BEMAD (β-elimination followed by Michael addition with DTT) (12, 13). However, none of the existing methods is ideally suited to the direct, high-throughput identification of *O*-GlcNAc proteins from tissues or cell lysates. For instance, the tritium methodology is labor intensive and lacks sensitivity, necessitating purification of relatively large amounts of protein. Enrichment of *O*-GlcNAc proteins using antibody or lectin chromatography has not afforded direct observation of *O*-GlcNAc-glycosylated peptides and thus cannot rule out false-positives (12). While metabolic labeling has been demonstrated to identify the known, highly glycosylated protein p62 (13), it has not been applied to map glycosylation sites. Moreover, cellular uptake requirements may limit its broad application to tissues. Although the BEMAD approach can be used to map sites on purified proteins and protein complexes, it is an inherently destructive technique that requires extensive controls to establish whether a peptide contains a phosphate, *O*-GlcNAc, or complex O-linked carbohydrate group (12).

A robust strategy for exploring the *O*-GlcNAc proteome would permit investigations into the breadth of the modification and its potential functions across various tissues and species. Direct detection of the *O*-GlcNAc moiety would enable conclusive identification of the glycoproteins and localize the modification to specific functional domains, a prerequisite for understanding the physiological role of the modification. Moreover, such an approach might

also allow for quantitative comparisons of glycosylation levels in cellular or disease states.

We recently reported a chemoenzymatic strategy for the rapid and sensitive detection of purified O-GlcNAc proteins (14). Here, we describe extension of the approach to the direct, high-throughput identification of O-GlcNAc proteins from the mammalian brain. Using this strategy, 25 O-GlcNAc-modified proteins have been identified, including regulatory proteins associated with gene expression, neuronal signaling, and synaptic plasticity. The diversity represented by this set of proteins provides insight into the role of O-GlcNAc in neuronal function.

## Materials and Methods

**Chemoenzymatic Labeling, Biotinylation, and Avidin Enrichment of α-Crystallin.** The A chain of bovine lens α-crystallin (8.7 μg, Sigma-Aldrich) was incubated with the unnatural UDP substrate (14) (750 μM), and Y289L GalT (15) in 20 mM Hepes (pH 7.9) containing 5 mM MnCl$_2$ and 100 mM NaCl for 12 h at 4°C. The reactions were then diluted 2-fold with saturated urea, 2.7 M NaOAc (pH 3.9) (50 mM final concentration, pH 4.8), and N-(aminoxyacetyl)-N′-(D-biotinoyl) hydrazine (5 mM final concentration, Dojindo, Gaithersburg, MD), and incubated with gentle shaking for 20–24 h at 23°C. The tagged αA-crystallin was excised from a Coomassie-stained gel and digested with trypsin (Promega) essentially as described by Shevchenko et al. (16). Avidin affinity chromatography and liquid chromatography (LC)–tandem MS (MS/MS) analysis were performed as described below.

**Preparation of Rat Forebrain Extracts.** The forebrains of Sprague–Dawley rats (Charles River Laboratories) were dissected on ice, lysed into 10 vol of homogenization buffer, and fractionated into nuclear and S100 cytoplasmic components as described by Dignam et al. (17), except that protease inhibitors, phosphatase inhibitors, and a hexosaminidase inhibitor (50 mM GlcNAc) (18) were added to the buffers. Before labeling, the extracts were dialyzed into 20 mM Hepes (pH 7.3), 0.1 M KCl, 0.2 mM EDTA, 0.2% Triton X-100, and 10% glycerol.

**Chemoenzymatic Labeling of Cellular Extracts.** Extract (1–10 mg; 1–3 mg/ml) was incubated with 5 mM MnCl$_2$, 1.25 mM ADP, 0.5 mM unnatural UDP substrate, and Y289L GalT (25 ng/μl) for 12–14 h at 4°C. After enzymatic labeling, extracts were dialyzed into denaturing buffer (5 M urea/50 mM NH$_4$HCO$_3$/100 mM NaCl, pH 7.8; 3 × 2 h). The pH was adjusted with 2.7 M NaOAc (pH 3.9) (final concentration 50 mM, pH 4.8). Aminoxy biotin (2.75 mM) was added, and the reactions were incubated as described for αA-crystallin. Extracts were diluted with 3 M NH$_4$HCO$_3$ (pH 9.6) (50 mM final concentration, pH 8) and dialyzed (1 × 2 h, 1 × 10 h) into 6 M urea, 50 mM NH$_4$HCO$_3$ (pH 7.8), and 100 mM NaCl, followed by either denaturing buffer (4 M urea/50 mM NH$_4$HCO$_3$/10 mM NaCl, pH 7.8) or nondenaturing buffer (50 mM NH$_4$HCO$_3$/10 mM NaCl, pH 7.8).

**Proteolytic Digestion and Cation Exchange/Avidin Affinity Chromatography.** Nondenatured extracts from the previous step were concentrated and denatured/reduced as described in the isotope-coded affinity tag protocol from Applied Biosystems. Proteins were then alkylated with 15 mM iodoacetamide for 45 min in the dark, diluted to 0.04% SDS with 50 mM NH$_4$HCO$_3$ (pH 7.8), and digested with trypsin or GluC (20–30 ng/μl) for 12–14 h at 37°C. Urea-denatured extracts were diluted with 50 mM NH$_4$HCO$_3$ (pH 7.8) after the reduction (10 min) and alkylation steps, and subjected to protease digestion as described above.

Proteolytic digests conducted in the presence of urea were desalted with peptide macrotrap cartridges (Michrom Bioresources, Auburn, CA). Digests conducted without urea were acidified with 1% aqueous trifluoroacetic acid and diluted into cation exchange load buffer (Applied Biosystems). Cation exchange chromatography was performed on 1–3 mg of lysate as described by the manufacturer, except that peptides were eluted with a step gradient of 40, 100, 200, and 350 mM KCl in 5 mM KH$_2$PO$_4$ containing 25% CH$_3$CN. Fractionated peptides were enriched by avidin chromatography (Applied Biosystems) as described by the manufacturer except that the washes were tripled in volume.

**β-Elimination/Michael Addition of Avidin-Purified Peptides.** After avidin chromatography, a portion of the S100 lysate fraction (40 mM KCl elution) was subjected to β-elimination/Michael addition (12) by using 25 mM butanethiol, and reactions were stopped with AcOH.

**LC-MS Analysis of Avidin-Enriched Biotinylated Peptides.** Automated nanoscale RP-HPLC/electrospray ionization (ESI)/MS was performed with an HPLC pump, autosampler (Agilent Technologies, Palo Alto, CA), and linear ion trap mass spectrometer (Thermo Electron, San Jose, CA) with a variation of the "vented column" approach described by Licklider et al. (19). For data-dependent experiments, the mass spectrometer was programmed to record a full-scan ESI mass spectrum ($m/z$ 500–2,000) followed by five data-dependent MS/MS scans (relative collision energy = 35%; 3.5-Da isolation window). Precursor ion masses for candidate glycosylated peptides were identified by a computer algorithm (Charge Loss Scanner; developed in-house with VISUAL BASIC 6.0) that inspected product ion spectra for peaks corresponding to losses of the ketone-biotin and ketone-biotin-GlcNAc moieties. Up to eight candidate peptides at a time were analyzed in subsequent targeted MS[4] experiments to derive sequence information. See Fig. 4, which is published as supporting information on the PNAS web site, for additional information on LC-MS and database querying.

## Results

**Approach Toward the Direct, High-Throughput Identification of O-GlcNAc-Modified Proteins.** Previously, we reported a chemoenzymatic strategy for the detection of purified O-GlcNAc-glycosylated proteins (14). Our approach took advantage of an engineered β-1,4-galactosyltransferase (GalT) enzyme to transfer a ketone-containing galactose analogue selectively to the C-4 hydroxyl of GlcNAc. Once transferred, the ketone functionality was reacted with an aminooxy biotin nucleophile, permitting the rapid, chemiluminescence detection of the O-GlcNAc-modified proteins. We reasoned that this biotin tagging approach could be extended to enrich O-GlcNAc-glycosylated species from complex mixtures (Fig. 1). Previous studies have demonstrated the importance of enrichment strategies for the detection of PTMs (20). In our case, proteins from cellular lysates would be selectively labeled with the ketone-biotin handle and proteolytically digested, and the glycopeptides would be captured by using avidin affinity chromatography. Mass spectrometric analysis of the enriched glycopeptides would afford the proteomewide identification of novel glycosylated proteins. Importantly, the approach would also permit the direct detection of modified peptides, enabling simultaneous mapping of O-GlcNAc to specific functional domains within a protein.

**Application of the Strategy to the A Chain of Bovine α-Crystallin.** We first demonstrated that O-GlcNAc-modified peptides could be selectively enriched by using peptide mixtures from αA-crystallin. αA-Crystallin contains one major site of glycosylation with an estimated stoichiometry of 10% (21). As such, it has proven to be a challenging target for MS analysis, requiring sophisticated quadrupole time-of-flight instrumentation (21) or in-line lectin affinity chromatography (22). αA-Crystallin was enzymatically labeled with the ketone functionality and chemically reacted with an aminooxy biotin derivative. After tryptic digestion and avidin chromatography, enrichment of the expected glycosylated species was observed (Fig. 2). LC-MS analysis indicated a peak corresponding to the mass of the O-GlcNAc-modified peptide [158]AIPVSREEKPSSAPSS[173]
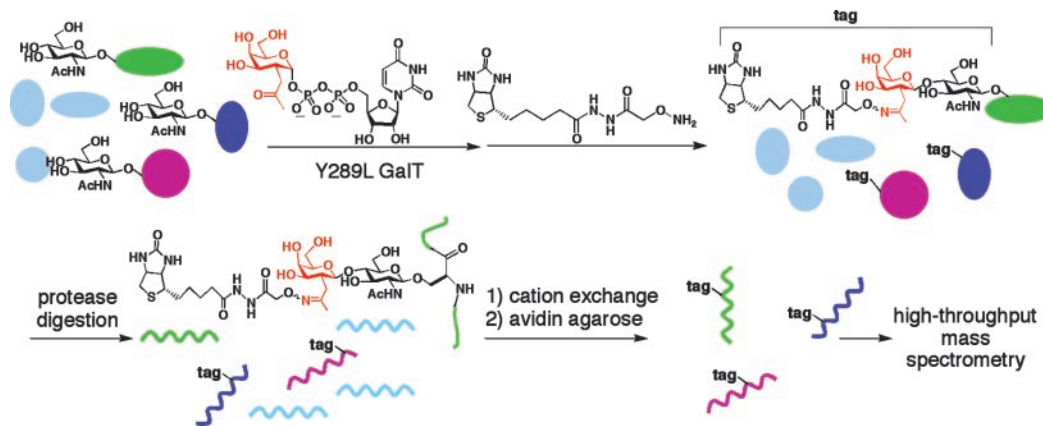
**Fig. 1.** Chemoenzymatic strategy for identifying *O*-GlcNAc-glycosylated proteins from cellular lysates.

labeled with the ketone-biotin tag ($m/z$ 787.86). Notably, the ketone-biotin moiety produced a unique fragmentation pattern upon collision-induced dissociation (CID), which provided unambiguous indication of an *O*-GlcNAc containing peptide. Specifi-



**Fig. 2.** Application of the strategy toward αA-crystallin. (*a*) MS analysis revealed the tagged *O*-GlcNAc peptide [158]AIPVSREEKPSSAPSS[173] ($m/z$ 787.86). The tag provided a diagnostic signature by MS/MS. The MS/MS spectrum of the triply charged precursor ion revealed the signature loss of the ketone-biotin moiety to yield the doubly charged GlcNAc-modified peptide ($m/z$ 922.85) as the predominant species. MS[3] analysis revealed the loss of the GlcNAc moiety to yield the unmodified peptide ($m/z$ 821.64) and several y and b fragment ions containing the GlcNAc moiety that were used to establish the glycosylation site as Ser-162. Glycosylated y and b ions are indicated with the subscript G. MS[4] analysis generated additional y and b ions as well as several internal fragment ions that were used to sequence the peptide. Note that the loss of the ketone-biotin moiety is associated with a loss of charge; therefore the major $m/z$ value of the fragment ion in the MS[2] spectrum is greater than that of the precursor ion observed by MS. (*b*) Summary of the y and b fragment ions.

cally, predominant loss of the ketone-biotin moiety (515.3 Da) was readily observed upon CID, followed by subsequent loss of the GlcNAc group (203.1 Da) during MS[3] experiments. MS[3] analysis localized the GlcNAc moiety on the peptide to the known glycosylation site, Ser-162 (21), and higher-order MS analysis afforded sequence confirmation of the peptide (Fig. 5 which is published as supporting information on the PNAS web site).
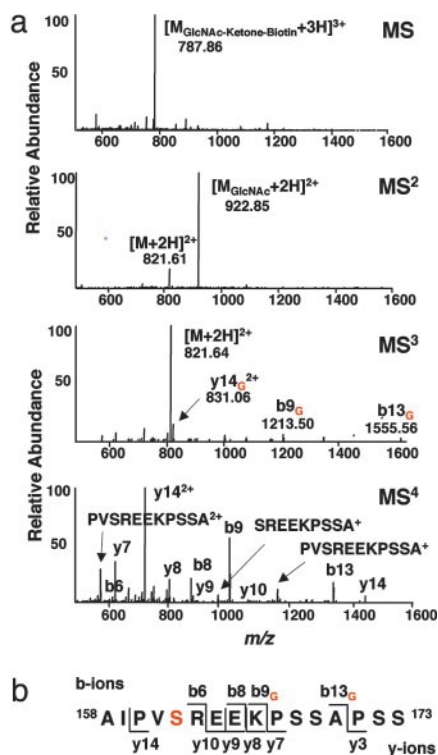
**Exploration of the *O*-GlcNAc Proteome of the Brain.** Having demonstrated the tagging and capture of an *O*-GlcNAc-glycosylated peptide from αA-crystallin, we applied the approach to the *O*-GlcNAc proteome of the mammalian brain. Rat brain lysates were separated into nuclear and S100 cytoplasmic fractions, labeled with the tag, and digested with trypsin. We also subjected a portion of the samples to proteolytic digestion with GluC to broaden the scope of analysis and generate confirmatory peptide sequences. Because of the overall complexity of the sample, the digested peptides were fractionated via strong cation exchange chromatography before avidin affinity chromatography.

Nearly 100 peptides containing the characteristic signature loss of the ketone-biotin tag were observed by LC-MS/MS. Fig. 3*a* shows an averaged electrospray ionization mass spectrum of ions eluting from the LC column with retention time of 17.0–18.1 min. Peaks corresponding to peptides with the diagnostic signature were subsequently selected for targeted MS[4] analysis. Notably, the vast majority of peaks in this region contained the GlcNAc-ketone-biotin moiety, demonstrating significant enrichment of modified peptides. The MS/MS spectrum of a representative peptide ($m/z$ = 789.2) (Fig. 3*b*), illustrates the characteristic loss of a ketone-biotin moiety ($m/z$ = 925.5) and GlcNAc-ketone-biotin moiety ($m/z$ 823.9). Higher-order MS analysis generated a definitive series of b and y ions (Fig. 3*c*), and database searching identified the peptide as belonging to the protein synaptopodin. Notably, alternative MS instrumentation and techniques such as a quadrupole time-of-flight mass spectrometer (21) can be used to obtain sequencing information of species exhibiting the characteristic loss signature.

Using this approach, we successfully sequenced 34 unique peptides corresponding to 25 proteins from rat brain (Table 1). Importantly, two of the proteins, microtubule-associated protein (MAP) 2B and host cell factor (HCF) have been reported to be *O*-GlcNAc-glycosylated (23, 24), providing strong validation of our methodology. In addition, our results extend earlier reports by establishing distinct amino acid stretches within each protein that bear the modification. Two sites of glycosylation were identified in the N-terminal region of MAP2B. In accordance with a demonstrated interaction between the N-terminal region of HCF and both a GlcNAc-specific antibody and lectin (24), we observed four distinct sites within three peptides in the N-terminal region of HCF.
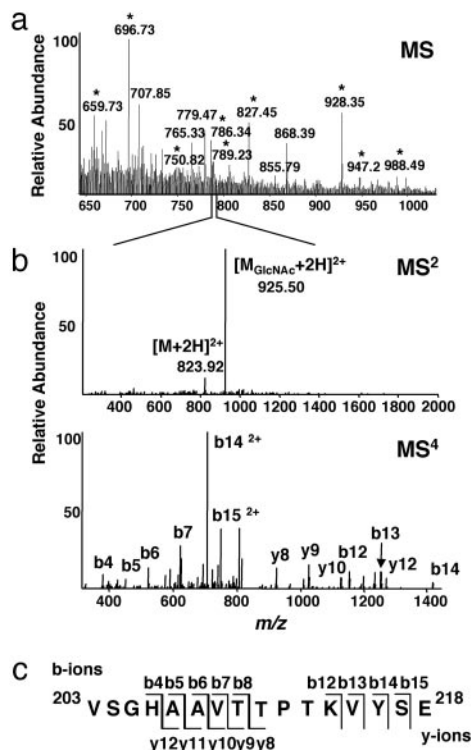
**Fig. 3.** Analysis of tagged *O*-GlcNAc peptides from brain lysates. (*a*) Summed *m/z* spectrum of ions eluting from the LC column with retention time 17.0–18.1 min. Peaks indicated by * represent peptides that yielded the diagnostic ketone-biotin and GlcNAc-ketone-biotin loss signature upon MS/MS. (*b*) MS/MS spectrum of a representative peak (*m/z* = 789.23), showing loss of a ketone-biotin moiety (*m/z* = 925.50) and GlcNAc-ketone-biotin moiety (*m/z* 823.92). Fragmentation during MS⁴ analysis yielded numerous b and y ions, which permitted sequencing of the peptide. (*c*) Prominent fragment ions used to identify the peptide as ²⁰³VSGHAAVTTPTKVYSE²¹⁸ from synaptopodin.

We also identified *O*-GlcNAc on erythrocyte protein band 4.1-like 3 within a region that shares significant sequence identity to a reported glycopeptide from human erythrocyte membrane protein band 4.1 (25).

In addition to known proteins, our approach enabled the identification of 23 additional *O*-GlcNAc-glycosylated proteins from the mammalian brain (Table 1). The proteins fall into a broad range of functional classes (26), including those involved in neuronal signaling, transcriptional regulation, and synaptic plasticity (Fig. 6, which is published as supporting information on the PNAS web site). Consistent with studies demonstrating the presence of *O*-GlcNAc on transcription factors and RNA polymerase II, we identified a large number of proteins involved in transcription. In addition to low-abundance transcription factors, we found *O*-GlcNAc on transcriptional coactivators, corepressors, and chromatin remodeling enzymes, which suggests expanded roles for *O*-GlcNAc in transcriptional control.

Notably, our methodology also afforded the simultaneous detection of multiple PTMs. For instance, we observed an *O*-GlcNAc-modified peptide with a characteristic loss of 98 Da upon collision-induced dissociation, consistent with phosphorylation within the same peptide (Fig. 7, which is published as supporting information on the PNAS web site). Moreover, two *O*-GlcNAc modifications were identified within the same peptide of HCF.

**Merging the Technology with β-Elimination Strategies to Map Glycosylation Sites.** The mapping of specific *O*-GlcNAc glycosylation sites is inherently difficult due to the lability of the glycosidic linkage upon collision-induced dissociation and the preference of OGT for sequences rich in serine, threonine, and proline residues. Although we successfully narrowed the sites of *O*-GlcNAc glycosylation to short amino acid sequences, the features noted above limited our ability to perform site identification on all but a few sequences. To address this issue, we combined precedented β-elimination strategies with our methodology to localize specific modification sites. Previous studies have shown that glycosylated and phosphorylated serine/threonine residues as well as carboxyamido-modified cysteine residues undergo β-elimination to form dehydroalanine/β-methyldehydroalanine under strong alkaline conditions (12, 27). Subsequent Michael addition of a thiol nucleophile generates a stable sulfide adduct. We first labeled S100 cytoplasmic lysates with our ketone-biotin tag and enriched the *O*-GlcNAc glycopeptides by using avidin chromatography as described. One of the enriched fractions was then selected for β-elimination, followed by butane-thiol addition (Figs. 8 and 9, which are published as supporting information on the PNAS web site). MS/MS analysis of the resultant peptides permitted localization of the glycosylation site on HIV-1 Rev binding protein from seven possible residues within the peptide ³⁶⁰APVGSVVSVPSHSSASSDK³⁷⁸ to Ser-372 or Ser-373. Notably, MS/MS analysis before β-elimination conclusively demonstrated that the original peptide was *O*-GlcNAc-glycosylated, rather than phosphorylated or modified with a complex carbohydrate. With further refinement of the β-elimination methodology toward complex mixtures, we anticipate that the combined ketone-labeling and β-elimination approaches will be a powerful tool for identifying specific *O*-GlcNAc modification sites.

## Discussion

Herein we report a direct, high-throughput analysis of *O*-GlcNAc glycosylated proteins from the mammalian brain. The proteins were identified by using a chemoenzymatic approach that exploits an engineered galactosyltransferase enzyme to selectively label GlcNAc proteins with a ketone-biotin tag. The tag provides both a straightforward means to enrich low-abundance *O*-GlcNAc peptides from complex mixtures and a unique signature upon MS/MS for unambiguous identification of the *O*-GlcNAc-glycosylated species. In contrast to reported antibody, lectin, and metabolic labeling methods (11–13), the strategy provides direct evidence of *O*-GlcNAc glycosylation and permits mapping of modification sites to short amino acid sequences. The ability to localize *O*-GlcNAc is essential for surveying its distribution across the proteome and understanding its functional significance on a given protein or family of proteins.

An exciting feature of the approach is its potential to explore the interplay among PTMs (2, 28). In this study, we identified two peptides that contained more than one PTM. For instance, the N-terminal domain of HCF showed two *O*-GlcNAc moieties within the same peptide, and a second peptide exhibited evidence of both phosphorylation and glycosylation. Notably, all *O*-GlcNAc proteins known to date are phosphoproteins, and increasing evidence suggests that glycosylation functionally antagonizes phosphorylation in many cases (3, 8). The approach reported herein involves a non-destructive technique that does not require the removal of other PTMs to study *O*-GlcNAc. As such, the strategy should permit a direct examination of whether specific glycosylation and phosphorylation events are mutually exclusive *in vivo*, as suggested for the C-terminal domain of RNA polymerase II (29), or whether the two modifications coexist, as recently reported for the transcription factor signal transducer and activator of transcription 5 (Stat5) (30). Thus, the strategy is complementary to top-down MS approaches that can be used to simultaneously interrogate multiple PTMs from intact proteins (31).

The chemoenzymatic approach can also be combined with existing β-elimination strategies, providing a powerful tool to identify precise sites of glycosylation. Notably, emerging MS techniques such as electron transfer dissociation, which has been successfully used to map phosphorylation sites, could also be

**Table 1. *O*-GlcNAc glycosylated proteins from the mammalian brain**

| Protein | NCBI entry | Function | Peptide sequence | Residues |
|---|---|---|---|---|
| **Transcriptional regulation** | | | | |
| Sox2 (sry-related high mobility group box 2) | 31543759* | Transcription factor | SEASSSPPVVTSSSHSR | 248–264 |
| ATF-2 | 13591926 | Transcription factor, histone acetyltransferase | AALTQQHPPVTDGDTVK | 262–278 |
| HCF | 34881756 | Transcriptional regulator, chromatin associated factor | TAAAQVGTSVSSAANTSTRPIITVHK[†] | 620–645 |
| HCF | 34881756 | Transcriptional regulator, chromatin associated factor | VMSVVQTK / SPITIITTK | 691–698 / 802–810 |
| SRC-1 (steroid receptor coactivator 1) | 34863079 | Transcriptional coactivator for nuclear receptors | INPSVNPGISPAHGVTR | 188–204 |
| CCR4-NOT4 | 34855140 | Global transcriptional regulator, mRNA metabolism | SNPVIPISSSNHSAR | 329–343 |
| CCR4-NOT subunit 2 | 34864872 | Global transcriptional regulator, mRNA metabolism | SLSQGTQLPSHVTPTTGVPTMSLHTPPSPSR | 79–109 |
| TLE-4 (transducin-like enhancer protein 4) | 9507191 | Transcriptional corepressor | TDAPTPGSNSTPGLRPVPGKPPGVDPLASSLR | 298–329 |
| RNA-binding motif protein 14 | 16307494* | Transcriptional coregulator for steroid receptors | AQPSVSLGAAYR | 239–250 |
| **Nucleic acid-binding proteins** | | | | |
| NFR-κB (nuclear factor-related κB) | 34862978 | DNA binding protein | VPVTATQTK | 896–904 |
| Zinc finger RNA-binding protein | 34854400 | RNA binding protein | AGYSQGATQYTQAQQAR | 58–74 |
| **Intracellular transport** | | | | |
| Hrb (HIV-1 Rev-binding protein) | 34859394 | RNA trafficking | APVGSVVSVPSHSSASSDK[‡] | 360–378 |
| GRASP55 (Golgi reassembly stacking protein 2) | 20301956 | Membrane protein transport, Golgi cisternae stacking | VPTTVEDR | 423–430 |
| **Cellular organization/dynamics** | | | | |
| Erythrocyte protein band 4.1-like 3 | 16758808 | Cytoskeletal protein | TITSETTSTTTTTHITK / TTSTTTTTHITKTVKGGISE | 1026–1042 / 1031–1050 |
| Erythrocyte protein band 4.1-like 1, isoform L | 11067407 | Cytoskeletal protein | DVLTSTYGATAETLSTSTTTHVTK | 1460–1483 |
| Erythrocyte protein band 4.1-like 1, isoform L | 11067407 | Cytoskeletal protein | TLSTSTTTHVTKTVKGGFSE | 1472–1491 |
| Spectrin beta chain (fodrin beta chain) | 34879632 | Axonal/pre-synaptic cytoskeletal protein | HDTSASTQSTPASSR | 2354–2368 |
| MAP1B | 19856246 | Axonogenesis | TTTKTTRSPDTSAYCYE | 2018–2034 |
| MAP2B | 111965 | Dynamic assembly of microtubules at dendrites | SSKDEEPQKDKADKVADVPVSE | 366–387 |
| MAP2B | 111965 | Dynamic assembly of microtubules at dendrites | KADKVADVPVSE / TSSESPFPAKE | 376–387 / 788–798 |
| **Cellular communication/signal transduction** | | | | |
| WNK-1 (lysine deficient protein kinase) | 16758634 | Signal transduction, ion homeostasis | DGTEVHVTASSSGAGVVK / MGGSTPISAASATSLGHFTK | 1584–1601 / 2043–2062 |
| PDZ-GEF | 34857578 | Guanine nucleotide exchange factor for RAP1/2 | ISSRSSIVSNSSFDSVPVSLHDE | 1211–1233 |
| PDZ-GEF | 34857578 | Guanine nucleotide exchange factor for RAP1/2 | SSFDSVPVSLHDER / SVPVSLHDE | 1221–1234 / 1225–1233 |
| Synaptopodin | 11067429 | Dendritic spine formation | VSGHAAVTTPTKVYSE | 203–218 |
| Bassoon | 9506427 | Synaptic vesicle cycling | VTQHFAK[§] | 1338–1444 |
| **Uncharacterized proteins** | | | | |
| Hypothetical protein FLJ31657 | 34855501 | Unknown | IGGDLTAAVTK | 196–206 |
| 1300019H17RIKEN protein | 34880180 | Unknown | EAALPSTK | 286–293 |
| KIAA1007 protein | 34851212 | Unknown | TVTVTKPTGVSFK | 1051–1063 |
| DACA-1 homolog | 34861007 | Unknown | IGDVTTSAVK | 271–280 |

*Mouse proteins identified in the National Center for Biotechnology Information (NCBI) database. Corresponding rat orthologs were identified in the Celera database.

[†]We identified two distinct sites of *O*-GlcNAc glycosylation on this peptide.

[‡]The site of modification was localized to Ser-372 or Ser-373 by using a combination of chemoenzymatic tagging and β-elimination.

[§]Confirmed by peptide synthesis and MS sequencing analysis (see Fig. 10, which is published as supporting information on the PNAS web site).

combined with our methodology to directly identify glycosylation sites and abrogate the need for β-elimination (32).

In this work, we demonstrated the power of our approach by identifying 25 *O*-GlcNAc-glycosylated proteins from the mammalian brain. Over the last 20 years, the *O*-GlcNAc modification has been established on ≈80 proteins (4). Thus, our findings represent a significant expansion in the number of known *O*-GlcNAc proteins, and they provide insights into the breadth of the modification and its potential functions in the brain.

Consistent with previous studies demonstrating an important role for *O*-GlcNAc in transcriptional regulation, we identified two unique transcription factors, sry-related high mobility group box 2 (Sox2) and activating transcription factor 2 (ATF-2). Sox2 is a member of the high mobility group box superfamily of minor groove DNA-binding proteins (33), proteins believed to govern cell fate decisions during diverse developmental processes. Although primarily known for its role in embryogenesis, Sox2 has also been detected in the adult central nervous system (34). ATF-2 is a

transcription factor that is enriched in the brain (35), and possesses an intrinsic histone acetyltransferase (HAT) activity required for activating transcription (36). As *O*-GlcNAc has been implicated in nutrient sensing and the development of insulin-resistant diabetes (3, 4), it is interesting that ATF-2 appears to play multiple roles in glucose homeostasis. For instance, ATF-2 has been shown to up-regulate transcription from the insulin promoter in human pancreatic $\beta$ cells in a $Ca^{2+}$/calmodulin-dependent protein kinase IV-dependent manner (37). Moreover, recent studies indicate that ATF-2 activates the gluconeogenic gene phosphoenolpyruvate carboxykinase in HepG2 hepatic cells upon retinoic acid induction (38). Notably, the region of glycosylation lies in a proline-rich stretch near a motif essential for the HAT activity of ATF-2. Phosphorylation in the N-terminal transactivation domain of ATF-2 (Thr-69 and Thr-71) up-regulates its HAT activity (36). It will be important to examine in this instance whether glycosylation and phosphorylation act in opposition.

Although transcription factors and RNA polymerase II have been shown to be glycosylated, other important elements of the transcriptional machinery have not been well documented. In this study, we demonstrated *O*-GlcNAc on unique transcriptional proteins, including coactivators and corepressors. This finding suggests broader roles for *O*-GlcNAc in regulating transcription than previously recognized. For instance, we found the modification on two proteins (including a ubiquitin ligase) in the carbon catabolite repression 4-negative on TATA-less (CCR4-NOT), a large protein complex involved in mRNA metabolism and the global control of gene expression (39). In addition, *O*-GlcNAc was identified on steroid receptor coactivator-1 (SRC-1), a protein involved in chromatin remodeling that functions as a transcriptional coactivator for estrogen, thyroid, and other nuclear receptors (40). Finally, *O*-GlcNAc was found on HCF, a chromatin-associated factor that interacts with both OGT and the Sin3A histone deacetylase complex *in vivo* (24). Mammalian Sin3A interacts with OGT and thereby synergistically represses transcription from both basal and Sp-1-driven promoters (41). Here, we identified four distinct sites of glycosylation within the N-terminal domain of HCF, a region required for interaction with both OGT and Sin3A (24). It will be interesting to examine the functional impact of HCF glycosylation on its binding to Sin3A and OGT and on gene silencing.

Importantly, our results demonstrate that a number of proteins involved in neuronal signaling and synaptic function are the targets of *O*-GlcNAc glycosylation. For instance, we identified the modification of PDZ-GEF, a guanine nucleotide exchange factor that activates the Ras-related GTPases Rap1 and Rap2 (42). PDZ-GEF contains a PDZ domain, a protein-interacting module often involved in the assembly of signal transduction complexes at the synapse (43). Another *O*-GlcNAc protein is WNK-1 (With No Lysine K), a serine/threonine protein kinase whose activation has been linked to ion transport and hypertension (44). Moreover, we identified two brain-enriched proteins important for synaptic function, synaptopodin and bassoon. The actin-associated protein synaptopodin is essential for dendritic spine formation, with synaptopodin-deficient mice exhibiting a lack of spine apparatuses and impaired long-term potentiation and spatial learning (45). Bassoon, a scaffolding protein of the cytomatrix assembled at the active zone, plays a critical role in synaptic vesicle cycling (46). Taken together, these findings reveal that *O*-GlcNAc glycosylation likely plays critical roles in neuronal communication and synaptic function.

In summary, we demonstrate a chemoenzymatic strategy for the high-throughput identification of *O*-GlcNAc-glycosylated proteins from the mammalian brain. The approach permits the enrichment and direct identification of *O*-GlcNAc-glycosylated peptides from complex mixtures and can be combined with existing technologies to map specific glycosylation sites. The generality of the method should enable explorations of the *O*-GlcNAc proteome in any cell type or tissue. Moreover, studies of the dynamic interplay among PTMs and future extension of the methodology to quantitative proteomics should be possible. Using the approach, we discovered 25 *O*-GlcNAc-glycosylated proteins from the brain, including regulatory proteins associated with gene expression, neuronal signaling, and synaptic plasticity. The functional diversity represented by this set of proteins suggests an expanded role for *O*-GlcNAc in regulating neuronal function. We anticipate that further investigations of the proteins identified in this study, coupled with the continued development of chemical tools, will provide insights into the physiological importance of this posttranslational modification.

1. Greengard, P. (2001) *Science* **294,** 1024–1030.
2. Fischle, W., Wang, Y. & Allis, C. D. (2003) *Nature* **425,** 475–479.
3. Slawson, C. & Hart, G. W. (2003) *Curr. Opin. Struct. Biol.* **13,** 631–636.
4. Whelan, S. A. & Hart, G. W. (2003) *Circ. Res.* **93,** 1047–1058.
5. Zhang, F., Su, K., Yang, X., Bowe, D. B., Paterson, A. J. & Kudlow, J. E. (2003) *Cell* **115,** 715–725.
6. Griffith, L. S. & Schmitz, B. (1999) *Eur. J. Biochem.* **262,** 824–831.
7. Iyer, S. P. N. & Hart, G. W. (2003) *Biochemistry* **42,** 2493–2499.
8. Lamarre-Vincent, N. & Hsieh-Wilson, L. C. (2003) *J. Am. Chem. Soc.* **125,** 6612–6613.
9. Cole, R. N. & Hart, G. W. (2001) *J. Neurochem.* **79,** 1080–1089.
10. Roquemore, E. P., Chou, T. Y. & Hart, G. W. (1994) *Methods Enzymol.* **230,** 443–460.
11. Cieniewski-Bernard, C., Bastide, B., Lefebvre, T., Lemoine, J., Mounier, Y. & Michalski, J. C. (2004) *Mol. Cell. Proteomics* **3,** 577–585.
12. Wells, L., Vosseller, K., Cole, R. N., Cronshaw, J. M., Matunis, M. J. & Hart, G. W. (2002) *Mol. Cell. Proteomics* **1,** 791–804.
13. Vocadlo, D. J., Hang, H. C., Kim, E. J., Hanover, J. A. & Bertozzi, C. R. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 9116–9121.
14. Khidekel, N., Arndt, S., Lamarre-Vincent, N., Lippert, A., Poulin-Kerstien, K. G., Ramakrishnan, B., Qasba, P. K. & Hsieh-Wilson, L. C. (2003) *J. Am. Chem. Soc.* **125,** 16162–16163.
15. Ramakrishnan, B. & Qasba, P. K. (2002) *J. Biol. Chem.* **277,** 20833–20839.
16. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. (1996) *Anal. Chem.* **68,** 850–858.
17. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **11,** 1475–1489.
18. Chou, T. Y., Hart, G. W. & Dang, C. V. (1995) *J. Biol. Chem.* **270,** 18961–18965.
19. Licklider, L. J., Thoreen, C. C., Peng, J. & Gygi, S. P. (2002) *Anal. Chem.* **74,** 3076–3083.
20. Kalume, D. E., Molina, H. & Pandey, A. (2003) *Curr. Opin. Chem. Biol.* **7,** 64–69.
21. Chalkley, R. J. & Burlingame, A. L. (2001) *J. Am. Soc. Mass Spectrom.* **12,** 1106–1113.
22. Haynes, P. A. & Aebersold, R. (2000) *Anal. Chem.* **72,** 5402–5410.
23. Ding, M. & Vandre, D. D. (1996) *J. Biol. Chem.* **271,** 12555–12561.
24. Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N. & Herr, W. (2003) *Genes Dev.* **17,** 896–911.
25. Inaba, M. & Maede, Y. (1989) *J. Biol. Chem.* **264,** 18149–18155.
26. Schoof, H., Zaccaria, P., Gundlach, H., Lemcke, K., Rudd, S., Kolesov, G., Arnold, R., Mewes, H. W. & Mayer, K. F. (2002) *Nucleic Acids Res.* **30,** 91–93.
27. Oda, Y., Nagasu, T. & Chait, B. T. (2001) *Nat. Biotechnol.* **19,** 379–382.
28. Khidekel, N. & Hsieh-Wilson, L. C. (2004) *Org. Biomol. Chem.* **2,** 1–7.
29. Kelly, W. G., Dahmus, M. E. & Hart, G. W. (1993) *J. Biol. Chem.* **268,** 10416–10424.
30. Gewinner, C., Hart, G., Zachara, N., Cole, R., Beisenherz-Huss, C. & Groner, B. (2004) *J. Biol. Chem.* **279,** 3563–3572.
31. Kelleher, N. L. (2004) *Anal. Chem.* **76,** 197A–203A.
32. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 9528–9533.
33. Pevny, L. H. & Lovell-Badge, R. (1997) *Curr. Opin. Genet. Dev.* **7,** 338–344.
34. Gure, A. O., Stockert, E., Scanlan, M. J., Keresztes, R. S., Jager, D., Altorki, N. K., Old, L. J. & Chen, Y. T. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 4198–4203.
35. Herdegen, T. & Leah, J. D. (1998) *Brain Res. Brain Res. Rev.* **28,** 370–490.
36. Kawasaki, H., Schiltz, L., Chiu, R., Itakura, K., Taira, K., Nakatani, Y. & Yokoyama, K. K. (2000) *Nature* **405,** 195–200.
37. Ban, N., Yamada, Y., Someya, Y., Ihara, Y., Adachi, T., Kubota, A., Watanabe, R., Kuroe, A., Inada, A., Miyawaki, K., *et al.* (2000) *Diabetes* **49,** 1142–1148.
38. Lee, M. Y., Jung, C. H., Lee, K., Choi, Y. H., Hong, S. & Cheong, J. (2002) *Diabetes* **51,** 3400–3407.
39. Collart, M. A. (2003) *Gene* **313,** 1–16.
40. Xu, J. & Li, Q. (2003) *Mol. Endocrinol.* **17,** 1681–1692.
41. Yang, X., Zhang, F. & Kudlow, J. E. (2002) *Cell* **110,** 69–80.
42. Rebhun, J. F., Castro, A. F. & Quilliam, L. A. (2000) *J. Biol. Chem.* **275,** 34901–34908.
43. Zhang, M. & Wang, W. (2003) *Acc. Chem. Res.* **36,** 530–538.
44. Wilson, F. H., Disse-Nicodeme, S., Choate, K. A., Ishikawa, K., Nelson-Williams, C., Desitter, I., Gunel, M., Milford, D. V., Lipkin, G. W., Achard, J. M., *et al.* (2001) *Science* **293,** 1107–1112.
45. Deller, T., Korte, M., Chabanis, S., Drakew, A., Schwegler, H., Stefani, G. G., Zuniga, A., Schwarz, K., Bonhoeffer, T., Zeller, R., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100,** 10494–10499.
46. Altrock, W. D., Tom Dieck, S., Sokolov, M., Meyer, A. C., Sigler, A., Brakebusch, C., Fassler, R., Richter, K., Boeckers, T. M., Potschka, H., *et al.* (2003) *Neuron* **37,** 787–800.

CHEMISTRY

BIOCHEMISTRY