

# Glycomic Approaches to Study GlcNAcylation: Protein Identification, Site-mapping, and Site-specific *O*-GlcNAc Quantitation

Zihao Wang · Gerald W. Hart

Published online: 17 June 2008  
© Humana Press 2008

## Abstract

**Background** *O*-Linked  $\beta$ -*N*-acetylglucosamine (*O*-GlcNAc) is an enzyme-catalyzed posttranslational modification of serine or threonine side chains of nuclear and cytoplasmic proteins. *O*-GlcNAc is present in all metazoans and in viruses that infect eukaryotic cells. GlcNAcylation is dynamic and has a high cycling rate on many proteins in response to cellular metabolism and various environmental stimuli. The rapid cycling of *O*-GlcNAc modulates many biological processes, including transcriptional regulation, stress responses, cell cycle regulation, and protein synthesis and turnover.

**Rationale** Despite the importance of *O*-GlcNAc, progress during the past two decades in this field has been slow. One of the major obstacles is the lack of simple and sensitive tools for efficient *O*-GlcNAc detection and localization. Recently developed *O*-GlcNAc derivatization and enrichment approaches, together with new techniques in mass spectrometric instrumentation and methods, have provided breakthroughs in *O*-GlcNAc site localization and site-specific quantitation. In this review, we will discuss how the current techniques are expanding our knowledge about *O*-GlcNAc proteomics/glycomics and functions.

**Keywords** Glycomics · *O*-GlcNAc · GlcNAcylation · Site-mapping · Chemoenzymatic tagging

## Introduction

Protein GlcNAcylation is distinct from classical “mucin-type” *O*-glycosylation. First, *O*-linked  $\beta$ -*N*-acetylglucosamine (*O*-GlcNAc) is found only on nuclear and cytoplasmic proteins, and not on secreted or cell surface glycoproteins. Second, *O*-GlcNAc is a monosaccharide modification and is generally not further elongated to complex structures. Similar to phosphorylation, GlcNAcylation is ubiquitous and abundant. Since its discovery more than 25 years ago [1, 2], about 600 proteins have been identified as *O*-GlcNAc-modified [3], and this number is growing rapidly with recent advances in *O*-GlcNAc detection. Cycling of *O*-GlcNAc is catalyzed by two unique enzymes: *N*-acetylglucosaminyltransferase (OGT) and *N*-acetyl- $\beta$ -*D*-glucosaminidase (*O*-GlcNAcase). OGT catalyzes the addition of *O*-GlcNAc using uridine 5'-diphosphate-GlcNAc (UDP-GlcNAc) as the donor substrate. UDP-GlcNAc is the major end product of hexosamine biosynthetic pathways and exists within the cells at concentrations ranging from of hundreds of micromolars to millimolars, which approaches the intracellular concentration of ATP [4]. Unlike phosphorylation, for which there are hundreds of site/protein-specific kinases [5], there is only one highly conserved gene encoding the OGT catalytic subunit in mammals and two encoding OGT in plants [6, 7]. Likewise, only a single highly conserved gene encoding *O*-GlcNAcase, the enzyme that removes *O*-GlcNAc, exists in mammals [8], and the plant *O*-GlcNAcase remains to be identified. *O*-GlcNAcase is distinct from lysosomal hexosaminidases in both its subcellular localization and pH optimum. Transient transfection, resulting in overexpression of *O*-GlcNAcase, can lead to dramatic decreases in overall *O*-GlcNAcylation within cells [3].

*O*-GlcNAc is relatively small and uncharged. Therefore, its addition and its removal usually do not affect a protein's

---

Z. Wang · G. W. Hart (✉)  
Department of Biological Chemistry,  
The Johns Hopkins University School of Medicine,  
725 N. Wolfe St.,  
Baltimore, MD 21205-2185, USA  
e-mail: gwhart@jhmi.edu

mobility in gel electrophoresis, even on high-resolution, two-dimensional gels. *O*-GlcNAc was first discovered in experiments designed to enzymatically label terminal GlcNAc on lymphocytic proteins with  $\beta$ 1,4-galactosyltransferase (GalT) and tritiated UDP-galactose (UDP-Gal) [1]. GalT labeling after PNGase F (an enzyme that cleaves *N*-linked carbohydrates) treatment, followed by  $\beta$ -elimination and analyses of the released disaccharide product, provides unambiguous evidence that the target protein is *O*-GlcNAc modified. This method remains the “gold” standard to detect *O*-GlcNAcylated proteins using non-mass spectrometric approaches. This original method also served as the basis for the earliest *O*-GlcNAc site-mapping, which involved time-consuming multiple-step procedures [9–15]. Briefly, a [<sup>3</sup>H] Gal-labeled peptide mixture is first separated by multiple rounds of reverse-phase high-performance liquid chromatography (HPLC), pooling only the radiolabeled fractions. Finally, *O*-GlcNAc site information is obtained by monitoring the release of radioactivity by gas-phase Edman automated sequencing or by manual Edman degradation [13–15]. Robustness of this method is complicated by multiple factors, such as performance of the HPLC purification and the intrinsic limitation of Edman degradation. Moreover, due to substantial sample loss during each HPLC cycle, fairly large amounts of starting material are required (nano- to submicromolar), which is usually not practical for low-abundance regulatory proteins with substoichiometric modification at any single site.

### Proteomic Approaches for Identification of GlcNAcylated Proteins

Besides [<sup>3</sup>H] Gal enzymatic labeling, other important tools for *O*-GlcNAc studies include pan-specific *O*-GlcNAc antibodies CTD 110.6 or RL2 [16, 17], which are able to detect *O*-GlcNAc-modified proteins by Western blotting. A straight-forward method to identify GlcNAcylated proteins is to run identical samples over two separate gels. One gel is used for Western blotting with *O*-GlcNAc antibodies. The other is used for protein staining. Corresponding protein bands/spots are then excised, protease digested, and identified by either mass fingerprint (matrix-assisted laser desorption ionization–time of flight, MALDI-TOF) or liquid chromatography–mass spectrometry (LC-MS/MS). This approach has been used in a number of studies to identify novel *O*-GlcNAc proteins in various biological tissues. The limitation of this approach lies in two aspects. Currently available pan-specific *O*-GlcNAc antibodies have relatively low binding affinity. Thus, in many cases, only multiply modified, higher-molecular-weight, or highly abundant proteins can be detected by the antibodies. In addition, there are chances for nonspecific binding. Sec-

ondly, and most importantly, it is very common that an apparent single-protein band may actually contain multiple proteins. This is true even when high-resolution two-dimensional gels have been employed. Thus, it is possible that a low abundant, highly GlcNAcylated protein is antibody-reactive, while in mass spectrometric analysis, only protein(s) of higher abundance are actually identified. In such cases, this approach will lead to erroneous results. A variation to this approach is to first purify *O*-GlcNAcylated proteins with immunoaffinity chromatography, and then identify the bound proteins after elution [18]. Similarly, due to nonspecific and indirect binding, all of such identifications can only be considered tentative until they are further confirmed by other methods. A simple approach to confirm GlcNAcylation is to immunoprecipitate the target protein, followed by immunoblotting with *O*-GlcNAc antibodies. Antibody specificity controls should also be performed to rule out nonspecific binding in immunoblotting, such as pretreatment of a part of the sample with hexosaminidase and/or by adding free GlcNAc in primary antibody solution to compete specific antibody binding [13–15, 18].

Recently, in addition to the immunoaffinity approaches, GlcNAcylated proteins can also be enriched and purified by chemical tagging. Vocadlo et al. [19] fed cells during cell culture with a peracetylated form of *N*-azidoacetylglucosamine, which is converted into UDP-azido-GlcNAc (UDP-GlcNAz) in vivo by the hexosamine salvage pathway. UDP-GlcNAz is a structural analog of UDP-GlcNAc with the *N*-acyl side chain modified into an azide group. In vitro OGT assays showed that recombinant OGT could tolerate this unnatural substrate and labeled proteins with GlcNAz. The azide moiety then allows facile conjugation to a biotinylated phosphine probe via Staudinger ligation, facilitating selective enrichment of *O*-GlcNAc-modified proteins by avidin chromatography [19]. Nandi et al. [20] adopted this strategy in their glycomic analysis for GlcNAcylated proteins, which they termed the “tagging-via-substrate approach,” and they successfully identified 199 putative *O*-GlcNAc-modified proteins, both known and novel. The advantage of this approach is that it provides superior binding affinity and selectivity over immunoaffinity-based methods for enrichment of *O*-GlcNAc proteins. The high affinity between biotin and streptavidin allows stringent washing and high selectivity, which allows hundreds of proteins to be identified in one simple analysis. However, the low efficiency of free GlcNAz conversion to UDP-GlcNAz and the relatively poor efficiency of UDP-GlcNAz utilization by endogenous OGT as a donor substrate, relative to the natural substrate, limit the utility of the method. More importantly, since OGT’s natural substrate, UDP-GlcNAc, approaches nearly millimolar level in concentration within the cells, most likely only a very small fraction of added UDP-GlcNAz can be actually utilized by OGT. As a consequence, only heavily GlcNAcylated proteins or highly abundant

GlcNAcylated proteins are readily labeled and enriched by this approach. Also, since the hexosamine biosynthesis pathway is closely linked to glucose metabolism, feeding cells with large amount of the GlcNAc analog, GlcNAz, may induce unknown alterations on in vivo metabolism.

### Proteomic/Glycomic Approaches to Map Site of *O*-GlcNAc Modification

An unambiguous alternative to identifying *O*-GlcNAc proteins is to directly map the modification sites. Mapping *O*-GlcNAc sites is essential for functional studies, including site-directed mutagenesis to investigate the role(s) of *O*-GlcNAc in protein function. Until recently, only very limited numbers of *O*-GlcNAc sites have been unambiguously identified [21]. The lack of site information imposes a major obstacle for studies of *O*-GlcNAcylation because it makes it impossible to design site-specific *O*-GlcNAc antibodies, which would provide valuable tools for understanding the regulatory roles of *O*-GlcNAc. Site-specific phosphorylation antibodies have played a key role in the elucidation of phosphorylation's functions in signaling cascades [22].

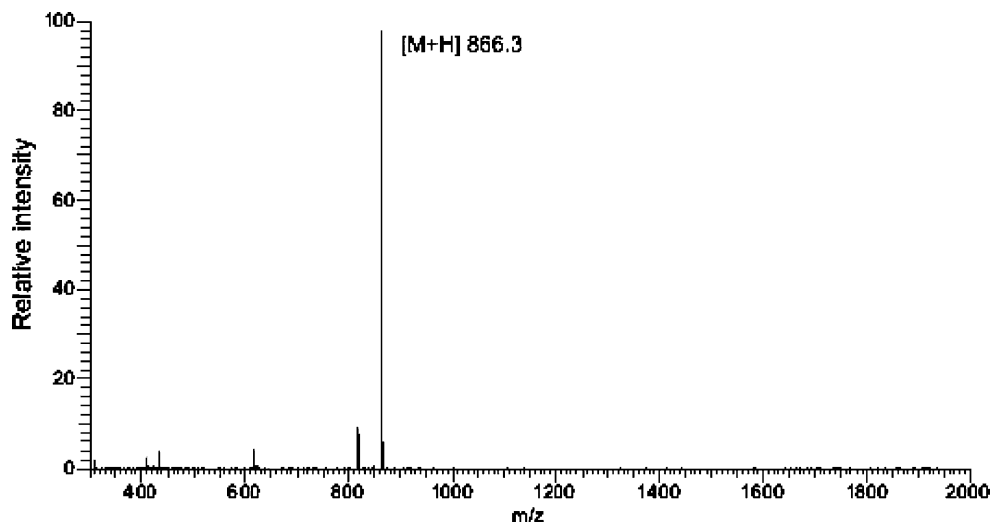
### Mass Spectrometry: *O*-GlcNAc is Labile in the Gas Phase

The difficulties of mapping *O*-GlcNAc lie in three major aspects: its intrinsic lability in gas phase, relatively low abundance, and severe ion suppression effect in the presence of unmodified peptides. When peptides carrying *O*-GlcNAc are fragmented by standard collisionally activated dissociation (CAD) with inert gas, most of the kinetic energy is lost by the cleavage at the *O*-GlcNAc moiety,

which easily falls off and generates a major neutral loss peak. This phenomenon is also well known for some Ser/Thr phosphorylated peptides. However, in case of *O*-GlcNAc, it is even worse because, unlike phosphopeptides, which usually still have certain degrees of backbone fragmentation, there is typically very little or no peptide backbone fragmentation for all *O*-GlcNAc peptides examined to date under standard CAD conditions (Fig. 1). Moreover, *O*-GlcNAc is also lost at the electrospray ionization interface and at the ion optics, presumably due to in-source fragmentation ([11]; data not shown). Of course, further fragmentation of the neutral loss peak in the ion trap ( $MS^3$ ) can achieve identification of the peptide. However, unless there is only one Ser or Thr residue in the peptide, unambiguous site-mapping is not possible because all modification site information has been lost.

Studies by Haynes and Aebersold [23] employed precursor ion scanning with low collision energy on a triple quadrupole instrument to first detect peptides that generated signature peaks corresponding to loss of the *O*-GlcNAc moiety. They then used CAD at higher energy to fragment and sequence the peptides that were initially *O*-GlcNAc-modified. Using this method, they managed to confine one *O*-GlcNAc site on bovine  $\alpha$ -crystallin to a 16-amino-acid carboxyl terminal peptide, which had five serine residues as potential *O*-GlcNAc sites. Chalkley and Burlingame [24] adopted a very similar strategy on a quadrupole time-of-flight instrument. These investigators used lowered collision energy to fragment preselected precursor ions that produced diagnostic GlcNAc oxonium ions under standard CAD condition, thus improving the possibility of generating GlcNAc-carrying fragments. Using this method, Chalkley et al. were able to locate several *O*-GlcNAc sites on synthetic peptides and on serum response factor. However, all these studies had to use rather large amounts of samples (synthetic *O*-GlcNAc peptides or

**Fig. 1** *O*-GlcNAc site information is lost before peptide backbone fragmentation. CAD MS/MS spectrum of GlcNAcylated peptide (YSPTgSPSK,  $[M+GlcNAc+2H]^{2+}$  535.2) showing a major peak  $[M+H]^+$  866.3 corresponding to loss of GlcNAc



overexpressed and purified recombinant *O*-GlcNAc proteins), which is not practical for most endogenous complex and substoichiometric biological samples.

### Enrichment is Key for *O*-GlcNAc Site-mapping

Recent proteomic analyses have been able to identify hundreds of phosphorylation sites in one experiment. This is in part due to extensive fractionation and, more importantly, the specific enrichment of phosphopeptides, such as by strong cation exchange [25], immobilized metal ion chromatography [26], titanium dioxide [27], etc. Without enrichment, it is difficult to detect phosphopeptides because they are usually in low abundance and suffer from low ionization efficiency. We observe similar (or even worse) ion suppression effects for *O*-GlcNAc peptides. *O*-GlcNAc peptide ionization is suppressed 5- to 10-fold under the presence of unmodified peptides in matrix-assisted laser desorption ionization (Fig. 2A). GlcNAcylated peptides usually coelute with the unmodified forms on reverse-phase HPLC (Fig. 2B). Electrospray ionization of GlcNAcylated peptides is suppressed in the presence of the unmodified peptides (Fig. 3C). Thus, it is obvious that the detection of *O*-GlcNAc peptides in substoichiometric biological samples, where the molar percentage of *O*-GlcNAc peptides is usually even lower, is almost impossible without specific enrichment.

### Direct Enrichment of *O*-GlcNAc Peptides by Affinity Chromatography

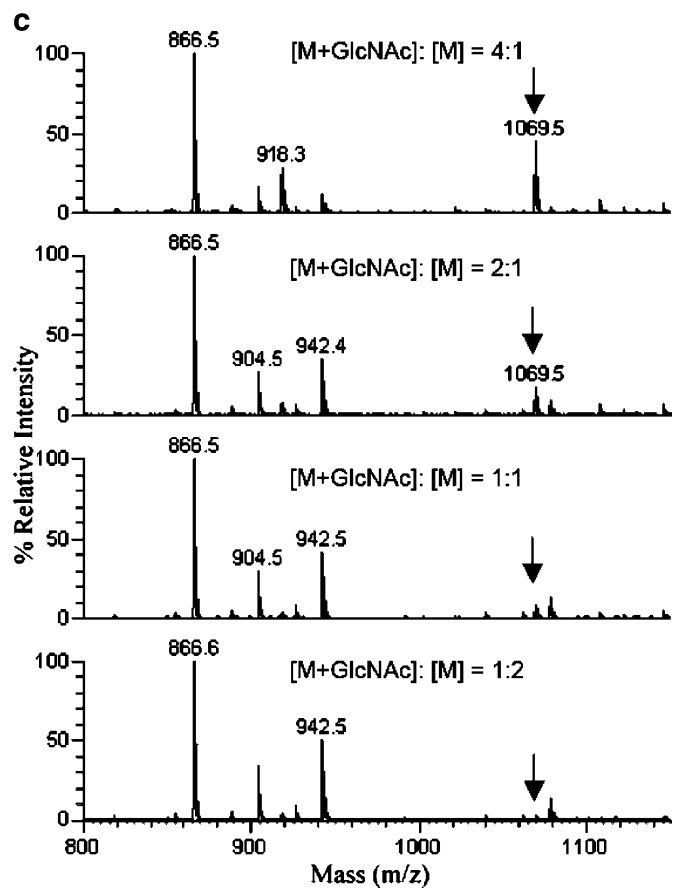
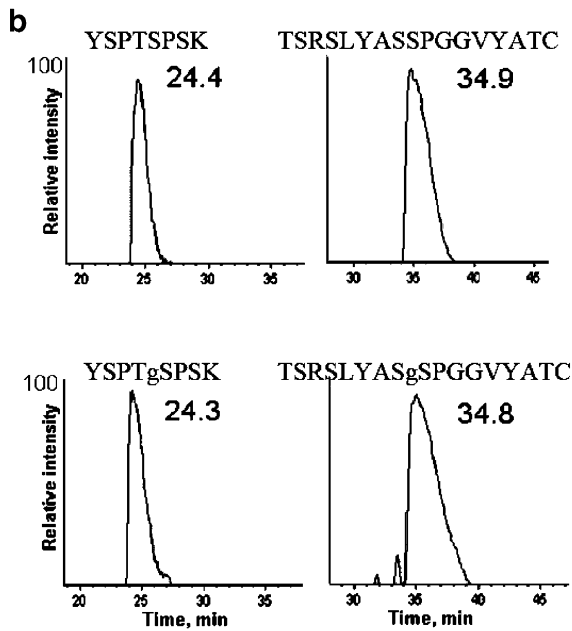
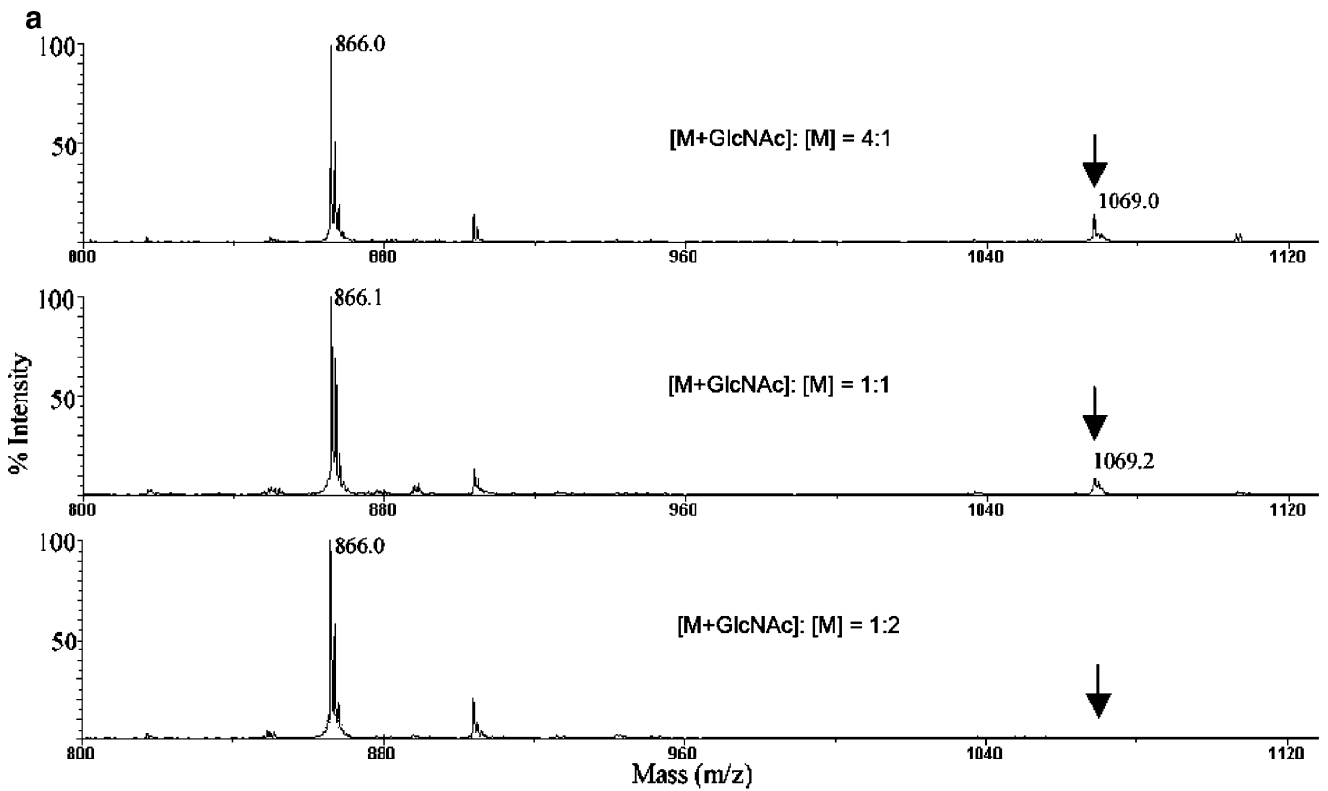
Initially, enrichment of *O*-GlcNAc peptides was based on immobilized *Ricinus communis* agglutinin I (RCA I), a lectin that binds weakly to oligosaccharides with terminal galactose residues. Hayes et al. [28] used RCA I to retard the movement of the GalT-labeled form of *O*-GlcNAc peptides during chromatography and thus separated them from mixtures of unmodified peptides, which reduced the number of HPLC rounds needed to purify *O*-GlcNAc peptides and greatly facilitated *O*-GlcNAc site-mapping by Edman degradation. Haynes and Aebersold [23] used the same method to enrich *O*-GlcNAcylated peptides from tryptic digests of bovine  $\alpha$ -crystallin before site-mapping, using a triple quadrupole mass spectrometer. Another plant lectin, wheat germ agglutinin (WGA), binds directly but very weakly to terminal GlcNAc residues and more strongly to sialic acids. Using WGA chromatography, Vosseller et al. [29] enriched 145 unique *O*-GlcNAc peptides from neuron postsynaptic density preparation. The problems of using lectins for enrichment include low specificity and weak binding affinity. Thus, it is not

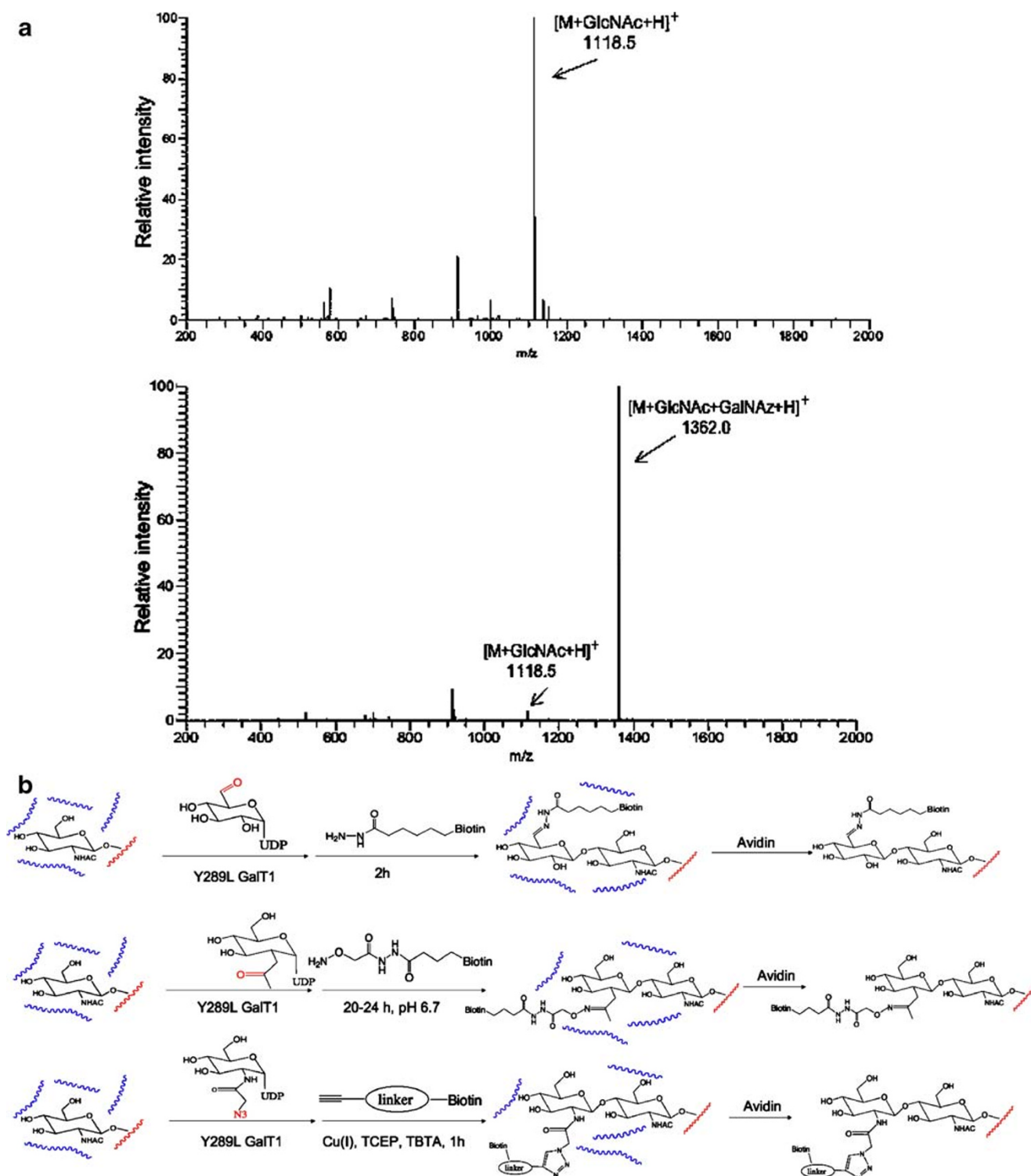
**Fig. 2** *O*-GlcNAc peptide ionization is suppressed in presence of unmodified peptides. **A** GlcNAcylated peptide YSPTgSPSK ( $[M+GlcNAc+H]^+$  1069.5) and its unmodified form ( $[M+H]^+$  866.4) are mixed at indicated molar ratios and analyzed by MALDI-TOF using  $\alpha$ -cyano-4-hydroxycinnamic acid as the matrix. *Arrows* point to peaks of  $m/z$  1069.5. **B** GlcNAcylated peptides coelute with their unmodified forms on reverse-phase HPLC. RF column is packed with a 10-cm-length of C18 beads. Gradient is 5–40% solvent B (solvent A: 0.1% formic acid; solvent B: 90% acetonitrile, 0.1% formic acid) in 60 min at a flow rate of 300 nl/min. Extracted ion chromatograms are shown. Time of elution peaks is as indicated. **C** GlcNAcylated peptide YSPTgSPSK ( $[M+GlcNAc+H]^+$  1069.5) is mixed with its unmodified form ( $[M+H]^+$  866.4) at indicated molar ratios and sprayed into LCQ mass spectrometer via electrospray ionization. Averaged full scan spectra are shown. *Arrows* point to peaks of  $m/z$  1069.5. Peaks of  $m/z$  904.5, 918.3, and 942.5 are due to peptide impurities

unexpected that peptides bearing “classical” *O*- or *N*-linked glycan will coelute with *O*-GlcNAc peptides and complicate the subsequent mass spectrometric analysis. For the same reason, Vosseller et al. [29] had to use a ~39-ft-long column to retard the elution of *O*-GlcNAc peptides during chromatography. However, by directly interfacing this long nanobore column into the mass spectrometer, they were able to map many *O*-GlcNAc sites in these experiments. Affinity enrichment of *O*-GlcNAc peptides based on pan-specific *O*-GlcNAc antibodies is yet to be proven generally useful, but little effort has been focused on this approach. While the antibodies have higher specificity, they also have low efficiency due to weak binding affinity.

### *O*-GlcNAc Mapping Based on $\beta$ -elimination

In spite of many other approaches available to selectively isolate phosphopeptides,  $\beta$ -elimination-based methods for mapping phosphorylation sites are still widely used to date. The well-established  $\beta$ -elimination under alkaline condition cleaves phospho groups from Ser or Thr residues and generates dehydroalanine (Ser) or dehydrobutyric acid (Thr). Michael addition with a nucleophilic reagent is then used to attach an affinity tag to the peptides, enabling selective enrichment. The same approach can be applied to enrich *O*-GlcNAc peptides because *O*-GlcNAc can be  $\beta$ -eliminated with even higher efficiency [1, 11–14], allowing for very mild conditions. Previously, our laboratory developed a method called  $\beta$ -elimination followed by Michael addition with dithiothreitol (BEMAD) to enrich and site-map *O*-GlcNAc [30]. With this method, tryptic peptides are first treated with phosphatases to remove interfering phosphogroups. After  $\beta$ -elimination, dithiothreitol (DTT) is used as the Michael addition donor. The attached DTT serves as a tag to enable selective enrichment by thiol-sepharose resins. This chemical derivatization is also stable in the mass spectrometer, allowing facile site-mapping by LC-MS/MS.





**Fig. 3** Chemoenzymatic approaches for O-GlcNAc peptide enrichment. **A** Full scan spectra of GlcNAcylated peptide (TAPTgSTIAPG,  $[M+GlcNAc+H]^+$  1118.5) solution before and after 12-h incubation with Y289L GalT1 and UDP-azido-Gal. **B** Flow charts representing

approaches using different functional groups for biotinylation [18, 35, 36]. *TCEP*, Tris(2-carboxyethyl)phosphine hydrochloride; *TBTA*, Tris [(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine

McLachlin and Chait [31] reported nonspecific  $\beta$ -elimination of free hydroxyl groups from Ser or Thr side chains with extended treatment of base. Li et al. [32] showed that this side reaction was not detected under mild

alkalinity at ambient temperature for up to 24 h. Thus, these data suggested that BEMAD experiments should be performed at room temperature with limited alkalinity. Barium hydroxide is preferred over sodium hydroxide as the

choice of base since barium provides faster reaction kinetics without compromising specificity [33]. Since BEMAD is an indirect method for site mapping, it is essential that a portion of the sample be used as a negative control that is pretreated with  $\beta$ -*N*-acetyl-hexosaminidase to remove all terminal hexosamines prior to analyses. Another issue of the current BEMAD method is the suboptimal enrichment efficiency using thiol-sepharose. A biotinylated reagent was tried as an alternative for nucleophilic attachment, but it resulted in other issues, such as low elution efficiency and reduced MS/MS performance due to fragmentation of the biotin moiety under CAD condition [30]. Design and synthesis of new nucleophilic reagents with cleavable biotin groups may provide a way to improve this technique.

### Enrichment Enabled by Chemoenzymatic Labeling

Ramakrishnan and Qasba engineered the  $\beta$ 1, 4-galactosyl-transferase 1 (GalT1) by mutating tyrosine 289 into a leucine. The single mutation leads to an enlarged donor substrate-binding pocket and allows the enzyme to tolerate UDP-GalNAc and its structural analogs [34]. This mutant enzyme sets the stage for enzymatic labeling of GlcNAc with UDP-GalNAc analogs that carry a functional group that can be subsequently conjugated to affinity tag, which enables specific enrichment. The Y289L GalT1 not only is able to label full-length GlcNAcylated proteins, but it also efficiently labels *O*-GlcNAc peptides. Overnight incubation of this enzyme with UDP-azido-Gal led to virtually complete labeling of CTD-GlcNAc peptide (Fig. 3A). This feature of the enzyme makes this chemoenzymatic approach not only suitable for isolation and detection of *O*-GlcNAcylated proteins, but also ideal for highly efficient enrichment of *O*-GlcNAc bearing peptides, which is critically important for *O*-GlcNAc site-mapping.

Bülter et al. [35] described a method of synthesizing UDP-aldose- $\alpha$ -D-galactose (UDP-aldose-Gal) by oxidizing UDP-Gal with galactose oxidase. Although it has not been experimentally demonstrated, UDP-aldose-Gal may possibly be used by GalT1 as the donor substrate to label GlcNAc. As shown in Fig. 3B, attachment of this functionalized aldose-Gal to the *O*-GlcNAc moiety could provide a chemical handle for subsequent biotinylation reaction with biotin- $\epsilon$ -amidocaproylhydrazide (BACH). After this chemoenzymatic tagging, *O*-GlcNAc peptides could then be easily isolated from unmodified peptides by avidin chromatography before mass spectrometric analysis. Khidekel et al. [36] synthesized UDP-Gal-ketone, which replaced the C-2 *N*-acetyl moiety with a ketone functional group that could be biotinylated by conjugating to an aminoxy biotin derivative after being transferred to *O*-GlcNAc residues (Fig. 3B). The authors streamlined this chemoenzymatic

approach and used it to successfully identify dozens of *O*-GlcNAcylated proteins from rat brains. Most recently, Wang et al. [18] used UDP-azido-Gal to directly label *O*-GlcNAc peptides by GalT1. The biotinylation was achieved through a highly selective cycloaddition reaction of alkyne to azide, or the so-called “Click” chemistry (Fig. 3B). This copper (I)-catalyzed Click reaction occurs rapidly under very mild conditions with virtually no side reaction [37]. This is in contrast to the aldo-BACH and ketone-aminoxy biotin reaction, which either suffer reaction incompleteness [35] or require 20–24 h incubation [36]. Biotin conjugation via Click chemistry generally is complete within 1 h at room temperature, which avoids potential *O*-GlcNAc loss during extended incubation.

For the chemoenzymatic approach, *O*-GlcNAc peptides enriched by avidin chromatography will carry an uncharged bulky tag, which includes a GlcNAc residue, a galactose residue, a linker arm, and a biotin moiety. The total mass of this tag on Ser/Thr side chain usually exceeds 700 Da. Data from our laboratory showed that these tagged peptides suffered even lower ionization efficiency than peptides carrying only *O*-GlcNAc. Also, the *O*-linkage remains very labile. Under tandem MS using CAD, the majority of these peptides will lose the biotin-Gal moiety, while the rest will just lose the entire tag. Neither case will result in sufficient peptide backbone fragmentation before *O*-GlcNAc site information is lost. To address this issue, Khidekel et al. [38] combined the recently developed electron-transfer dissociation (ETD), a mild fragmentation technique that preserves the *O*-linkage, with the chemoenzymatic enrichment approach, and unambiguously mapped a number of *O*-GlcNAc sites. Alternatively, Wang et al. [18] performed BEMAD on the enriched peptides, which replaced the labile GlcNAc-Gal-Biotin tag with a stable nucleophilic reagent. This simple additional treatment improved ionization efficiency and permitted relatively easy *O*-GlcNAc site-mapping on intermediate filament protein vimentin using a conventional CAD tandem mass spectrometer.

### New Mass Spectrometric Techniques for *O*-GlcNAc Site-mapping

In contrast to the widely used low-energy CAD, which preferentially breaks *O*-linkage before fragmenting peptide backbone, electron capture dissociation (ECD) in the magnetic field of a Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR-MS) usually preserve posttranslational modifications on amino acid side chains, including phosphorylation, glycosylation, and sulfonation [39]. ETD enables the ECD-type of fragmentation in the widely used, lower-cost RF quadrupole ion trap mass spectrometers [40]. ETD has been successfully used in

studies of phosphoproteomes. However, as a fragmentation technique only, ETD does not circumvent the necessity of enriching phosphopeptides. Similarly, ETD alone is not the ultimate solution for the *O*-GlcNAc site-mapping problem. Because *O*-GlcNAc peptides are substoichiometric and suffer from severe ion suppression effect, enrichment of *O*-GlcNAc peptides is still required before site-mapping by mass spectrometry.

A potential alternative to circumvent enrichment is to apply top-down proteomics [41]. Intact proteins or briefly digested long peptides are highly charged. Thus, the negative effect of a single *O*-GlcNAc moiety on ionization efficiency may be alleviated. Schroeder et al. [42] manually analyzed the ETD spectra of an abundant peptide that the search engine failed to assign a sequence to. Based on its fragmentation pattern together with high mass accuracy and partial de novo sequencing conferred by FT-MS, they identified Ser74 as *O*-GlcNAcylated from a 55-amino-acid peptide with a charge state of +8. The current database search engines lack the matching power to assign sequence or modification status to highly charged peptides based on their complex MS/MS spectra, which feature also highly charged fragmentation ions. Hence, manual inspection and analysis (together with a certain amount of prior knowledge of the target protein) is usually required. Because of this, the effectiveness of the top-down proteomic approach for site-mapping has only been demonstrated on pure, recombinant proteins. Applying the same approach to substoichiometric proteins in complex biological mixtures is challenging.

### Site-specific *O*-GlcNAc Dynamics

Rapid and dynamic regulation of *O*-GlcNAcylation is essential in many signaling events. Accumulating evidences suggest that *O*-GlcNAc is a nutrient/stress sensor that modulates many biological functions, including transcription regulation, stress response, cell cycle regulation, and protein turnover [3]. Needless to say, investigating site-specific *O*-GlcNAc dynamics will greatly help us to understand the roles of GlcNAcylation. Fortunately, the current chemoenzymatic approach for *O*-GlcNAc peptide enrichment and site-mapping can be easily combined with isotopic labeling for quantitative analysis of site-specific *O*-GlcNAcylation by mass spectrometry. For example, Khidekel et al. [38] differentially labeled N-terminal amines and  $\epsilon$ -amino groups of lysine residues of the enriched *O*-GlcNAc peptides with either hydrogen or deuterium. We have used natural or deuterated DTT for BEMAD derivatization after *O*-GlcNAc peptides are enriched from different biological states. Since dysregulated *O*-GlcNAcylation is closely related to diseases like cancer, neurodegenerative

disease, and diabetes [3, 43], quantitative proteomic analysis of *O*-GlcNAc dynamics therefore holds great potential for the discovery of disease mechanisms and biomarkers.

**Acknowledgements** We thank Dr. Frank Comer and Ms. Shino Shimoji for synthesizing *O*-GlcNAc peptides and members of the Hart laboratory for critical reading of the manuscript. The work is supported by National Institutes of Health (NIH) grants CA42486, HD13563, DK61671, and DK71280 and NIH contract N01-HV-28180. Dr. Hart receives a share of royalty received by the university on sales of the CTD 110.6 antibody. Terms of this arrangement are managed by Johns Hopkins University School of Medicine.

### References

1. Torres C-R, Hart GW. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J Biol Chem* 1984;259:3308–17.
2. Holt G, Hart GW. Subcellular distribution of terminal N-acetylglucosamine residues on oligosaccharides. Distribution of the novel protein-saccharide structure—O-linked GlcNAc. *J Biol Chem* 1986;261:8049–57.
3. Hart G, Housley M, Slawson C. Cycling of O-linked b-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 2007;446:1017–22.
4. McClain D, Crook E. Hexosamines and insulin resistance. *Diabetes* 1996;45:1003–9.
5. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science* 2002;298:1912–34.
6. Kreppel L, Hart G. Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. *J Biol Chem* 1999;274:32015–22.
7. Lubas W, Hanover J. Functional expression of O-linked GlcNAc transferase. Domain structure and substrate specificity. *J Biol Chem* 2000;275:10983–8.
8. Gao Y, Wells L, Comer F, Parker G, Hart G. Dynamic O-glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a neutral, cytosolic b-N-acetylglucosaminidase from human brain. *J Biol Chem* 2001;276:9838–45.
9. Roquemore E, Chou T, Hart G. As GalT labels terminal GlcNAc sugars, experiments using PNGaseF and b-elimination serve as valuable controls. *Methods Enzymol* 1994;230:443–60.
10. Greis K, Gibson W, Hart G. Site-specific glycosylation of the human cytomegalovirus tegument basic phosphoprotein (UL32) at serine 921 and serine 952. *J Virol* 1994;68:8339–49.
11. Greis K, Hayes B, Comer F, Kirk M, Barnes S, Lowary T, Hart G. Selective detection and site-analysis of O-GlcNAc-modified glycopeptides by beta-elimination and tandem electrospray mass spectrometry. *Anal Biochem* 1996;234:38–49.
12. Cole RN, Hart GW. Glycosylation sites flank phosphorylation sites on synapsin I: O-linked N-acetylglucosamine residues are localized within domains mediating synapsin I interactions. *J Neurochem* 1999;73:418–28.
13. Zachara NE, Gao Y, Cole RN, Hart GW. Detection and analysis of proteins modified by O-linked N-acetylglucosamine. *Curr Protoc Protein Sci* 2001;2:12.8.1–25.
14. Zachara NE, Cheung WD, Hart GW. Nucleocytoplasmic glycosylation, O-GlcNAc: identification and site mapping. In: Dickson



- RC, Mendenhall MD, editors. *Methods in molecular biology—signal transduction protocols*. Clifton: Humana; 2003. p. 175–94.
15. Whelan SA, Hart GW. Identification of O-GlcNAc sites on proteins. *Methods Enzymol* 2006;415:113–33.
  16. Holt G, Snow C, Senior A, Haltiwanger R, Gerace L, Hart G. Nuclear pore complex glycoproteins contain cytoplasmically disposed O-linked N-acetylglucosamine. *J Cell Biol* 1987;104:1157–64.
  17. Comer F, Vosseller K, Wells L, Accavitti MA, Hart GW. Characterization of a mouse monoclonal antibody specific for O-linked N-acetylglucosamine. *Anal Biochem* 2001;293:169–77.
  18. Wang Z, Pandey A, Hart G. Dynamic interplay between O-Linked N-acetylglucosaminylation and glycogen synthase kinase-3-dependent phosphorylation. *Mol Cell Proteomics* 2007;6:1365–79.
  19. Vocadlo D, Hang H, Kim E, Hanover J, Bertozzi C. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc Natl Acad Sci U S A* 2003;100:9116–21.
  20. Nandi A, Sprung R, Barma D, Zhao Y, Kim S, Falck J, Zhao Y. Global identification of O-GlcNAc-modified proteins. *Anal Chem* 2006;78:452–8.
  21. Zachara NE, Hart GW. Cell signaling, the essential role of O-GlcNAc. *Biochim Biophys Acta* 2006;1761:599–617.
  22. Nagata K, Izawa I, Inagaki M. A decade of site- and phosphorylation state-specific antibodies: recent advances in studies of spatiotemporal protein phosphorylation. *Genes Cells* 2001;6:653–64.
  23. Haynes P, Aebersold R. Simultaneous detection and identification of O-GlcNAc-modified glycoproteins using liquid chromatography tandem mass spectrometry. *Anal Chem* 2000;72:5402–10.
  24. Chalkley RJ, Burlingame AL. Identification of novel sites of O-Nacetylglucosamine modification of serum response factor using quadrupole time-of-flight mass spectrometry. *Mol Cell Proteomics* 2003;2:182–90.
  25. Beausoleil S, Jedrychowski M, Schwartz D, Elias J, Villén J, Li J, Cohn M, Cantley L, Gygi S. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 2004;101:12130–5.
  26. Chi A, Huttenhower C, Geer L, Coon J, Syka J, Bai D, Schabanowitz J, Burke D, Troyanskaya O, Hunt D. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci U S A* 2007;104:2193–8.
  27. Larsen M, Thingholm T, Jensen O, Roepstoff P, Jørgensen T. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumn. *Mol Cell Proteomics* 2005;4:873–86.
  28. Hayes B, Greis K, Hart G. Specific isolation of O-linked N-acetylglucosamine glycopeptides from complex mixtures. *Anal Biochem* 1995;228:115–22.
  29. Vosseller K, Trinidad J, Chalkley R, Specht C, Thalhammer A, Lynn A, Snedecor J, Guan S, Medzihradsky K, Maltby D, Schoepfer R, Burlingame A. O-Linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 2006;5:923–34.
  30. Wells L, Vosseller K, Cole R, Cronshaw J, Matunis M, Hart G. Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol Cell Proteomics* 2002;1:791–804.
  31. McLachlin D, Chait B. Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal Chem* 2003;75:6826–36.
  32. Li W, Backlund P, Boykins R, Wang G, Chen H. Susceptibility of the hydroxyl groups in serine and threonine to beta-elimination/Michael addition under commonly used moderately high-temperature conditions. *Anal Biochem* 2003;323:94–102.
  33. Rusnak F, Zhou J, Hathaway G. Reaction of phosphorylated and O-glycosylated peptides by chemically targeted identification at ambient temperature. *J Biomol Tech* 2004;15:296–304.
  34. Ramakrishnan B, Qasba P. Structure-based design of b1,4-galactosyltransferase I (b4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity. Point mutation broadens b4gal-T1 donor specificity. *J Biol Chem* 2002;277:20833–9.
  35. Bültner T, Schumacher T, Namdjou D, Gutierrez Gallego R, Clausen H, Elling L. Chemoenzymatic synthesis of biotinylated nucleotide sugars as substrates for glycosyltransferases. *Chem-biochem* 2001;2:884–94.
  36. Khidekel N, Ficarro S, Peters E, Hsieh-Wilson L. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci U S A* 2004;101:13132–7.
  37. Kolb H, Finn M, Sharpless K. Click chemistry: diverse chemical function from a few good reactions. *Angew Chem* 2001;40:2004–21.
  38. Khidekel N, Ficarro S, Clark P, Bryan M, Swaney D, Rexach J, Sun Y, Coon J, Peters E, Hsieh-Wilson L. Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nat Chem Biol* 2007;3:339–48.
  39. Zubarev R, Kelleher N, McLafferty F. ECD of multiply charged protein cations. A non-ergodic process. *J Am Chem Soc* 1998;120:3265–6.
  40. Syka J, Coon J, Schroeder M, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 2004;101:9528–33.
  41. Chi A, Bai D, Geer L, Shabanowitz J, Hunt D. Analysis of intact proteins on a chromatographic time scale by electron transfer dissociation tandem mass spectrometry. *Int J Mass Spectrom* 2007;259:197–203.
  42. Schroeder M, Webb D, Shabanowitz J, Horwitz A, Hunt D. Methods for the detection of paxillin post-translational modifications and interacting proteins by mass spectrometry. *J Proteome Res* 2005;4:1832–41.
  43. Dias W, Hart G. O-GlcNAc modification in diabetes and Alzheimer's disease. *Mol Biosyst* 2007;3:766–72.