Journal of proteome research

Article

pubs.acs.org/jpr

Proteome Wide Purification and Identification of *O*-GlcNAc-Modified Proteins Using Click Chemistry and Mass Spectrometry

Hannes Hahne,[†] Nadine Sobotzki,[†,⊥] Tamara Nyberg,[‡] Dominic Helm,[†] Vladimir S. Borodkin,[§] Daan M. F. van Aalten,[§] Brian Agnew,[‡] and Bernhard Kuster*[,†,∥]

[†]Chair for Proteomics and Bioanalytics, Center of Life and Food Sciences, Weihenstephan, Technische Universität München, Freising, Germany

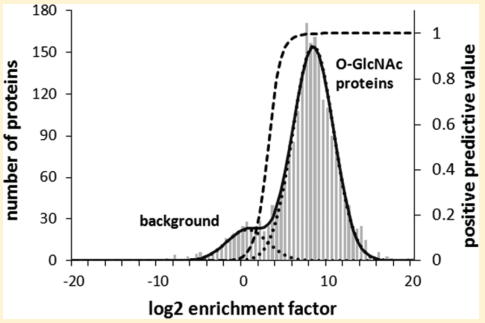
[‡]Life Technologies, Eugene, Oregon 97402, United States

[§]The MRC Protein Phosphorylation Unit, College of Life Sciences, University of Dundee, Dundee, Scotland

[∥]Center for Integrated Protein Science, Munich, Germany

Supporting Information

ABSTRACT: The post-translational modification of proteins with *N*-acetylglucosamine (*O*-GlcNAc) is involved in the regulation of a wide variety of cellular processes and associated with a number of chronic diseases. Despite its emerging biological significance, the systematic identification of *O*-GlcNAc proteins is still challenging. In the present study, we demonstrate a significantly improved *O*-GlcNAc protein enrichment procedure, which exploits metabolic labeling of cells by azide-modified GlcNAc and copper-mediated Click chemistry for purification of modified proteins on an alkyne-resin. On-resin proteolysis using trypsin followed by LC−MS/MS afforded the identification of around 1500 *O*-GlcNAc proteins from a single cell line. Subsequent elution of covalently resin bound *O*-GlcNAc peptides using selective *β*-elimination enabled the identification of 185 *O*-GlcNAc modification sites on 80 proteins. To demonstrate the practical utility of the developed approach, we studied the global effects of the *O*-GlcNAcase inhibitor GlcNAcstatin G on the level of *O*-GlcNAc modification of cellular proteins. About 200 proteins including several key players involved in the hexosamine signaling pathway showed significantly increased *O*-GlcNAcylation levels in response to the drug, which further strengthens the link of *O*-GlcNAc protein modification to cellular nutrient sensing and response.



KEYWORDS: *proteomics, mass spectrometry, post-translational modifications, O-GlcNAc, metabolic labeling, Click chemistry, β-elimination*

INTRODUCTION

O-linked *N*-acetylglucosamine (*O*-GlcNAc) is an emerging dynamic post-translational modification (PTM) of serine and threonine residues of proteins and was first discovered in 1984 by Torres and Hart.[1] Since then, *O*-GlcNAc has been found on a wide range of nuclear and cytoplasmic proteins involved in almost all cellular processes including signaling, cell cycle regulation, transcription and translation regulation, protein trafficking, and protein quality control, as well as stress and survival.[2−4] The evolutionary conserved enzyme *O*-GlcNAc transferase (OGT) catalyzes the attachment of GlcNAc from uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) to specific protein residues. This reaction can be reversed by the nuclear and cytoplasmic *O*-GlcNAcase (OGA), which is a bifunctional protein with an N-terminal glycosidase domain, a C-terminal histone acetyltransferase domain,[5] and an interjacent caspase-3 cleavage site.[6] The *O*-GlcNAcase activity of OGA can be inhibited using small molecule inhibitors resulting in a dramatic increase in overall cellular *O*-GlcNAc levels.[7] OGA inhibitors have been extensively used to study the role of *O*-GlcNAc in nutrient-sensing and diabetes, protection from cellular stress or cellular signaling, and the interplay of *O*-GlcNAcylation and phosphorylation.[8]

Despite the emerging relevance of *O*-GlcNAc in a multitude of cellular processes, the systematic discovery of *O*-GlcNAc proteins is still challenging. Like for many other PTMs, liquid chromatography tandem mass spectrometry (LC−MS/MS) is the method of choice and abundant *O*-GlcNAc proteins can sometimes be identified directly from full proteome digests.[9] However, this task is more commonly achieved by adding a selective enrichment step. A number of different conceptual approaches have been developed for this purpose both at the level of modified peptides and proteins. For proteome wide applications, the most promising technologies so far are based on lectin affinity chromatography using wheat germ agglutinin[10,11] and a chemoenzymatic approach that tags endogenous *O*-GlcNAc moieties with azide-labeled galactose and allows Click chemistry-based enrichment of the tagged proteins using a photocleavable biotin probe.[12,13] The conceptual advantage of

Received: October 14, 2012
Published: January 9, 2013

the above approaches is that they leave the modification on the peptide and thus, in principle, allow not only the identification of *O*-GlcNAc peptides and proteins but also the direct determination of the site of modification within the peptide or protein. Direct site determination is, however, often complicated by the fact that the modification is labile during the standard mass spectrometric readout of collision-induced dissociation (CID) and therefore the site information is usually lost. This shortcoming can, in principle, be overcome by the use of electron capture dissociation (ECD) or electron transfer dissociation (ETD) mass spectrometry,[10,14,15] but these techniques also have shortcomings, notably a rather poor overall sensitivity. As a result, alternative strategies that resort to semidirect or even indirect measures of modification identification and site localization have been developed.

For example, several groups have employed metabolic labeling of *O*-GlcNAc proteins by azide or alkyne-tagged *N*-acetylglucosamine[16] (GlcNAz and GlcNAlk, respectively) and subsequently coupled the modified proteins to an affinity probe via copper-catalyzed azide/alkyne Click chemistry (CuAAC) or Staudinger ligation. The affinity enriched *O*-GlcNAc proteins can then be identified by mass spectrometry.[17−20] However, these approaches did not enable the direct identification of a single *O*-GlcNAc site, hence, rendering the information regarding the *O*-GlcNAc modification rather indirect. As an alternative, β-elimination of *O*-GlcNAc moieties followed by Michael addition (BEMAD) has been employed for the enrichment and site identification of *O*-GlcNAc proteins.[10,21,22] In the BEMAD approach, *O*-GlcNAc moieties are eliminated under strong alkaline conditions resulting in an α,β-unsaturated carbonyl group (a so-called Michael system), which can subsequently be modified using a strong nucleophile. The addition of a stable nucleophile tags the former *O*-GlcNAc site, which can be then recognized in the MS experiment. The BEMAD approach has been used frequently and has enabled the identification and quantification of numerous rodent brain proteins along with their sites.[10,21,22] A clear downside of the BEMAD approach is that phosphorylated and, to a lesser extent, unmodified serine, threonine and alkylated cysteine residues are also susceptible to β-elimination under certain experimental conditions,[22−25] necessitating additional means to control false-positive *O*-GlcNAc site assignments.

In the present study, we demonstrate that the combination of the above biochemical methods (notably metabolic GlcNAz labeling, Click chemistry, on resin proteolysis, and selective β-elimination) enables the efficient enrichment and identification of *O*-GlcNAc proteins along with their sites. The approach has, in principle, already been described, but suffered from ineffective biochemical enrichment, identification of only small numbers of potential *O*-GlcNAc proteins, and did not enable the direct identification of *O*-GlcNAc sites.[26−28] However, when complemented with additional means to reduce and control unspecific protein background (notably ultracentrifugation of the cell lysate, washing with a strong copper chelator, and rigorous label-free quantification), the method allowed the identification of around 1500 high-confidence *O*-GlcNAc-modified proteins from a single cell line along with >180 modification sites. Furthermore, we were able to demonstrate the practical utility of the developed approach by studying the effect of the OGA inhibitor GlcNAcstatin G on the *O*-GlcNAc proteome, which led to the identification of several key signaling proteins.

MATERIALS AND METHODS

Peptide Synthesis and Assessment of β-Elimination/Michael Addition Conditions

O-GlcNAc- and phosphopeptides for the systematic assessment of β-elimination/Michael addition conditions were synthesized in our laboratory using standard solid-phase peptide synthesis.[15,29] Beta-elimination reactions were performed on dried peptides using 1% triethylamine and 0.1% NaOH in 20% ethanol at different temperatures and for various amounts of time.[21] In addition, β-elimination was performed using the GlycoProfile β-elimination kit (Sigma-Aldrich, Taufkirchen Germany) according to the manufacturer's instructions. Michael addition was performed using β-mercaptoethanol, dithiothreitol, or 1-propanethiol at different reagent concentrations (Table S1, Supporting Information). The β-elimination/Michael addition reaction was quenched with 1% trifluoroacetic acid (TFA). Peptides were dried in vacuo, desalted using C_{18} StageTips,[30] and reconstituted in 20 μL of 0.1% formic acid (FA) prior to LC−MS/MS analysis (for details, see Supporting Information). Beta elimination of resin-bound *O*-GlcNAc peptides was eventually performed using the GlycoProfile β-elimination kit (for details, see below).

Cell Culture, Metabolic Labeling, and Inhibitor Treatment

HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM; PAA, Pasching, Austria) containing 1.0 g/L glucose supplemented with 10% (v/v) fetal bovine serum (FBS; PAA, Pasching, Austria) at 37 °C with humidified air and 5% CO_2. For metabolic labeling, HEK293 cells were treated with 200 μM tetraacetylated GlcNAz (Ac_4GlcNAz; Life Technologies, Eugene, OR) for 18 h. In the case of GlcNAcstatin G-treated cells, HEK293 cells were metabolically labeled and treated with 20 μM GlcNAcstatin G for two hours before cell lysis.[31]

Click Chemistry-Based Enrichment of *O*-GlcNAc Proteins

The detailed procedure is described in the Supporting Information section. Briefly, cell lysis and Click chemistry-based enrichment were performed using the Click-iT protein enrichment kit according to the manufacturer's instructions with two relevant modifications. We incorporated an ultracentrifugation step to clear the lysate from fine insoluble matter and an additional wash step with the strong copper chelator diethylene triamine pentaacetic acid (DTPA) after CuAAC. In order to control the selectivity of the *O*-GlcNAc protein purification procedure, we performed the procedure in parallel with GlcNAz-labeled and unlabeled HEK293 cells and used the same amount of protein starting material for the CuAAC enrichment. Briefly, following cell lysis by sonification in a urea buffer (8 M urea, 200 mM TrisHCl pH 8, 4% CHAPS, and 1 M NaCl), cell debris were pelleted (10 000*g*, 15 min, 4 °C), and the samples were subjected to ultracentrifugation (145 000*g*, 60 min, 4 °C). Three milligrams of total protein in 800 μL lysis buffer was alkylated with 10 mM iodo acetamide (IAM, 60 min, room temperature) and used for the subsequent CuAAC reaction according to the manufacturer's instructions using 200 μL of the alkyne resin slurry. After overnight Click reaction, the resin was washed 3× with 1.5 mL of 10 mM DTPA before proteins were reduced (10 mM dithiothreitol, 30 min, 55 °C) and alkylated (50 mM IAM, 60 min, room temperature). Following extensive washing with 5× 2 mLof SDS wash buffer (100 mM TrisHCl pH 8, 1% SDS, 250 mM NaCl, and 5 mM EDTA), 5× 2 mL of urea buffer (8 M urea and 100 mM

dx.doi.org/10.1021/pr300967y | J. Proteome Res. 2013, 12, 927−936

TrisHCl pH 8), and 5× 2 mL of 20% ACN, the resin-bound proteins were digested in 50 mM TrisHCl (pH 7.6) in two steps, first for two hours and then overnight, using each time 0.5 μg of trypsin. After on-resin digestion, the supernatant was aspirated and desalted using C_{18} StageTips before LC−MS/MS analysis, for which 25% of the sample was used.

On-Resin Dephosphorylation and Elution of *O*-GlcNAc Peptides by β-Elimination

Following on-resin digestion, the resin was washed twice with 1.8 mL of double-distilled water (ddH_2O) and once with 1.8 mL of dephosphorylation buffer (50 mM TrisHCl pH 7.6, 100 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$, and 1 mM $MnCl_2$). Dephosphorylation was performed at 37 °C for 6 h in 400 μL using 800 U λ phosphatase and 20 U calf intestine phosphatase (New England Biolabs, Frankfurt a. M., Germany). Following dephosphorylation, the resin was washed twice with 1.8 mL of ddH_2O, and the slurry volume was adjusted to 300 μL with ddH_2O before β-elimination using the GlycoProfile β-elimination kit. The β-elimination reaction was incubated on an end-over-end shaker with extensive mixing at 4 °C and quenched after 24 h with 1% TFA. β-eliminated peptides were desalted and concentrated with C_{18} StageTips before LC−MS/MS analysis, for which 50% of the sample was used.

Liquid Chromatography and Mass Spectrometry

Mass spectrometry was performed on an LTQ Orbitrap XL or an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, Germany) connected to a nanoLC Ultra 1D+ liquid chromatography system (Eksigent, CA) using an in-house packed precolumn (20 mm × 75 μm ReproSil-Pur C18, Dr. Maisch, Germany) and analytical column (400 mm × 50 μm ReproSil-Pur C18, Dr. Maisch, Germany). The mass spectrometer was equipped with a nanoelectrospray ion source (Proxeon Biosystems, DK), and the electrospray voltage was applied via a liquid junction. The mass spectrometer was operated in data-dependent mode and all measurements were performed in positive ion mode. Intact peptide mass spectra were acquired at a resolution of 60 000 (at m/z 400) and an automatic gain control (AGC) target value of 10^6, followed by fragmentation of the most intense ions by collision-induced dissociation (CID; LTQ Orbitrap XL) or higher energy-collision induced dissociation (HCD; LTQ Orbitrap Velos). Full scans were acquired in profile mode, whereas all tandem mass spectra were acquired in centroid mode. CID was performed for up to 15 MS/MS (2 h gradient) or 8 MS/MS (4 h gradient) per full scan with 35% normalized collision energy (NCE), and an AGC target value of 5000. HCD was performed for up to 10 MS/MS per full scan with 40% NCE and an AGC target value of 35 000. Singly charged ions and ions without assigned charge state were excluded from fragmentation, and fragmented precursor ions were dynamically excluded (2 h gradient, 10 s; 4 h gradient, 30 s). Internal calibration was performed using the polysiloxane ion signal at m/z 445.1200 present in ambient laboratory air.

Protein Identification and Quantification

Protein identification and intensity based label free quantification from on-resin digestion experiments was achieved with Mascot version 2.3.02 (Matrix Science, U.K.) in combination with Progenesis LC-MS 4.0 (Nonlinear Dynamics, U.K.). LC−MS/MS experiments were manually prealigned and then submitted to the automated alignment routine in Progenesis. Features with two isotopes or less were discarded, and tandem MS spectra were required to have a minimal ion count of 100. Further processing steps included deisotoping and deconvolution of tandem MS spectra. Resulting peaklists were searched against the UniProtKB complete human proteome set (download date 26.10.2010; 110 550 sequences) combined with sequences of common contaminants using Mascot. The target-decoy option of Mascot was enabled, and search parameters included a precursor tolerance of 10 ppm and a fragment tolerance of 0.5 Da for CID spectra and 0.02 Da for HCD spectra. Enzyme specificity was set to trypsin, and up to two missed cleavage sites were allowed. The Mascot ^{13}C option, which accounts for the misassignment of the monoisotopic precursor peak, was set to 1. The following variable modifications were considered: oxidation of Met and carbamidomethylation of Cys. The database search results were processed using the built-in Mascot Percolator option[32,33] and filtered at a score of 13, which corresponds to the posterior error probability of 0.05. This results in a peptide FDR of 0.94% and 1.06% for the global *O*-GlcNAc profiling data set and the OGA inhibition data set, respectively. The Mascot results were imported into Progenesis LC−MS for protein grouping and quantification.

Protein identification from β-elimination experiments was performed with Mascot using the Mascot Distiller version 2.3 (Matrix Science, U.K.) for data processing. Search parameters were set as detailed above except that dehydration of Ser and Thr as well as β-elimination of Cys (Δm = −33.9877 Da) was used as variable modification. Mascot search results were processed using the Mascot Percolator stand-alone software[32,33] and imported into Scaffold version 3.5.1 (Proteome Software, OR). Mascot Percolator results were filtered at a score of 13. Ascore-based localization probabilities[34] for β-eliminated peptides were calculated with Scaffold PTM 2.0.0 (Proteome Software, Portland, OR). Peptides for which the β-elimination site could not be localized to either a Ser or Thr residue were discarded.

Mass spectrometry raw data, peak list files, and Mascot and Progenesis result files can be downloaded from the ProteomeXchange repository (http://www.proteomexchange.org, accession no. PXD000061).

Data Analysis

Biochemical *O*-GlcNAc protein enrichment factors were determined based on label-free protein quantification comparing GlcNAz-labeled samples and samples without metabolic labeling. Briefly, the biochemical enrichment factor has been calculated as the ratio of intensities of a given protein in the GlcNAz-labeled sample compared to the sample without metabolic labeling. Protein intensities represent summed peptide intensities of unique peptides. Missing values were replaced with arbitrarily chosen small intensity values (0.008) to avoid zero and infinite ratios. Furthermore, proteins representing biochemical noise (i.e., very low intensity in all samples of an experiment) were discarded, when they did not show a minimum intensity in at least one sample. In the case of the global *O*-GlcNAc proteome profiling of HEK293 cells, the minimum intensity was set to 10^5. For the *O*-GlcNAc proteome profiling in response to OGA inhibition, the minimum intensity was set to 10^4. The differences are due to the different mass spectrometers used.

A positive predictive value (PPV) of representing an *O*-GlcNAc protein was calculated for each quantified protein based on the bimodal $\log_2$ distribution of biochemical

dx.doi.org/10.1021/pr300967y | J. Proteome Res. 2013, 12, 927−936

enrichment factors. Briefly, the $\log_2$ distribution of biochemical enrichment factors can be approximated with two Gaussian distributions, one for background proteins and one for specifically enriched *O*-GlcNAc proteins. The resulting sum of Gaussian distributions (a Gaussian mixture model) was fitted to the observed distribution of $\log_2$ enrichment factors using the method of least-squares. A PPV was then calculated based on the resulting modeled Gaussian distributions of background proteins and *O*-GlcNAc proteins. Similarly, a PPV for the GlcNAcstatin G-responsive proteins was calculated based on the distribution of $\log_2$ ratios (±GlcNAcstatin G) with a mixture model approximating nonresponsive proteins with a Gaussian distribution and GlcNAcstatin G-responsive proteins with a gamma distribution.

To assess the novelty in our data, we compiled a comprehensive list of reported *O*-GlcNAc proteins from a number of sources.[11,20,35−41] To reduce redundancy in terms of identifiers and orthologs, the proteins were processed as well as compared to *O*-GlcNAc proteins identified using IPA (Ingenuity Systems, www.ingenuity.com). Biological functions and pathways enriched in the generated *O*-GlcNAc data were assessed using IPA. A right-tailed Fisher's exact test was used to calculate a *p*-value determining the probability that each biological function assigned to that data set is due to chance alone. The *p*-value was corrected for multiple hypothesis testing using the method of Benjamini−Hochberg.

Sequence motifs were extracted with X-motif[42] and produced with PhosphositePlus.[37]

RESULTS AND DISCUSSION

Experimental Strategy for Global *O*-GlcNAc Proteome-Profiling in HEK293 Cells

To enable the efficient enrichment and identification of *O*-GlcNAc proteins along with their sites of modification, we developed a straightforward experimental strategy based on azide/alkyne Click chemistry using a commercially available alkyne resin (Figure 1).[28] *O*-GlcNAc proteins are metabolically labeled with GlcNAz, covalently conjugated to an alkyne agarose via CuAAC, and then purified from the vast background of unmodified proteins. Several precautions were taken to maintain selectivity toward *O*-GlcNAc purification. Cells were grown under low glucose conditions to reduce azide tagging of O- and N-linked glycans.[20] The cell lysate was subjected to ultracentrifugation prior to CuAAC to clear the lysate from fine insoluble matter, thereby significantly reducing unspecific protein background. To further minimize the copper-mediated protein background, which is frequently observed during CuAAC,[43] it turned out that the washing step with a strong copper chelator such as DTPA is absolutely required. After CuAAC-based purification, the resin bound *O*-GlcNAc proteins are digested with trypsin, thereby allowing for MS-based identification of those parts of *O*-GlcNAc proteins, which are not covalently bound to the resin. To minimize the risk of false-positive *O*-GlcNAc site assignments during *β*-elimination, the remaining resin bound *O*-GlcNAc peptides are extensively dephosphorylated before they are released by *β*-elimination, which enables the concomitant tagging and MS-based identification of the former *O*-GlcNAc sites. The selectivity of this procedure is mainly conferred by the metabolic labeling of *O*-GlcNAc proteins and the bioorthogonality of the Click chemistry reaction.

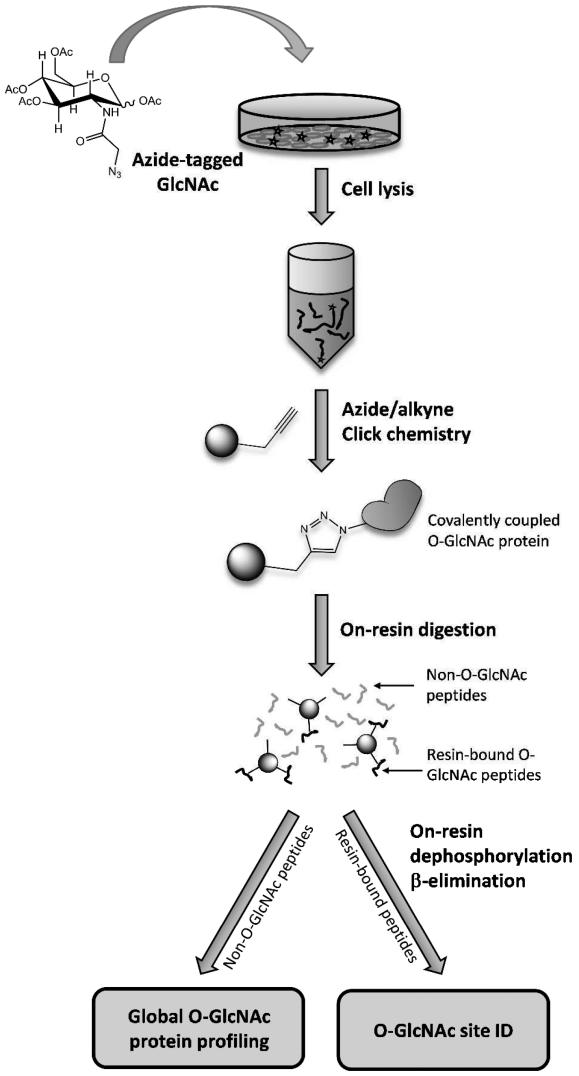


Figure 1. Experimental strategy for the Click chemistry-based enrichment and identification of *O*-GlcNAc-modified proteins. Note the potential interference with azide-tagged N- or O-linked glycoproteins.

To evaluate the merits of the strategy, we initially profiled the global *O*-GlcNAc proteome of HEK293 cells. To assess the selectivity of the enrichment procedure, we performed the experiment in parallel with GlcNAz-labeled and unlabeled HEK293 cells. Label-free intensity-based quantification was used to obtain a biochemical enrichment factor for every protein identified in the on-resin digest. The summed intensities of proteins identified from the GlcNAz-labeled sample was 1.9×10^{10}, which is >60-fold higher than the summed intensity of the negative control (Figure 2A), and the median protein enrichment factor across all proteins was 260. Together, these figures clearly show that *O*-GlcNAc modified proteins can be efficiently enriched from metabolically GlcNAz-labeled samples. Interestingly, the biochemical enrichment factors follow a bimodal distribution on a logarithmic scale (Figure 2C). The distribution centered around zero likely represents the background proteome unselectively bound by the alkyne resin. This background binding might be in part copper-mediated, but proteins may also unspecifically bind as a result of noncovalent interactions with the alkyne moieties on the beads or the agarose backbone of the beads themselves. In

dx.doi.org/10.1021/pr300967y | J. Proteome Res. 2013, 12, 927−936

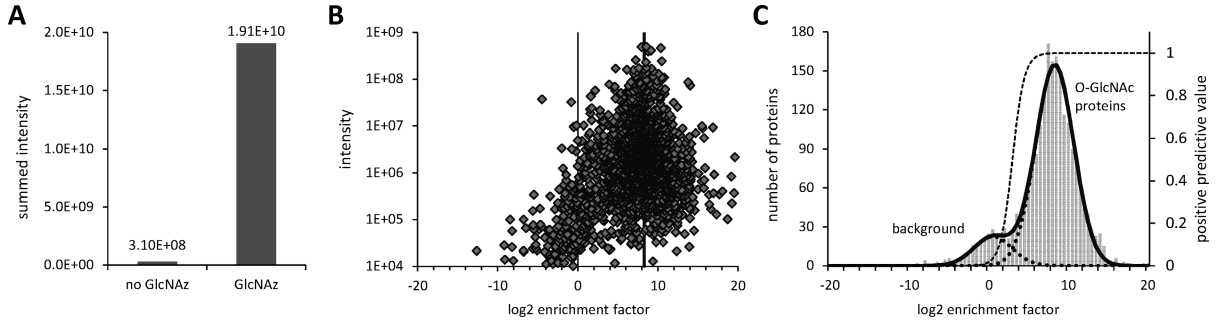


Figure 2. Global identification of *O*-GlcNAc proteins from HEK293 cells. (A) Summed intensities of enriched proteins with and without metabolic labeling of cells by GlcNAz. (B) Scatterplot of intensity and $\log_2$ biochemical enrichment of identified *O*-GlcNAc proteins. (C) The bimodal distribution of biochemical enrichment factors allows the calculation of positive predictive values for CuAAC-enriched *O*-GlcNAc proteins.

contrast, the specifically enriched *O*-GlcNAc proteome is represented by the strongly right-shifted distribution of biochemical enrichment factors. To describe the selectivity of the enrichment more quantitatively, we approximated the observed bimodal distribution by a Gaussian mixture model, which enabled us to calculate a positive predictive value (PPV) for *O*-GlcNAc proteins. For instance, a PPV of 0.99 indicates that 99% of all proteins with 82-fold biochemical enrichment originate from the distribution of *O*-GlcNAc proteins and only 1% from the distribution of background proteins. Following this rationale, we identified 1535 selectively enriched proteins with a PPV > 0.99 (1746 proteins at PPV > 0.95), representing the largest collection of *O*-GlcNAc proteins identified in a single experiment so far (Table S2 and S3 and Figure S1, Supporting Information).

The result of the GO term analysis for subcellular localization of the identified proteins is consistent with the common notion that *O*-GlcNAc is primarily a nuclear and cytoplasmic PTM. Indeed, 39% of all identified proteins are supposedly nuclear and a further 41% cytoplasmic (Figure 3A). However, we also note a small number of extracellular proteins (1%), which is presumably not *O*-GlcNAc modified, but selectively enriched as GlcNAz can be incorporated into O- and N-linked glycans of extracellular and membrane proteins.[20] Similarly, we cannot rule out the possibility that some of the identified membrane proteins might also not be *O*-GlcNAc modified.

A comparison to 1269 *O*-GlcNAc proteins compiled from comprehensive studies and resources[11,20,35−41] revealed that 74 of the 100 most abundant *O*-GlcNAc proteins in our data set (or their mouse orthologs) have been previously identified underscoring the biochemical validity of the approach. The total overlap of our data with the cited studies is 27% (338 reported *O*-GlcNAc proteins), which appears reasonable given that our data stems from a human cell line, whereas the reported 1269 *O*-GlcNAc proteins were identified from different mammalian species, tissues, and cell types, thus representing a mixture of cell-type specific *O*-GlcNAcylation profiles.

A broad range of biological functions (Figure 3B and Table S4, Supporting Information) and cellular pathways (Table S5, Supporting Information) is associated with the identified *O*-GlcNAc proteins. Notably, our data comprises a wealth of biologically interesting proteins that are often not identified in global proteome profiling studies. Examples for such proteins include the transcription factors p53, SP1, FoxO3, CREB, STAT1, and NFκB, proteins involved in epigenetic regulation such as HCF-1, Sirtuin-1, NCOA-1, −2 and −3, HDAC1, HDAC2, MLL, and proteins required for microRNA maturation (e.g., Dicer and TARPB2). Other modified proteins are involved in ubiquitination, RAN-mediated nuclear transport, aminoacyl-tRNA synthesis, and several signaling pathways. Taken together, the identified *O*-GlcNAc proteins reflect the large variety of regulatory functions *O*-GlcNAc may exert in the cell[4,44−46] and provides a method by which these roles may be further studied in the future.

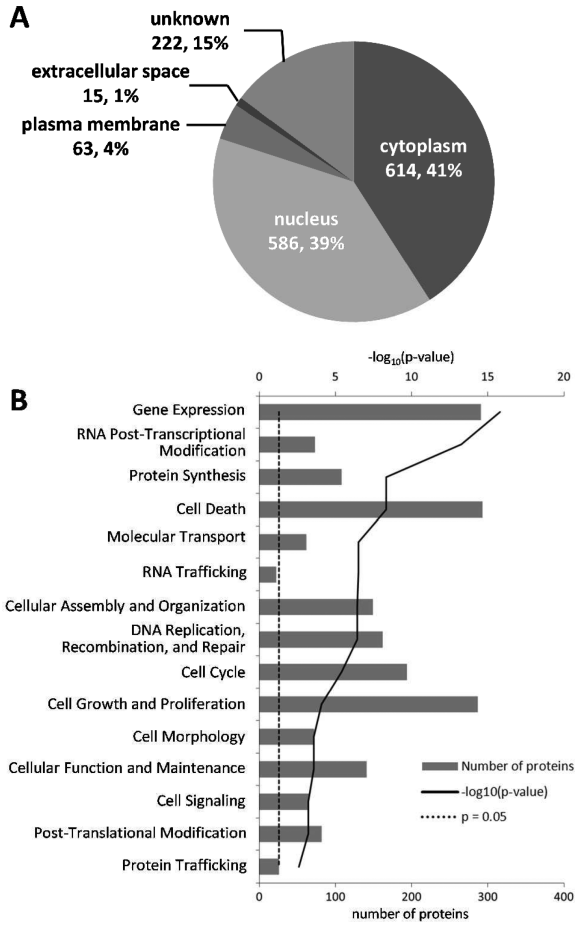


Figure 3. Subcellular localization and functional categories of identified *O*-GlcNAc proteins. (A) Presumed subcellular localization and (B) significantly enriched cellular functions as revealed by Ingenuity Pathways Analysis ($p < 0.05$). The complete list of overrepresented functional categories and pathways is provided in Table S4 and Table S5 (Supporting Information), respectively.

dx.doi.org/10.1021/pr300967y | *J. Proteome Res.* 2013, 12, 927−936

Identification of *O*-GlcNAc Sites

While the previous section provides good evidence for the selectivity of the developed *O*-GlcNAc protein enrichment method, this section addresses the identification of the corresponding modification sites based on selective β-elimination. Chemical β-elimination is a rather unselective procedure because it does not discriminate well between *O*-GlcNAc and phosphate moieties on Ser and Thr residues and can also affect unmodified residues and alkylated cysteins. This is further complicated by the frequent co-occurrence of both modifications and the approximately 10-fold higher abundance of phosphorylation over *O*-GlcNAcylation.[11,38] To discriminate between phosphorylation and *O*-GlcNAcylation, we introduced an on-resin dephosphorylation step between the on-resin proteolytic digest and the on-resin β-elimination. By way of the Click reaction, all peptides bound to the alkyne resin (ideally) should be *O*-GlcNAc modified but could also be additionally phosphorylated. The latter case could lead to misinterpretation of the data obtained after β-elimination, while removing the phosphate groups first largely eliminates this issue. To obtain an efficient as well as selective β-elimination procedure, we screened a broad range of β-elimination and Michael addition conditions with respect to reaction time, temperature, and Michael addition donors (Table S1, Supporting Information) using a synthetic reference peptide library comprising 72 *O*-GlcNAc peptides[15] and 48 phosphopeptides.[29] From this work, we concluded that the commercial GlycoProfile β-elimination kit represents a sound compromise between efficiency and selectivity as it enables efficient *O*-GlcNAc β-elimination, while maintaining a low number of false-positive identifications resulting from β-elimination of phosphosites (Table S1, Supporting Information). Importantly, we observed that BEMAD conditions often led to massive degradation of the protein backbone (Figure S2, Supporting Information), whereas this was not the case for the β-elimination kit. Unfortunately, we do not know the exact composition of the kit reagents, which precludes us from providing details on the reasons why. From this set of experiments, we also saw no advantage in adding a nucleophile after β-elimination as this step is not required for the detection of the modified amino acid by mass spectrometry (which can be done by detecting the dehydro form of the amino acid) and instead only increases the chemical complexity of the sample (e.g., as a result of incomplete or side reactions).

Following these results, we analyzed HEK293 cells and, overall, identified 635 *O*-GlcNAc spectra representing 185 *O*-GlcNAc sites on 80 proteins (triplicate analysis on an LTQ Orbitrap XL and a single analysis on an LTQ Orbitrap Velos platform; Tables S6–S8, Supporting Information). For 85 of the *O*-GlcNAc sites, the site could be determined with a localization probability of better than 90%, while for 100 sites, the former *O*-GlcNAc sites could not be mapped with certainty. Nearly 80% of the identified *O*-GlcNAc proteins and around 30% of the unambiguously identified sites have been reported previously, which is well within expectation and underpins the validity of the employed procedure. In addition, two significantly enriched sequence motifs have been identified (Figure S3, Supporting Information), including the known PVST and a similar VPTS motif.

We identified *O*-GlcNAc sites on a number of novel *O*-GlcNAc proteins (Table S6, Supporting Information). Probably the most noteworthy novel protein is the E3 ubiquitin-protein ligase CBL. CBL is a negative regulator of multiple receptor tyrosine kinase signaling pathways and a reported oncogene.[47] It has been found as *O*-GlcNAc modified at Ser-601.

Even though the methods reported in this study led to a significant number of *O*-GlcNAc sites identified from a single cell line, there still is a striking discrepancy between the number of identified *O*-GlcNAc proteins and the corresponding modification sites. While the on-resin digest revealed 1535 high-confidence *O*-GlcNAc proteins (PPV >0.99), site evidence was obtained only for 80 proteins. These 80 proteins are, by and large, among the most abundant *O*-GlcNAc proteins (Figure S4, Supporting Information) indicating that the identification of CuAAC-captured *O*-GlcNAc proteins can be readily accomplished, while the identification of their *O*-GlcNAc sites is rather challenging. We identified several general and experiment-specific reasons that may explain the observed bias. Clearly, the identification of resin-bound *O*-GlcNAc proteins is strongly facilitated by the large number and diversity of tryptic peptides generated from intact *O*-GlcNAc proteins. While any of the peptides support an identification, the modification site assignment clearly requires the detection of the specific modified peptide. In complex samples, the chance of this happening is likely less than 10–20%. In addition, azide-tagged *O*-GlcNAc proteins may be conjugated to the alkyne resin via any one of their *O*-GlcNAc sites but very unlikely by all sites. While this would lead to a stoichiometry of at least one for the captured protein, the stoichiometry of a particular captured *O*-GlcNAc peptide will likely be fairly (or even very) low, rendering the identification of *O*-GlcNAc sites difficult. Furthermore, *O*-GlcNAc sites have been found to occur predominantly in regions with low compositional complexity,[11,38] suggesting that *O*-GlcNAc sites may often not be amenable for common proteomic workflows using trypsin. The use of alternative or multiple proteases can, therefore, add significant value to the analysis of *O*-GlcNAc sites. Last, but not least, the employed β-elimination procedure has been optimized for *O*-GlcNAc peptides in solution. However, the on-resin reaction is likely less efficient owing to kinetic or spatial imperfections. It appears that serine *O*-GlcNAc residues are more susceptible to β-elimination than threonine residues, which is consistent with a less acidic C_α–H. The serine/threonine ratio of unambiguous *O*-GlcNAc sites is around 4:1, while reported *O*-GlcNAc sites typically show 1:1 to 2:1 distribution.[11,35−41] We also note a significant degree of β-elimation of alkylated cysteine residues (1295 peptides), which, again, shows the susceptibility of alkylated Cys residue to β-elimination-based approaches.

Still, compared to previous approaches utilizing some form of β-elimination/Michael addition for the identification of *O*-GlcNAc proteins and sites,[10,21,22] our approach represents a considerable improvement. Because the reagents are commercially available, the methods should also be easily adaptable by other laboratories.

Global Identification of GlcNAcstatin G-Responsive Proteins

To demonstrate the practical utility of the developed *O*-GlcNAc protein enrichment procedure for the global analysis of *O*-GlcNAc proteins, we studied the effect of OGA inhibition on a proteome-wide scale using GlcNAcstatin G.[31] GlcNAcstatin G is a potent OGA inhibitor that exhibits selectivity over related lysosomal hexosaminidases and has been shown to induce hyper-*O*-GlcNAcylation of cellular proteins in the nanomolar range, but the spectrum of proteins actually

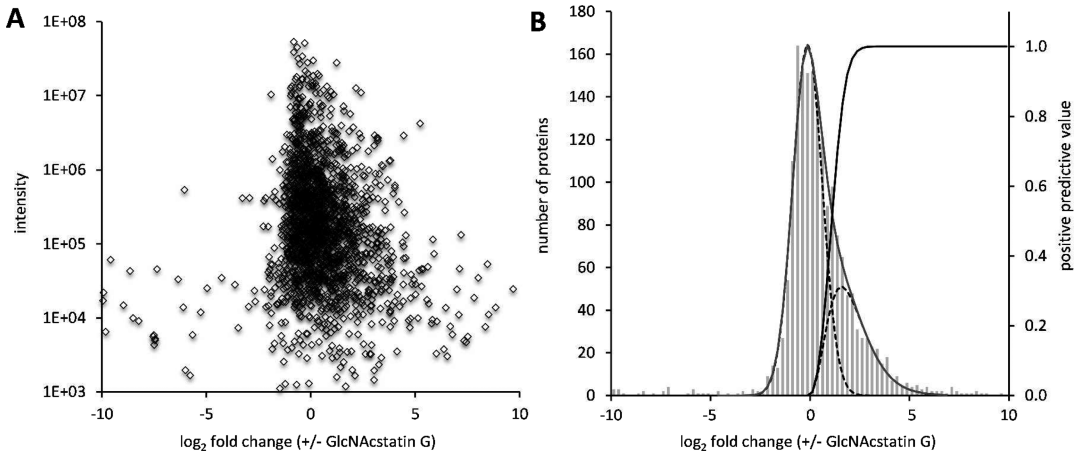


Figure 4. Global profiling of GlcNAcstatin G-responsive proteins. (A) Scatterplot of intensity and $\log_2$-fold change of identified *O*-GlcNAc proteins. (B) The distribution of $\log_2$ ratios (±GlcNAcstatin G) of identified *O*-GlcNAc proteins was used to calculate positive predictive values for GlcNAcstatin G-responsive proteins.

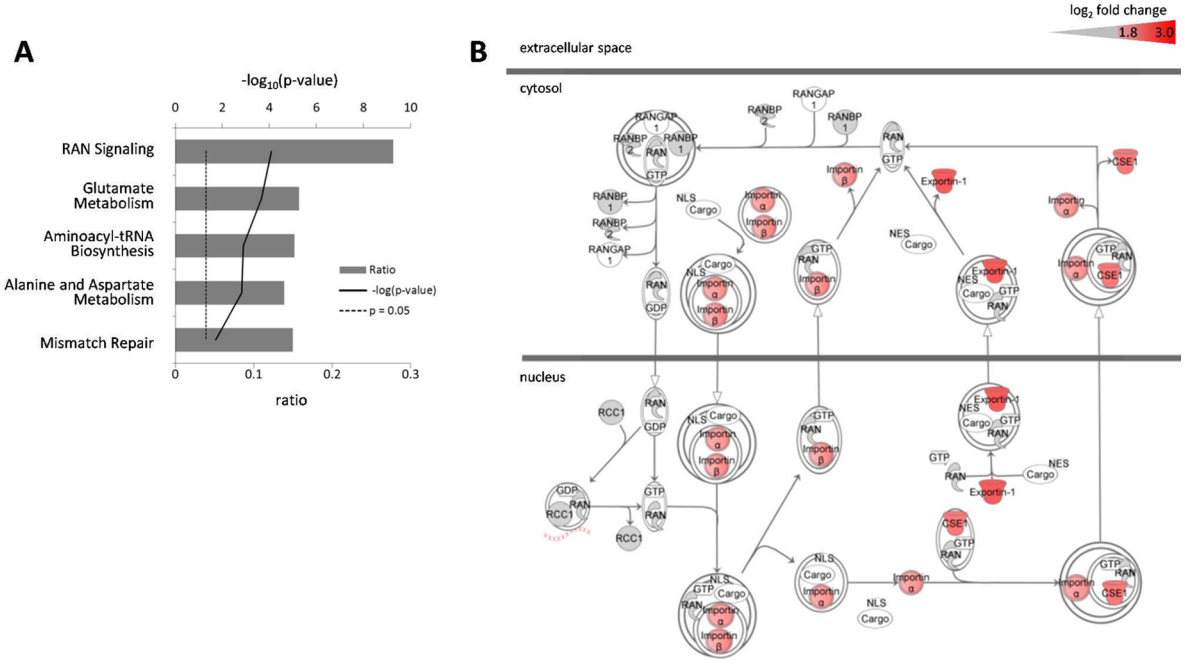


Figure 5. Cellular pathways responding to OGA inhibition by GlcNAcstatin G. (A) Significantly enriched biochemical and signaling pathways as revealed by Ingenuity Pathways analysis ($p < 0.05$). (B) RAN nuclear transport pathway. Almost all proteins of this pathway were found to be *O*-GlcNAc modified (gray), and several members of this cellular machinery show increased levels of *O*-GlcNAc upon OGA inhibition (red).

responding to this treatment has not yet been systematically determined. To do so, HEK293 cells were labeled with GlcNAz, treated with GlcNAcstatin G or vehicle control (DMSO) for two hours, and *O*-GlcNAc proteins were enriched and quantified as above. The selectivity of the enrichment was controlled using unlabeled HEK293 cells as negative control. In this experiment, analyzed on a LTQ Orbitrap XL, we identified 1675 proteins, of which 1031 can be considered high-confidence *O*-GlcNAc proteins (PPV > 0.99; Figure S5 and Table S9 and S10, Supporting Information). Note that, in this experiment, we calculated the PPV for *O*-GlcNAc modified proteins based on the OGA inhibitor treated sample. Otherwise, *O*-GlcNAc proteins with an initially low *O*-GlcNAc stoichiometry (and hence low biochemical enrichment factor) but significantly increased *O*-GlcNAcylation upon treatment would not be taken into account, thus leaving out a very relevant group of GlcNAcstatin G-responsive proteins.

Following a two hour drug treatment, a significant number of proteins exhibit clearly increased *O*-GlcNAc levels (Figure 4A). To obtain a reliable list of affected proteins, we calculated an additional PPV for GlcNAcstatin G-responsive proteins using a mixture model assuming a Gaussian distribution around zero for unaffected proteins and a gamma distribution for GlcNAcstatin G-responsive proteins. This mixture model can be rationalized as follows: the intensity of unaffected proteins reflects solely biological and technical variation that can be approximated by a Gaussian distribution. In contrast, the intensity of GlcNAcstatin G-responsive proteins also increases to a varying extent upon OGA inhibition. To account for the inherent skewness of such a distribution, we modeled the distribution of affected proteins using a right-tailed gamma

dx.doi.org/10.1021/pr300967y | *J. Proteome Res.* 2013, 12, 927−936

distribution. Overall, this model provides a good fit for the observed right-tailed distribution of fold changes on a logarithmic scale (Figure 4B). Following this rationale, 187 *O*-GlcNAc proteins were affected by the drug treatment (PPV > 0.9).

As depicted in Figure S6, Supporting Information, the overlap to the global *O*-GlcNAc proteome data set is fairly large. Nearly 60% of the *O*-GlcNAc proteins have been identified in the previous data set, while 434 proteins have only been identified in this experiment. Among these are 92 proteins, which are strongly GlcNAcstatin G-responsive (PPV > 0.9). In contrast, more than 900 proteins have been exclusively identified in the global *O*-GlcNAc proteome data set, which is clearly due to the higher sensitivity of the LTQ Orbitrap Velos utilized in that study.

Overall, the fact that less than 20% of all high confidence *O*-GlcNAc proteins were GlcNAcstatin G-responsive suggests that most of these proteins are already modified to a high stoichiometry. This is consistent with recent findings that *O*-GlcNAc sites of the most abundant *O*-GlcNAc proteins exhibit a relative site occupancy of around 90%[9] and also with in vitro experiments showing that some OGT substrates can be constitutively modified.[48] An important technical aspect of this experiment is that the identification of 187 GlcNAcstatin G-responsive proteins further underscores the validity of the developed *O*-GlcNAc enrichment approach.

GlcNAcstatin G-Responsive Pathways and Selected Proteins

Biological functions and cellular pathways significantly over-represented among the GlcNAcstatin G-responsive proteins are depicted in Figure 5A and S7 as well as in Table S11 and S12 in the Supporting Information. Interestingly, there are several key players involved in nutrient responsive *O*-GlcNAc cycling, some of which have not been reported as *O*-GlcNAc modified before. Several metabolic enzymes, most of which catalyze committed steps of anabolic pathways requiring glutamine as amino donor, were identified to be OGA inhibitor-responsive. Notably, we identified glutamine-fructose-6-phosphate transaminase (GFPT), which controls the flux of glucose into the HBP and, eventually, the intracellular level of UDP-GlcNAc. AMP-activated protein kinase (AMPK) also showed increased *O*-GlcNAc levels following inhibitor treatment. AMPK is the central hub in energy responsive AMPK signaling and down-regulates multiple anabolic pathways upon energy deprivation.[49,50] This is consistent with previous results that an increased flux through the HBP by overexpression of GFPT results in increased AMPK phosphorylation and *O*-GlcNAcylation leading to the activation of energy replenishing pathways.[51] Another drug responsive protein is the deacetylase Sirtuin-1. Interestingly, it has been hypothesized that *O*-GlcNAcylation and sirtuin-dependent deacetylation may exert opposing functions in situations of nutrient excess or starvation.[45] A striking finding is that almost all members of the RAN-dependent nuclear transport system, which mediates the transport of proteins, tRNAs, and ribosomal subunits across the nuclear membrane, are modified by *O*-GlcNAc and that several of the key proteins show considerable increase in *O*-GlcNAc levels upon OGA inhibition (Figure 5B). *O*-GlcNAc residues on nuclear transport factors have been associated with important recognition events in nuclear transport.[52] It is, therefore, tempting to speculate that RAN signaling may represent an additional module of the nutrient responsive *O*-GlcNAc cycling system. Clearly, further work is required to identify the site(s) of modification on these proteins as well as and their potential functional significance.

CONCLUSIONS

In this article, we demonstrate the refinement and validation of a recently introduced method for the enrichment of *O*-GlcNAc modified proteins based on metabolic labeling, Click chemistry, and quantitative mass spectrometry that represents a significant improvement over, and a useful complement to, existing methods. Although most *O*-GlcNAc sites remain undetectable, the approach routinely enables the identification of >1000 modified proteins, which opens up many lines of investigation into the cellular role of this emerging post-translational protein modification. Because the reagents are all commercially available, the approach should be readily adoptable by other laboratories and may even be combined with existing *O*-GlcNAc enrichment approaches such as chemoenzymatic *O*-GlcNAc tagging. Indeed, the reach of the method may extend even further to any type of azide-tagged protein or modification thereof.

ASSOCIATED CONTENT

Supporting Information

Additional experimental procedures, figures, and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: kuster@tum.de. Tel: +49 8161 715696. Fax: +49 8161 715931.

Present Address

[⊥]Department of Biology, Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule (ETH) Zürich, Zürich, Switzerland.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge the Studienstiftung des deutschen Volkes e. V. for a Ph.D. fellowship to H.H., and the support of the Faculty Graduate Center Weihenstephan of TUM Graduate School at the Technische Universität München, Germany.

ABBREVIATIONS

BEMAD, β-elimination followed by Michael addition; CID, collision-induced dissociation; CuAAC, copper-catalyzed azide/alkyne Click chemistry; ECD, electron capture dissociation; ETD, electron transfer dissociation; HBP, hexosamine biosynthetic pathway; MS, mass spectrometry; LC−MS/MS, liquid chromatography coupled to tandem mass spectrometry; *O*-GlcNAc, O-linked β-*N*-acetylglucosamine; PTM, post-translational modification; UDP-GlcNAc, uridine diphosphate *N*-acetylglucosamine

REFERENCES

(1) Torres, C. R.; Hart, G. W. 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* (5), 3308−3317.

(2) Love, D. C.; Hanover, J. A. The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Sci. STKE* **2005**, *2005* (312), re13.
(3) Hart, G. W.; Housley, M. P.; Slawson, C. Cycling of O-linked beta-*N*-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **2007**, *446* (7139), 1017−1022.
(4) Hart, G. W.; Slawson, C.; Ramirez-Correa, G.; Lagerlof, O. Cross talk between *O*-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu. Rev. Biochem.* **2011**, *80*, 825−858.
(5) Toleman, C.; Paterson, A. J.; Whisenhunt, T. R.; Kudlow, J. E. Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable *O*-GlcNAcase and HAT activities. *J. Biol. Chem.* **2004**, *279* (51), 53665−53673.
(6) Wells, L.; Gao, Y.; Mahoney, J. A.; Vosseller, K.; Chen, C.; Rosen, A.; Hart, G. W. Dynamic *O*-glycosylation of nuclear and cytosolic proteins: further characterization of the nucleocytoplasmic beta-*N*-acetylglucosaminidase, *O*-GlcNAcase. *J. Biol. Chem.* **2002**, *277* (3), 1755−1761.
(7) Haltiwanger, R. S.; Grove, K.; Philipsberg, G. A. Modulation of O-linked *N*-acetylglucosamine levels on nuclear and cytoplasmic proteins in vivo using the peptide *O*-GlcNAc-beta-N-acetylglucosaminidase inhibitor *O*-(2-acetamido-2-deoxy-D-glucopyranosylidene)-amino-*N*-phenylcarbamate. *J. Biol. Chem.* **1998**, *273* (6), 3611−7.
(8) Macauley, M. S.; Vocadlo, D. J. Increasing *O*-GlcNAc levels: An overview of small-molecule inhibitors of *O*-GlcNAcase. *Biochim. Biophys. Acta* **2010**, *1800* (2), 107−121.
(9) Hahne, H.; Gholami, A. M.; Kuster, B. Discovery of *O*-GlcNAc-modified proteins in published large-scale proteome data. *Mol. Cell. Proteomics* **2012**, *11* (10), 843−850.
(10) Vosseller, K.; Trinidad, J. C.; Chalkley, R. J.; Specht, C. G.; Thalhammer, A.; Lynn, A. J.; Snedecor, J. O.; Guan, S.; Medzihradszky, K. F.; Maltby, D. A.; Schoepfer, R.; Burlingame, A. L. O-linked *N*-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol. Cell. Proteomics* **2006**, *5* (5), 923−934.
(11) Trinidad, J. C.; Barkan, D. T.; Gulledge, B. F.; Thalhammer, A.; Sali, A.; Schoepfer, R.; Burlingame, A. L. Global identification and characterization of both *O*-GlcNAcylation and phosphorylation at the murine synapse. *Mol. Cell. Proteomics* **2012**, *11* (8), 215−229.
(12) Wang, Z.; Udeshi, N. D.; O'Malley, M.; Shabanowitz, J.; Hunt, D. F.; Hart, G. W. Enrichment and site mapping of O-linked *N*-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. *Mol. Cell. Proteomics* **2010**, *9* (1), 153−160.
(13) Alfaro, J. F.; Gong, C. X.; Monroe, M. E.; Aldrich, J. T.; Clauss, T. R.; Purvine, S. O.; Wang, Z.; Camp, D. G., II; Shabanowitz, J.; Stanley, P.; Hart, G. W.; Hunt, D. F.; Yang, F.; Smith, R. D. Tandem mass spectrometry identifies many mouse brain *O*-GlcNAcylated proteins including EGF domain-specific *O*-GlcNAc transferase targets. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (19), 7280−7285.
(14) Mirgorodskaya, E.; Roepstorff, P.; Zubarev, R. A. Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal. Chem.* **1999**, *71* (20), 4431−4436.
(15) Hahne, H.; Kuster, B. A novel two-stage tandem mass spectrometry approach and scoring scheme for the identification of *O*-GlcNAc modified peptides. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (5), 931−942.
(16) Vocadlo, D. J.; Hang, H. C.; Kim, E. J.; Hanover, J. A.; Bertozzi, C. R. A chemical approach for identifying *O*-GlcNAc-modified proteins in cells. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100* (16), 9116−9121.
(17) Sprung, R.; Nandi, A.; Chen, Y.; Kim, S. C.; Barma, D.; Falck, J. R.; Zhao, Y. Tagging-via-substrate strategy for probing *O*-GlcNAc modified proteins. *J. Proteome Res.* **2005**, *4* (3), 950−957.
(18) Nandi, A.; Sprung, R.; Barma, D. K.; Zhao, Y.; Kim, S. C.; Falck, J. R. Global identification of *O*-GlcNAc-modified proteins. *Anal. Chem.* **2006**, *78* (2), 452−458.
(19) Gurcel, C.; Vercoutter-Edouart, A. S.; Fonbonne, C.; Mortuaire, M.; Salvador, A.; Michalski, J. C.; Lemoine, J. Identification of new *O*-GlcNAc modified proteins using a click-chemistry-based tagging. *Anal. Bioanal. Chem.* **2008**, *390* (8), 2089−2097.
(20) Zaro, B. W.; Yang, Y. Y.; Hang, H. C.; Pratt, M. R. Chemical reporters for fluorescent detection and identification of *O*-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4−1. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (20), 8146−8151.
(21) Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell. Proteomics* **2002**, *1* (10), 791−804.
(22) Vosseller, K.; Hansen, K. C.; Chalkley, R. J.; Trinidad, J. C.; Wells, L.; Hart, G. W.; Burlingame, A. L. Quantitative analysis of both protein expression and serine/threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* **2005**, *5* (2), 388−398.
(23) Li, W.; Backlund, P. S.; Boykins, R. A.; Wang, G.; Chen, H. C. 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* (1), 94−102.
(24) Herbert, B.; Hopwood, F.; Oxley, D.; McCarthy, J.; Laver, M.; Grinyer, J.; Goodall, A.; Williams, K.; Castagna, A.; Righetti, P. G. Beta-elimination: an unexpected artefact in proteome analysis. *Proteomics* **2003**, *3* (6), 826−831.
(25) McLachlin, D. T.; Chait, B. T. Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal. Chem.* **2003**, *75* (24), 6826−3686.
(26) Hart, C.; Chase, L. G.; Colquhoun, D.; Agnew, B. In *Identification and Characterization of O-GlcNAc Modification of Galectin-1 in Mesenchymal Stem Cells Using Click Chemistry*, Annual Conference of the Society for Glycobiology, St. Pete Beach, FL, 2010, Society for Glycobiology.
(27) Nyberg, T.; Qian, X.; Slade, P.; Huang, W.; Agnew, B. In *Universal Click Chemistry-Based Enrichment of Multiple PTM Subclasses Coupled with Rapid Modification Site Identification*, 58th ASMS Conference on Mass Spectrometry and Allied Topics, Salt Lake City, UT, 2010.
(28) Hart, C.; Chase, L. G.; Hajivandi, M.; Agnew, B. Metabolic labeling and click chemistry detection of glycoprotein markers of mesenchymal stem cell differentiation. *Methods Mol. Biol.* **2011**, *698*, 459−484.
(29) Savitski, M. M.; Lemeer, S.; Boesche, M.; Lang, M.; Mathieson, T.; Bantscheff, M.; Kuster, B. Confident phosphorylation site localization using the Mascot Delta Score. *Mol. Cell. Proteomics* **2011**, *10* (2), M110 003830.
(30) Rappsilber, J.; Mann, M.; Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2007**, *2* (8), 1896−1906.
(31) Dorfmueller, H. C.; Borodkin, V. S.; Schimpl, M.; Zheng, X.; Kime, R.; Read, K. D.; van Aalten, D. M. Cell-penetrant, nanomolar *O*-GlcNAcase inhibitors selective against lysosomal hexosaminidases. *Chem. Biol.* **2010**, *17* (11), 1250−1255.
(32) Kall, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4* (11), 923−925.
(33) Brosch, M.; Yu, L.; Hubbard, T.; Choudhary, J. Accurate and sensitive peptide identification with Mascot Percolator. *J. Proteome Res.* **2009**, *8* (6), 3176−3181.
(34) Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **2006**, *24* (10), 1285−1292.
(35) Wang, Z.; Udeshi, N. D.; Slawson, C.; Compton, P. D.; Sakabe, K.; Cheung, W. D.; Shabanowitz, J.; Hunt, D. F.; Hart, G. W. Extensive crosstalk between *O*-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci. Signaling* **2010**, *3* (104), ra2.

(36) Wang, J.; Torii, M.; Liu, H.; Hart, G. W.; Hu, Z. Z. dbOGAP: an integrated bioinformatics resource for protein *O*-GlcNAcylation. *BMC Bioinf.* **2011**, *12* (1), 91.
(37) Hornbeck, P. V.; Kornhauser, J. M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **2012**, *40* (database issue), D261−D270.
(38) Hahne, H.; Gholami, A. M.; Kuster, B. Discovery of *O*-GlcNAc-modified proteins in published large-scale proteome data. *Mol. Cell. Proteomics* **2012**, *11* (10), 843−850.
(39) Myers, S. A.; Panning, B.; Burlingame, A. L. Polycomb repressive complex 2 is necessary for the normal site-specific *O*-GlcNAc distribution in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (23), 9490−9495.
(40) Chalkley, R. J.; Thalhammer, A.; Schoepfer, R.; Burlingame, A. L. Identification of protein *O*-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (22), 8894−8899.
(41) Zhao, P.; Viner, R.; Teo, C. F.; Boons, G. J.; Horn, D.; Wells, L. Combining high-energy C-trap dissociation and electron transfer dissociation for protein *O*-GlcNAc modification site assignment. *J. Proteome Res.* **2011**, *10* (9), 4088−4104.
(42) Schwartz, D.; Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **2005**, *23* (11), 1391−1398.
(43) Speers, A. E.; Cravatt, B. F. Profiling enzyme activities in vivo using click chemistry methods. *Chem. Biol.* **2004**, *11* (4), 535−546.
(44) Hu, P.; Shimoji, S.; Hart, G. W. Site-specific interplay between *O*-GlcNAcylation and phosphorylation in cellular regulation. *FEBS Lett.* **2010**, *584* (12), 2526−2538.
(45) Hanover, J. A.; Krause, M. W.; Love, D. C. The hexosamine signaling pathway: *O*-GlcNAc cycling in feast or famine. *Biochim. Biophys. Acta* **2010**, *1800* (2), 80−95.
(46) Butkinaree, C.; Park, K.; Hart, G. W. O-linked beta-*N*-acetylglucosamine (*O*-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim. Biophys. Acta* **2009**, *1800* (2), 96−106.
(47) Lipkowitz, S.; Weissman, A. M. RINGs of good and evil: RING finger ubiquitin ligases at the crossroads of tumour suppression and oncogenesis. *Nat. Rev. Cancer* **2011**, *11* (9), 629−643.
(48) Shen, D. L.; Gloster, T. M.; Yuzwa, S. A.; Vocadlo, D. J. Insights into *O*-GlcNAc processing and dynamics through kinetic analysis of *O*-GlcNAc transferase and *O*-GlcNAcase activity on protein substrates. *J. Biol. Chem.* **2012**, *287* (19), 15395−15408.
(49) Carling, D.; Mayer, F. V.; Sanders, M. J.; Gamblin, S. J. AMP-activated protein kinase: nature's energy sensor. *Nat. Chem. Biol.* **2011**, *7* (8), 512−518.
(50) Hardie, D. G.; Ross, F. A.; Hawley, S. A. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat. Rev. Mol. Cell. Biol.* **2012**, *13* (4), 251−262.
(51) Luo, B.; Parker, G. J.; Cooksey, R. C.; Soesanto, Y.; Evans, M.; Jones, D.; McClain, D. A. Chronic hexosamine flux stimulates fatty acid oxidation by activating AMP-activated protein kinase in adipocytes. *J. Biol. Chem.* **2007**, *282* (10), 7172−7180.
(52) Yu, S. H.; Boyce, M.; Wands, A. M.; Bond, M. R.; Bertozzi, C. R.; Kohler, J. J. Metabolic labeling enables selective photocrosslinking of *O*-GlcNAc-modified proteins to their binding partners. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (13), 4834−4839.