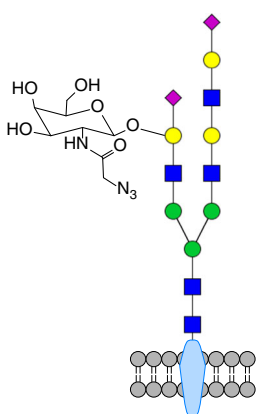


# Mass Spectrometric Analysis of the Cell Surface *N*-Glycoproteome by Combining Metabolic Labeling and Click Chemistry

Johanna M. Smeekens, Weixuan Chen, Ronghu Wu

School of Chemistry and Biochemistry and the Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332, USA



**Abstract.** Cell surface *N*-glycoproteins play extraordinarily important roles in cell–cell communication, cell–matrix interactions, and cellular response to environmental cues. Global analysis is exceptionally challenging because many *N*-glycoproteins are present at low abundances and effective separation is difficult to achieve. Here, we have developed a novel strategy integrating metabolic labeling, copper-free click chemistry, and mass spectrometry (MS)-based proteomics methods to analyze cell surface *N*-glycoproteins comprehensively and site-specifically. A sugar analog containing an azido group, *N*-azidoacetylglucosamine, was fed to cells to label glycoproteins. Glycoproteins with the functional group on the cell surface were then bound to dibenzocyclooctyne-sulfo-biotin via copper-free click chemistry under physiological conditions. After protein extraction and digestion, glycopeptides with the biotin tag were enriched by NeutrAvidin conjugated beads. Enriched glycopeptides were deglycosylated with peptide-*N*-glycosidase F in heavy-oxygen water, and in the process of glycan removal, asparagine was converted to aspartic acid and tagged with  $^{18}\text{O}$  for MS analysis. With this strategy, 144 unique *N*-glycopeptides containing 152 *N*-

glycosylation sites were identified in 110 proteins in HEK293T cells. As expected, 95% of identified glycoproteins were membrane proteins, which were highly enriched. Many sites were located on important receptors, transporters, and cluster of differentiation proteins. The experimental results demonstrated that the current method is very effective for the comprehensive and site-specific identification of the cell surface *N*-glycoproteome and can be extensively applied to other cell surface protein studies.

**Keywords:** Cell surface *N*-glycoproteome, Metabolic labeling, Click chemistry, MS-based proteomics, HEK293T cells

Received: 22 August 2014/Revised: 2 October 2014/Accepted: 3 October 2014/Published Online: 26 November 2014

## Introduction

With the advent of soft ionization sources, especially electrospray ionization (ESI) [1] and matrix-assisted laser desorption/ionization (MALDI) [2], mass spectrometry (MS) has become an increasingly powerful tool to investigate bio-macromolecules [3–8]. MS-based proteomics techniques offer a unique opportunity to systematically study proteins [9, 10], which is beyond the reach of conventional biochemical

methods. Classic antibody-based protein identification methods depend entirely on the availability and quality of antibodies. In addition, antibodies are expensive and the experimental procedures are typically time-consuming and labor-intensive. Alternatively, MS-based techniques can assist in the confident global identification and quantification of proteins without the use of antibodies [11–13].

Protein modifications are essential in biological systems and involved in nearly every cellular event [14–18]. There has been great interest in the comprehensive and quantitative analysis of modified proteins to obtain a better understanding of protein functions and disease mechanisms, which could lead to the discovery of effective biomarkers and drug targets [19–21]. Although MS-based proteomics techniques have been applied to globally identify modified proteins, pinpoint the

**Electronic supplementary material** The online version of this article (doi:10.1007/s13361-014-1016-7) contains supplementary material, which is available to authorized users.

Correspondence to: Ronghu Wu; e-mail: ronghu.wu@chemistry.gatech.edu

modification sites, and quantify the abundance changes [22–26], it is still extraordinarily challenging to achieve these goals in complex biological samples for multiple reasons [27, 28]. Many modifications are substoichiometric and occur on low-abundance proteins that only have several copies per cell, such as glycosylated receptor proteins. In addition, reversible modifications are very dynamic, and labile modification groups are not stable in the gas phase or during sample preparation. Finally, each modified group is structurally different; thus, there is no universal method that is suitable for the analysis of all protein modifications. Therefore, effective separation and enrichment methods are mandatory for the large-scale analysis of modified proteins.

Protein *N*-glycosylation is one of the most common and important modifications; it frequently initiates cell signal transduction and regulates cell–cell communication and cell–matrix interactions [29, 30]. Based on predictions and computational results, about half of all mammalian proteins are glycosylated at any given time [31]. Additionally, the vast majority of cell surface and secretome proteins are glycosylated. Aberrant glycosylation is often correlated with human diseases, including cancer and infectious diseases [21, 32]. Glycoproteins on the cell surface can serve as effective biomarkers for the early detection of disease and excellent drug targets for disease treatment because of their easy accessibility by small molecules and macromolecules. However, comprehensive identification of cell surface glycoproteins has not been thoroughly investigated, especially compared with intracellular proteins. Cell surface glycoprotein enrichment is even more challenging than typical proteomics experiments, which magnifies the difficulties associated with modified proteins mentioned above. Several years ago, a very elegant cell surface capturing method was reported for cell surface glycoprotein analysis [33]. Glycoproteins on living cells were oxidized, bound to biotin through a bifunctional linker molecule, biocytin hydrazide, and enriched by avidin for MS analysis. However, the main drawback of the method is that the oxidation reaction is not very fast or efficient and it can also act as an external stimulus to cells. Effective methods will profoundly advance the analysis of cell surface glycoproteins and provide insight into glycoprotein function.

In this work, we have developed an effective MS-based method to identify cell surface glycoproteins comprehensively and site-specifically. A sugar analog containing a biologically inert but chemically functional azido group was fed to cells to label cell surface glycoproteins according to a previous method [34]. Surface glycoproteins containing the functional group were subsequently bound to biotin through copper-free click chemistry under mild physiological conditions. Further separation and enrichment by exploiting the strong interaction between biotin and avidin allowed the global analysis of cell surface glycoproteins.

## Experimental

### *Cell Culture and Labeling*

HEK293T cells were grown in Dulbecco's modified Eagle medium (DMEM) containing 10% fetal bovine serum (FBS).

Once cells reached 10% confluency, medium was changed to DMEM containing 10% FBS and 50  $\mu$ M *N*-azidoacetylgalactosamine (GalNAz). Cells were incubated for 3 d until confluency was  $\sim$ 80%. After metabolic labeling, cells were washed with phosphate buffered saline (PBS) two times and 100  $\mu$ M dibenzocyclooctyne (DBCO)-sulfo-biotin in PBS was added into the culture flasks. Cells were incubated for 1 h with gentle agitation at 4°C and then harvested by scraping and centrifugation at 300g for 5 min. The supernatant was discarded and the cell pellet containing about  $4 \times 10^7$  cells was washed twice with PBS containing 10 mM dithiothreitol (DTT).

### *Cell Lysis and Membrane Protein Extraction*

The cell pellet was incubated in a buffer containing 150 mM NaCl, 50 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) (pH = 7.4), 25  $\mu$ g/mL digitonin, and Roche (Indianapolis, IN, USA) protease inhibitor (1 tablet per 10 mL) with end-over-end rotation for 10 min at 4°C. After incubation, samples were centrifuged at 2000g for 10 min. Cell pellets were washed with the buffer twice and subsequently lysed with the MiniBeadbeater (Biospec, Bartlesville, OK, USA) in a buffer containing 10 mM HEPES, 1.5 mM MgCl<sub>2</sub>, and 10 mM KCl. The resulting solutions were centrifuged at 2500g for 10 min, and the supernatant was collected and centrifuged at 16,000g for 30 min. The membrane-rich pellet was collected and washed two times with the lysis buffer. The pellet was further incubated in 0.1 M sodium carbonate solution containing 1 mM ethylenediaminetetraacetic acid (EDTA) on ice for 30 min, followed by centrifugation at 16,000g for 15 min. Sodium carbonate incubations and subsequent centrifugations were repeated once. The membrane rich pellets were incubated with shaking in a buffer containing 4 M urea, 100 mM NaCl, 10 mM HEPES, and 1 mM EDTA for 30 min at room temperature, and then samples were centrifuged at 16,000g for 15 min. The urea buffer wash was repeated once. Solubilization buffer containing 100 mM PBS and 1% NP-40 was added to cell pellets and incubated end-over-end overnight at room temperature.

### *Glycopeptide Preparation and Enrichment*

Solubilized samples were centrifuged at 16,000g for 15 min and supernatants were collected. Disulfide bonds within proteins were reduced with 5 mM DTT (56°C, 25 min) and subsequently alkylated with 14 mM iodoacetamide (room temperature, 30 min in the dark). After reduction and alkylation, proteins were purified with the methanol chloroform protein precipitation method [35]. Four volumes of methanol, one volume of chloroform and three volumes of water were added to one volume of the protein sample, and the mixture was vortexed. The sample was centrifuged at 5000g for 20 min. The proteins remained at the phase boundary between the methanol and chloroform layers. The methanol layer above the sample was removed. Four volumes of methanol were added, and the mixture was vortexed again. The sample was

centrifuged again at 5000g for 15 min. The supernatant was removed without disturbing the pellet, and the pellet was dried.

The resulting ~2 mg protein samples were digested overnight at a protein:trypsin ratio of ~100:1 in 2 mL of buffer containing 50 mM HEPES (pH = 8.5), 0.1 M urea, and 5% ACN. The next day, the digestion was quenched with trifluoroacetic acid (TFA) and purified on a 200 mg Sep-Pak C18 cartridge. Purified samples were dried and enriched with 200  $\mu$ L of NeutrAvidin bead slurry end-over-end for 30 min at 37°C. After enrichment, the peptide sample was transferred to a spin column and beads were washed 10 times with 400  $\mu$ L PBS and once with water. Peptides were eluted from the beads twice by 2-min incubations with 200  $\mu$ L of 8 M guanidine (pH = 1.5) at 56°C. Combined eluates were purified on a 50 mg Sep-Pak C18 cartridge. Enriched peptides were dried thoroughly before enzymatic deglycosylation with eight units of peptide-*N*-glycosidase F (PNGase F; Sigma-Aldrich, St. Louis, MO, USA) in 40  $\mu$ L of buffer containing 40 mM  $\text{NH}_4\text{HCO}_3$  in  $\text{H}_2^{18}\text{O}$  for 3 h at 37°C. The deglycosylation reaction was quenched with formic acid (FA) and purified with the stage tip method. Proteins were eluted into three samples using 20%, 50%, and 80% ACN each containing 1% HOAc.

### LC-MS/MS Analysis

Purified samples were dried and resuspended in a solvent containing 5% ACN and 4% FA, and 4  $\mu$ L were loaded onto a C18-packed microcapillary column (Magic C18AQ, 5  $\mu$ m, 200  $\text{\AA}$ , 100  $\mu\text{m}\times 16$  cm) using a WPS-3000TPL RS autosampler (Thermostatted Pulled Loop Rapid Separation Nano/Capillary Autosampler; Dionex, Sunnyvale, CA, USA). Peptides were separated by reversed-phase chromatography using an UltiMate 3000 binary pump with a 110 min gradient that varied for three samples. The first sample had a gradient of 4%–25% ACN (0.125% FA), the second sample's gradient was 10%–38% ACN (0.125% FA), and the third sample had a gradient of 15%–50% ACN (0.125% FA). Samples were detected in a hybrid dual-cell quadrupole linear ion trap-Orbitrap mass spectrometer (LTQ Orbitrap Elite; ThermoFisher, Waltham, MA, USA) using a data-dependent Top 20 method. Each cycle included one full MS scan (resolution: 60,000) in the Orbitrap at the automatic gain control (AGC) target of 1 million, followed by up to 20 MS/MS of the most intense ions in the LTQ. Selected ions were excluded from being further sequenced for 90 s. Ions with a single or unassigned charge were not fragmented. Maximum ion accumulation times were 1000 ms for each full MS scan and 50 ms for MS/MS scans.

### Database Searches and Data Filtering

The raw files recorded by MS were converted into mzXML format. Precursors for MS/MS fragmentation were checked for incorrect monoisotopic peak assignments while refining precursor ion mass measurements. The SEQUEST algorithm [36] (ver. 28) was used to search and match all MS/

MS spectra against a database encompassing sequences of all proteins in the Uniprot Human (*Homo sapiens*) Database (downloaded in February 2014) containing common contaminants such as keratins. Each protein sequence was listed in both forward and reverse orientations to estimate the false discovery rate (FDR) of peptide identification [37, 38]. A 20 ppm precursor mass tolerance and 1.0 Da product ion mass tolerance were used in the database search. Other selected parameters were full tryptic digestion, up to two missed cleavages, variable modifications including oxidation of methionine (+15.9949) and  $^{18}\text{O}$  tag of Asn (+2.9883), and fixed modifications including carbamidomethylation of cysteine (+57.0214).

In order to evaluate and further control FDRs of glycopeptide identification, the target-decoy method was employed [37, 38]. Linear discriminant analysis (LDA) was utilized to distinguish correct and incorrect peptide identifications using numerous parameters, including XCorr,  $\Delta\text{Cn}$ , and precursor mass error [39]. Separate linear discriminant models were trained for each raw file using forward and reverse peptide sequences to provide positive and negative training data. This approach is similar to other methods in the literature [40, 41]. After scoring, peptides less than six amino acids in length were discarded and peptide spectral matches were filtered to a less than 1% FDR based on the number of decoy sequences in the final data set. The dataset was restricted to glycopeptides when determining FDRs.

### Glycosylation Site Localization

To localize glycosylation sites and obtain a level of confidence corresponding to the identification, we applied a probabilistic algorithm [42] that considers all potential glycosylation sites on a peptide and uses the presence or absence of experimental fragment ions unique to each to obtain a ModScore. The ModScore, which is similar to Ascore [42], indicates the likelihood that the best site match is correct compared with the next best match. If only one glycosylation site is possible, a value of 1000 is assigned to the site. We considered sites with a score  $\geq 19$  ( $P \leq 0.01$ ) to be confidently localized.

## Results and Discussion

### Labeling and Enrichment Method

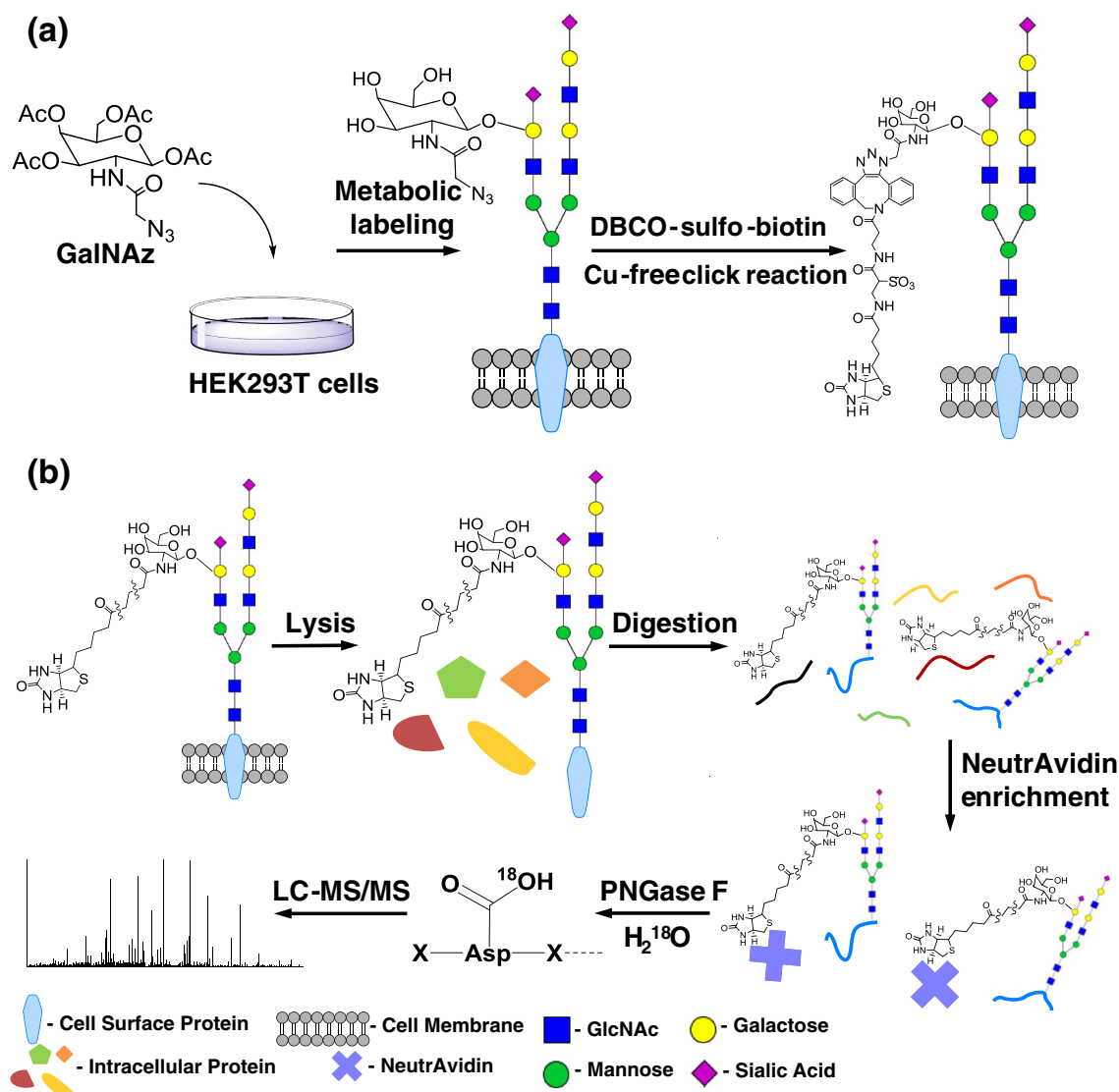
Sugar analogs have frequently been used to discover glycotransferase inhibitors [43]. Some sugar analogs can be used by glycosyltransferases to modify glycans within mammalian proteins. By taking advantage of this discovery, scientists have extensively investigated culturing cells with sugar analogs bearing various functional groups, including azido, alkyl, and aldehyde groups [34, 44–46]. The azido functional group in the sugar analog serves as a chemical handle for the click chemistry reaction, which will provide further insight into protein function and cellular activities.

In recent years, metabolic labeling has been employed for *in vivo* imaging experiments and valuable information has been obtained regarding the location of the sugar analogs and their relative abundance changes based on the fluorescence signal and corresponding intensity changes [45, 47]. In previous studies, GalNAz has been used in mammalian cells to modify proteins, resulting in azido-containing proteins. In this experiment, this strategy was applied to label glycoproteins that contain *N*-acetylgalactosamine (GalNAc). GalNAz was added into DMEM and incubated with HEK293T cells in order to metabolically label glycoproteins [48].

Cell surface glycoproteins with the functional azido group were selectively bound to biotin through copper-free click chemistry, as shown in Figure 1a. The copper-free click chemistry reaction between DBCO and the functional azido group is

very specific, rapid, and efficient [49, 50]. More importantly, the reaction can occur under physiological conditions, which allows cell surface glycoproteins on living cells to be labeled and minimizes external stimuli. After metabolic labeling, cells were lysed and proteins were extracted and digested. Based on the strong interaction between biotin and avidin, glycopeptides containing biotin were selectively enriched by incubation with NeutrAvidin conjugated agarose beads. After incubation, the beads were washed to remove non-biotinylated peptides, and enriched glycopeptides were subsequently eluted from the beads.

Enriched glycopeptides were dried for at least 24 h in a vacuum concentrator and then treated with PNGase F in heavy-oxygen water. This enzyme cleaves *N*-glycans and converts asparagine (Asn) to aspartic acid (Asp) in the process. The deamination of Asn can also occur *in vivo* and during sample



**Figure 1.** Principle of the cell surface glycoprotein enrichment method including (a) the metabolic labeling and click chemistry reaction, and (b) glycopeptide separation and analysis

preparation. By performing this enzymatic deglycosylation in heavy-oxygen water, newly formed Asp is labeled with  $^{18}\text{O}$ , which helps distinguish bona fide glycosylation sites from those caused by non-enzymatic deamination [51, 52]. As a result, *N*-glycopeptides and their corresponding glycosylation sites were confidently identified and localized.

### Glycopeptide Identification and Site Localization

Using this novel strategy, 144 unique *N*-glycosylated peptides were identified. Figure 2 shows the tandem mass spectra of three identified *N*-glycopeptides, EN#TSDPSLVIAFGR, GHTLTLN#FTR, and YSVQLMSFVYN#LSDTHLFPN#ASSK (# denotes the glycosylation site) from lysosome-associated membrane glycoprotein 1 (Lamp1), which is a single-pass type I membrane protein. It presents carbohydrate ligands to selectins and is involved in tumor cell metastasis [53]. These peptides were confidently identified with XCorr values of 4.11, 2.45, and 4.50, respectively. The high mass accuracy of each identification is represented by their ppm values of 0.84, 0.14, and  $-0.16$ . Corresponding glycosylation sites were confidently localized at N84, N103, N121, and N130, with ModScore values of 1000 for each. Additionally, all sites contain the consensus motif Asn-X-Ser/Thr, where X is any amino acid residue except proline.

The high resolution and mass accuracy of the Orbitrap mass spectrometer allowed for very confident *N*-glycopeptide identifications. The mass accuracy distribution of all identified *N*-glycopeptides is displayed in Figure 3a. Clearly, the vast majority of identified glycopeptides have a mass accuracy of less than 3 ppm. The mass accuracy results show the high confidence level associated with glycopeptide identifications. This is further confirmed by the XCorr values, which are shown in Figure 3b. Most of the XCorr values are greater than 3, and only several of them are less than 2. We manually checked glycopeptide identifications with relatively low XCorr values, and all of them were relatively short peptides whose fragments matched very well with corresponding theoretical spectra. The low XCorr values assigned to short peptides are due to the fact that the peptide matches are normally biased for longer peptides, i.e., higher XCorr values for longer peptides, since there are more potential fragments and, therefore, more possible matches [54].

Based on the same probability principle of Ascore [42], ModScore evaluates the confidence associated with each site localization [55, 56]. A value greater than 19 indicates a higher than 99% confidence level associated with the site localization, and a score greater than 13 corresponds to a higher than 95% confidence level. Among all the identified sites, 87% have a ModScore greater than 19, and only a very small portion of them (8%) have a ModScore less than 13, as shown in Figure 4c.

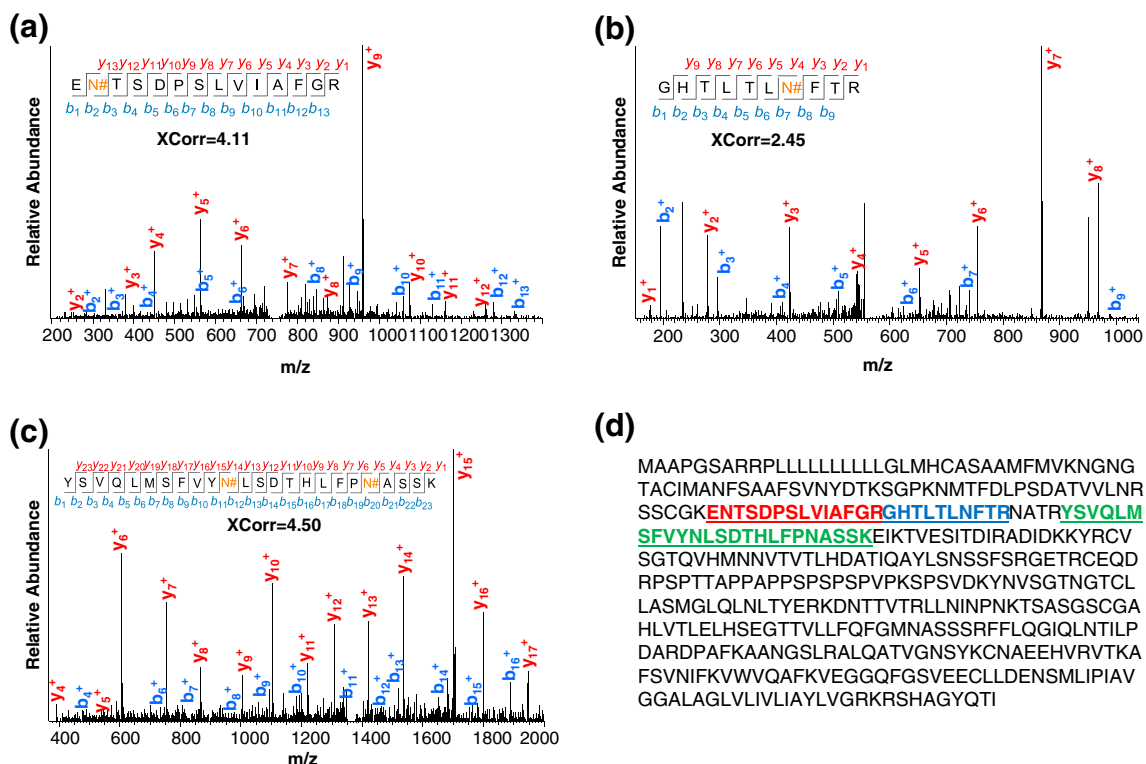
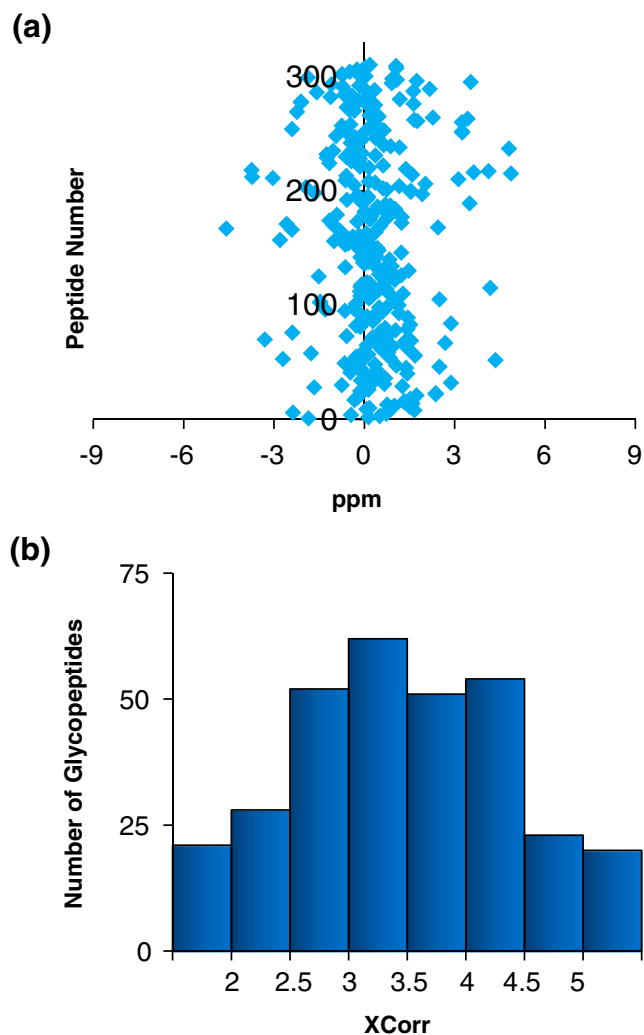
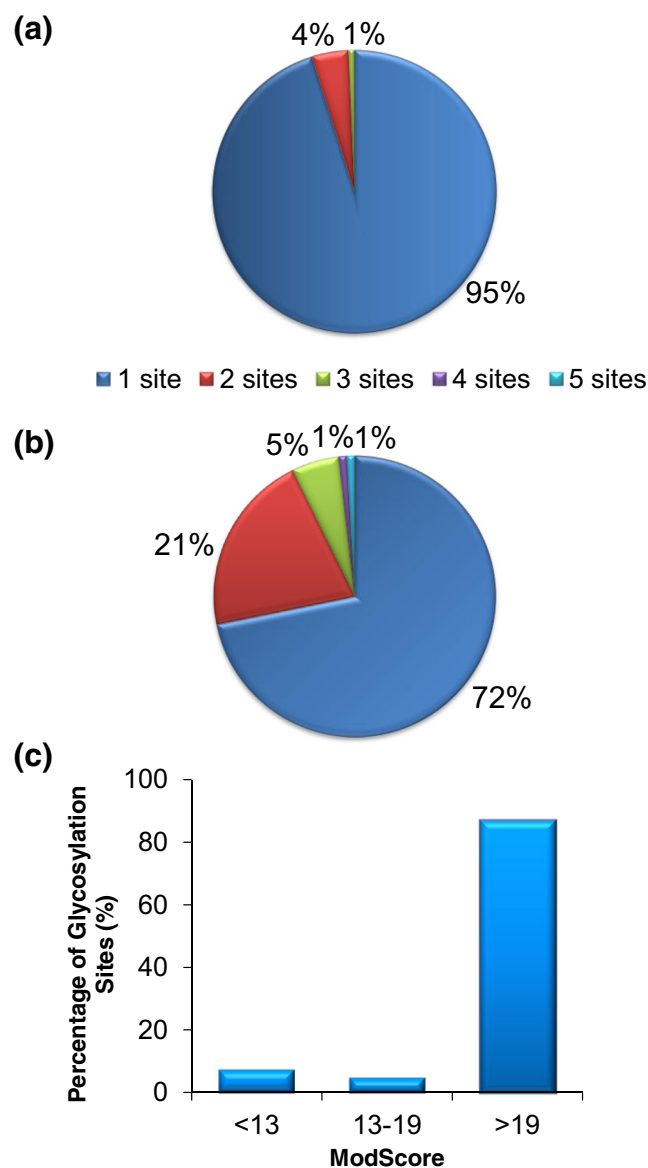


Figure 2. Tandem mass spectra of three peptides from the Lamp1 protein, including (a) EN#TSDPSLVIAFGR, (b) GHTLTLN#FTR, and (c) YSVQLMSFVYN#LSDTHLFPN#ASSK (# denotes the glycosylation site). The complete protein sequence with the highlighted identified peptides is shown in (d)



**Figure 3.** Distribution of (a) ppm and (b) XCorr values assigned for each glycopeptide identification. The bin size for XCorr analysis is 0.5

Overall, 152 sites were identified on 110 cell surface glycoproteins in HEK293T cells, which are listed in Supplemental Table 1. The major advantage of the current method is that cell surface glycoproteins containing the azido group can be labeled under mild physiological conditions without toxic heavy metal ions (Cu(I) and Cu(II)) that are frequently used as a catalyst in the traditional click chemistry reaction, copper-catalyzed azide-alkyne cycloaddition (CuAAC) [57]. The limitations of the current method include the fact that the metabolic labeling method is not suitable for clinical tissue samples, but it can be employed for cultured cells and also model animals such as mice [58] and zebra fish [47]. In addition, this method can only be used to identify *N*-glycopeptides containing GalNAc, but even with this limitation, more than 150 unique glycosylation sites were identified in over 100 cell-surface proteins. The



**Figure 4.** Number of *N*-glycosylation sites identified in (a) peptides and (b) proteins. The ModScore distribution for each site localization is shown in (c)

number of glycosylated proteins identified with the current method is comparable to previous results, in which 110 glycoproteins in Jurkat T-lymphocyte cells were identified with the cell surface-capturing method [33].

The majority of *N*-glycopeptides identified (95%) contained only one glycosylation site. Figure 4a shows the distribution of glycosylation sites in peptides; very few (six peptides) have two sites, and only one contains three *N*-glycosylation sites. Similarly, most proteins (72%) are singly glycosylated, and 21% of them contain two *N*-glycosylation sites (Figure 4b). The protein that contains five glycosylation sites was identified as nicastrin (Ncstn), which is an essential subunit of the gamma-secretase complex that is widely expressed in many different tissues [59].

**Table 1.** *N*-Glycosylation Sites in Transporter Proteins Identified in the Current Cell Surface Experiment

Protein	Site	ModScore	Peptide	XCorr	ppm	Annotation
Slc1a4	201	1000	VVTQN#SSSGN#VTHEK	3.4	-0.45	Neutral amino acid transporter; transports alanine, serine, cysteine and threonine
	206	1000				
Slc1a5	212	1000	SYSTTYEERN#ITGTR	3.2	-0.11	Neutral amino acid transporter; accepts all neutral amino acids
Slc2a1	45	1000	VIEEFYN#QTWVHR	3.6	0.38	Facilitated glucose transporter member
Slc30a6	352	17.0	DDWIRPALLSGPVAANVLN#FSDH HVIPMPLLK	3.0	-0.94	Zinc transporter
Slc39a10	339	106.2	DLNEDDHHHECLN#VTQLLK	3.7	-0.08	Zinc transporter
	266	23.6	NTNENPQECFN#ASK	2.8	0.65	
Slc39a6	283	24.6	LLTSHGMGIQVPLN#ATEFNLYLCP AIINQIDAR	5.0	-0.55	Zinc-influx transporter
	67	24.4	YGEN#NLSLVEGFRK	3.8	0.23	
	365	1000	DASSFLAEWQN#ITK	3.1	0.19	
Slc3a2	381	181.2	LLIAGTN#SSDLQQLSLESNK	5.3	-0.46	Involved in sodium-independent neutral amino acid transport including phenylalanine, tyrosine, leucine, arginine and tryptophan
	424	59.3	SLVTQYLN#ATGNR	3.5	-0.75	
	187	1000	GVLNVGN#ETTYEDGHGSRK	3.5	3.24	
Slc44a2	417	84.6	TCNPETFPSSN#ESR	2.5	1.76	Choline transporter-like protein
	776	13.4	LTSYSCVCTEPPNPSN#ETLAQWK	3.8	3.43	
Slc4a7	776	13.4	LTSYSCVCTEPPNPSN#ETLAQWK	3.8	3.43	Sodium bicarbonate cotransporter
Slc6a15	187	1000	N#ASHTFVEPECEQSSATYYWYR	5.0	0.48	Sodium-dependent neutral amino acid transporter, specifically branched-chain amino acids

The red highlights the glycosylation sites

### Cell Surface *N*-Glycoprotein Clustering

All glycoproteins identified in this work were clustered using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) [60]. Membrane proteins were highly enriched with a *P* value of 1.37E-22. Of the 110 total proteins identified, 104 were categorized as membrane proteins. Six proteins were not defined as membrane proteins in the DAVID analysis, which could be due to incorrect annotations of these proteins, contamination, and/or experimental errors. Alternatively, they could be non-membrane proteins that are located on the cell surface through an anchor such as glycosylphosphatidylinositol (GPI) or other modified lipid groups.

All identified glycoproteins were categorized according to a biological process using DAVID. The biological processes with the highest enrichment include cell adhesion, cell motion, structure morphogenesis, transport, interspecies interaction,

positive immune regulation, cell projection organization, stimulus response, and cell recognition (Figure 5a), all of which are consistent with the biological processes associated with cell surface proteins. Glycoproteins were also clustered based on molecular function, and the protein functions that were most highly enriched include signal transducer activity, transporter activity, and carbohydrate binding, as shown in Figure 5b. Cell surface proteins are known to participate in these biological functions and processes [61].

### *N*-Glycosylation Sites on Cell Surface Transporters

Membrane transporter proteins compose a large group of cell surface proteins and they regulate the input and output of many molecules and ions. About 10% of human genes are transporter-related, which corresponds well with the biological significance of transporters and their roles in cell homeostasis [62]. These membrane

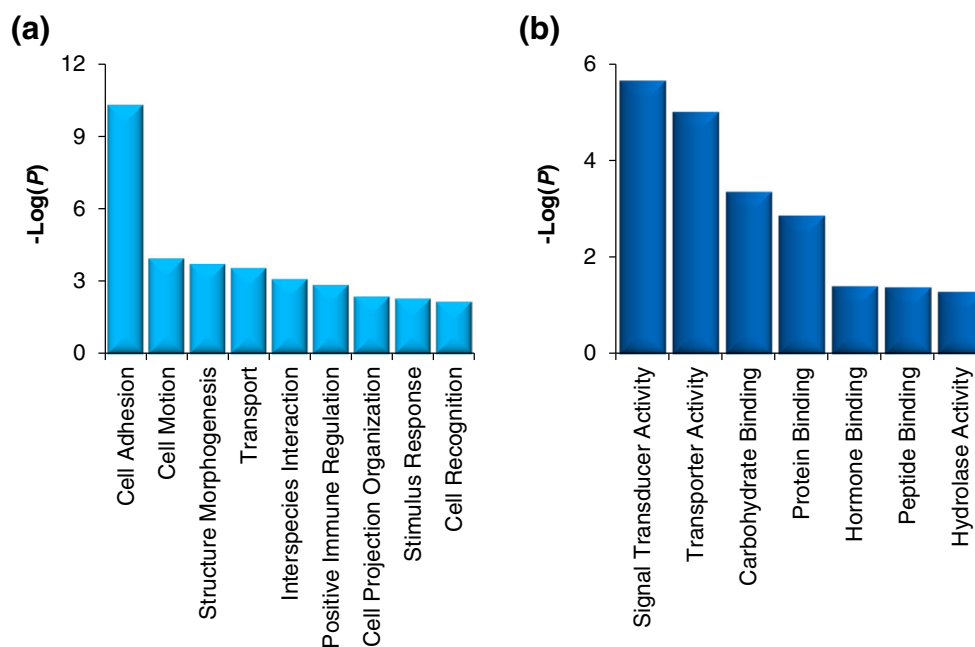


Figure 5. Clustering of glycoproteins based on (a) biological process and (b) molecular function

transporter proteins can serve as potential biomarkers and therapeutic targets. There were 10 transporter proteins identified in this dataset, which are listed in Table 1.

The protein solute carrier family 3 member 2 (Slc3a2), also designated as cluster of differentiation 98 (CD98), is the 4F2 cell-surface antigen heavy chain and is required for the function of light chain amino-acid transporters [63]. In this experiment, three *N*-glycopeptides were confidently identified from Slc3a2, i.e., DASSFLAEWQN#ITK, LLIAGTN#SSDLQQLSLLSENK, and SLVTQYLN#ATGNR with corresponding *N*-glycosylation sites at N365, N381 and N424.

*N*-glycosylation may play determinant roles in trafficking these proteins to the cell surface since glycosylation has been proven to regulate the classic secretory pathway [64]. Glycans within proteins can significantly increase the overall hydrophilicity and impact protein interactions with other small molecules or macromolecules. Reversible glycan binding on these cell surface transporters may also participate in the regulation of molecular transport through the cell membrane.

### *N*-glycosylation Sites on Cluster of Differentiation Proteins

Molecules presented on the cell surface can differentiate and classify the cell type. Cluster of differentiation (CD) molecules are a group of cell surface molecules that are selected to distinguish cell type. Traditionally these CDs are targets for immunophenotyping [65] and often have a wide variety of functions. For example, some CDs are receptors or ligands and some contribute to cell adhesion.

To date, around 350 CDs have been designated to different types of human cells. In this experiment, 35 *N*-glycosylation sites were located on 24 CDs in HEK293T cells (Supplemental Table 2) and some selected sites are listed in Table 2.

Integrins are cell adhesion receptors that are conserved between species, and play critical roles in developmental and pathological processes. The integrin family is heavily involved in mediating the attachment of cells to the extracellular matrix and also takes part in specialized cell–cell interactions [66]. Integrin alpha-1 (Itga1), designated as CD49a, acts as a receptor for laminin and collagen and participates in anchorage-dependent negative regulation of epidermal growth factor-stimulated cell growth. Integrin alpha-2 (Itga2), CD49b, is a receptor for laminin, collagen, collagen C-propeptides, fibronectin, and E-cadherin. It is also accountable for adhesion of platelets and other cells to collagen, the modulation of collagen and collagenase gene expression, and organization of newly synthesized extracellular matrix. Integrin alpha-5 (Itga5), CD49e, is a receptor for fibronectin and fibrinogen. Glycosylation sites were identified on each of these three cell-surface integrins in HEK293T cells.

CD molecules are crucial for classifying cells, but their glycosylation remains largely unstudied. Considering the importance of these molecules and protein glycosylation on the cell surface, information regarding CD glycosylation can provide insight into the functions of these molecules. Cell-surface protein glycosylation analysis can offer valuable information that will lead to the identification of potential drug targets and biomarkers for diseases.



**Table 2.** Selected CD Proteins Identified and Their Corresponding *N*-glycosylation Sites

Protein	CD	Site	ModScore	Peptide	XCorr	ppm	Annotation
Atp1b3	CD298	124	1000	N#LTVCPD <b>G</b> ALFEQ K	3.1	-0.12	Na/K-transporting ATPase subunit beta-3
Bsg	CD147	160	1000	ILLTCSLN#DSATE VTGHR	2.3	0.17	Basigin; participates in the targeting of monocarboxylate transporters to the plasma membrane
		268	1000	ALMN#GSESR	2.0	-0.11	
CD276	CD276	104	1000	TALFPDLLAQGN# ASLR	4.0	-0.42	May be involved in regulation of T-cell-mediated immune response
		215	1000	VVLGAN#GTYSCL VR	3.4	0.86	
CD46	CD46	83	1000	GYFYIPPLATH <b>T</b> IC DRN#HTWLPVSD <b>D</b> ACYR	3.1	1.69	Membrane cofactor protein; cofactor for complement factor I
CD97	CD97	371	1000	N#VTMGQSSAR	1.4	0.13	Receptor possibly involved in adhesion and signaling after leukocyte activation
Insr	CD220	445	1000	HN#LTITQ <b>G</b> K	2.2	0.06	Insulin receptor; receptor tyrosine kinase that mediates the action of insulin
Itga1	CD49a	840	1000	FN#VSLTV <b>K</b>	1.7	0.09	Integrin alpha-1; receptor for laminin and collagen
Itga2	CD49b	105	68.0	LNLQTST <b>S</b> IPN#VTE MK	3.5	0.99	Integrin alpha-2; receptor for laminin, collagen, collagen C-propeptides, fibronectin and E-cadherin
Itga5	CD49e	182	1000	EPLSDPVG <b>T</b> CYLST DN#F <b>T</b> R	4.2	-1.51	Integrin alpha-5; receptor for fibronectin and fibrinogen
Mcam	CD146	56	1000	CGLSQSQGN#LSH VDWFSV <b>H</b> K	3.5	0.19	Cell surface glycoprotein MUC18; participates in cell adhesion

The red highlights the glycosylation sites

## Conclusions

*N*-glycoproteins on the cell surface are essential for a wide range of cellular events, and their global analysis is exceptionally challenging. A novel method has been developed that integrates metabolic labeling, copper-free click chemistry and mass spectrometry-based proteomics techniques to comprehensively analyze cell surface *N*-glycoproteins. Labeling proteins with the azido functional group allows the selective enrichment and separation of cell surface glycoproteins, while the copper-free click chemistry reaction tags cell surface glycoproteins with biotin under physiological conditions. In this experiment, 152 *N*-glycosylation sites were identified in 110 cell surface proteins in HEK293T cells by MS. The main

functions assigned to identified cell surface *N*-glycoproteins, including signal transducer, transporter, binding, and catalytic activity, are consistent with the documented functions of cell surface proteins. The current strategy provides an effective method for large-scale analysis of the cell surface *N*-glycoproteome, and can be extensively applied for further cell surface studies.

## Acknowledgments

The authors gratefully acknowledge Dr. M.G. Finn for his valuable advice regarding click chemistry, and Dr. Gang Bao for the HEK293T cell line. This work was supported by start-

up funds from Georgia Institute of Technology. R.W. is also supported by the Blanchard Assistant Professorship.

## References

- Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M.: Electrospray ionization for mass-spectrometry of large biomolecules. *Science* **246**, 64–71 (1989)
- Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y.: Protein and polymer analysis up to  $m/z$  100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988)
- Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E., Engelhard, V.H.: Characterization of peptides bound to the class-I MHC molecule HLA-A2.1 by mass-spectrometry. *Science* **255**, 1261–1263 (1992)
- Holliman, C.L., Rempel, D.L., Gross, M.L.: Detection of high mass-to-charge ions by Fourier-transform mass-spectrometry. *Mass Spectrom. Rev.* **13**, 105–132 (1994)
- Loo, J.A.: Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **16**, 1–23 (1997)
- Hebert, A.S., Merrill, A.E., Bailey, D.J., Still, A.J., Westphall, M.S., Strieter, E.R., Pagliarini, D.J., Coon, J.J.: Neutron-encoded mass signatures for multiplexed proteome quantification. *Nat. Methods* **10**, 332–334 (2013)
- Skinner, O.S., Catherman, A.D., Early, B.P., Thomas, P.M., Compton, P.D., Kelleher, N.L.: Fragmentation of integral membrane proteins in the gas phase. *Anal. Chem.* **86**, 4627–4634 (2014)
- Zhou, M.W., Wysocki, V.H.: Surface induced dissociation: dissecting noncovalent protein complexes in the gas phase. *Acc. Chem. Res.* **47**, 1010–1018 (2014)
- Yates, J.R.: The revolution and evolution of shotgun proteomics for large-scale proteome analysis. *J. Am. Chem. Soc.* **135**, 1629–1640 (2013)
- Neilson, K.A., Scafaro, A.P., Chick, J.M., George, I.S., Van Sluyter, S.C., Gygi, S.P., Atwell, B.J., Haynes, P.A.: The influence of signals from chilled roots on the proteome of shoot tissues in rice seedlings. *Proteomics* **13**, 1922–1933 (2013)
- Ong, S.E., Mann, M.: Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252–262 (2005)
- Yao, X.D., Freas, A., Ramirez, J., Demirev, P.A., Fenselau, C.: Proteolytic O-18 labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* **73**, 2836–2842 (2001)
- Zhu, Z.K., Go, E.P., Desaire, H.: Absolute quantitation of glycosylation site occupancy using isotopically labeled standards and LC-MS. *J. Am. Soc. Mass Spectrom.* **25**, 1012–1017 (2014)
- Wang, Y.C., Jobe, S.M., Ding, X.K., Choo, H.J., Archer, D.R., Mi, R.J., Ju, T.Z., Cummings, R.D.: Platelet biogenesis and functions require correct protein O-glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16143–16148 (2012)
- Dai, L.Z., Peng, C., Montellier, E., Lu, Z.K., Chen, Y., Ishii, H., Debernardi, A., Buchou, T., Rousseaux, S., Jin, F.L., Sabari, B.R., Deng, Z.Y., Allis, C.D., Ren, B., Khochbin, S., Zhao, Y.M.: Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nat. Chem. Biol.* **10**, 365–370 (2014)
- Ma, J.F., Hart, G.W.: Protein O-GlcNAcylation in diabetes and diabetic complications. *Exp. Rev. Proteom.* **10**, 365–380 (2013)
- Xiao, Y., Lee, T., Latham, M.P., Warner, L.R., Tanimoto, A., Pardi, A., Ahn, N.G.: Phosphorylation releases constraints to domain motion in ERK2. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2506–2511 (2014)
- Bullen, J.W., Balsbaugh, J.L., Chanda, D., Shabanowitz, J., Hunt, D.F., Neumann, D., Hart, G.W.: Cross-talk between two essential nutrient-sensitive enzymes O-GlcNAc transferase (OGT) and AMP-activated protein kinase (AMPK). *J. Biol. Chem.* **289**, 10592–10606 (2014)
- Nie, S., Lo, A., Wu, J., Zhu, J.H., Tan, Z.J., Simeone, D.M., Anderson, M.A., Shedden, K.A., Ruffin, M.T., Lubman, D.M.: Glycoprotein biomarker panel for pancreatic cancer discovered by quantitative proteomics analysis. *J. Proteome Res.* **13**, 1873–1884 (2014)
- Wu, J., Xie, X.L., Nie, S., Buckanovich, R.J., Lubman, D.M.: Altered expression of sialylated glycoproteins in ovarian cancer sera using lectin-based ELISA assay and quantitative glycoproteomics analysis. *J. Proteome Res.* **12**, 3342–3352 (2013)
- Gilgunn, S., Conroy, P.J., Saldova, R., Rudd, P.M., O'Kennedy, R.J.: Aberrant PSA glycosylation—a sweet predictor of prostate cancer. *Nat. Rev. Urol.* **10**, 99–107 (2013)
- Oda, Y., Nagasu, T., Chait, B.T.: Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat. Biotechnol.* **19**, 379–382 (2001)
- Ficarro, S.B., McClelland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J., Hunt, D.F., White, F.M.: Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301–305 (2002)
- Yang, W.M., Shah, P., Eshghi, S.T., Yang, S., Sun, S.S., Ao, M.H., Rubin, A., Jackson, J.B., Zhang, H.: Glycoform analysis of recombinant and human immunodeficiency virus envelope protein GP120 via higher energy collisional dissociation and spectral-aligning strategy. *Anal. Chem.* **86**, 6959–6967 (2014)
- Chen, W.X., Smeekens, J.M., Wu, R.H.: Comprehensive analysis of protein *N*-glycosylation sites by combining chemical deglycosylation with LC-MS. *J. Proteome Res.* **13**, 1466–1473 (2014)
- Wu, R., Haas, W., Dephoure, N., Huttlin, E.L., Zhai, B., Sowa, M.E., Gygi, S.P.: A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat. Methods* **8**, 677–683 (2011)
- Mann, M., Jensen, O.N.: Proteomic analysis of post-translational modifications. *Nat. Biotechnol.* **21**, 255–261 (2003)
- Witze, E.S., Old, W.M., Resing, K.A., Ahn, N.G.: Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods* **4**, 798–806 (2007)
- Zhao, Y.Y., Takahashi, M., Gu, J.G., Miyoshi, E., Matsumoto, A., Kitazume, S., Taniguchi, N.: Functional roles of *N*-glycans in cell signaling and cell adhesion in cancer. *Cancer Sci.* **99**, 1304–1310 (2008)
- Ohtsubo, K., Marth, J.D.: Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867 (2006)
- Apweiler, R., Hermjakob, H., Sharon, N.: On the frequency of protein glycosylation, as deduced from analysis of the Swiss-Prot database. *Biochim Biophys Acta Gen Subj* **1473**, 4–8 (1999)
- Vigerust, D.J.: Protein glycosylation in infectious disease pathobiology and treatment. *Cent. Eur. J. Biol.* **6**, 802–816 (2011)
- Wollscheid, B., Bausch-Fluck, D., Henderson, C., O'Brien, R., Bibel, M., Schiess, R., Aebersold, R., Watts, J.D.: Mass-spectrometric identification and relative quantification of *N*-linked cell surface glycoproteins. *Nat. Biotechnol.* **27**, 378–386 (2009)
- Chang, P.V., Prescher, J.A., Hangauer, M.J., Bertozzi, C.R.: Imaging cell surface glycans with bioorthogonal chemical reporters. *J. Am. Chem. Soc.* **129**, 8400–8401 (2007)
- Wessel, D., Flugge, U.I.: A method for the quantitative recovery of protein in dilute-solution in the presence of detergents and lipids. *Anal. Biochem.* **138**, 141–143 (1984)
- Eng, J.K., McCormack, A.L., Yates, J.R.: An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994)
- Peng, J.M., Elias, J.E., Thoreen, C.C., Licklider, L.J., Gygi, S.P.: Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2**, 43–50 (2003)
- Elias, J.E., Gygi, S.P.: Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**, 207–214 (2007)
- Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villen, J., Haas, W., Sowa, M.E., Gygi, S.P.: A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010)
- Du, X., Callister, S.J., Manes, N.P., Adkins, J.N., Alexandridis, R.A., Zeng, X., Roh, J.H., Smith, W.E., Donohue, T.J., Kaplan, S., Smith, R.D., Lipton, M.S.: A computational strategy to analyze label-free temporal bottom-up proteomics data. *J. Proteome Res.* **7**, 2595–2604 (2008)
- 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* **4**, 923–925 (2007)
- 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.* **24**, 1285–1292 (2006)
- Asano, N.: Naturally occurring iminosugars and related compounds: structure, distribution, and biological activity. *Curr. Top. Med. Chem.* **3**, 471–484 (2003)

44. 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.* **108**, 8146–8151 (2011)
45. Hsu, T.L., Hanson, S.R., Kishikawa, K., Wang, S.K., Sawa, M., Wong, C.H.: Alkynyl sugar analogs for the labeling and visualization of glycoconjugates in cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2614–2619 (2007)
46. Laughlin, S.T., Bertozzi, C.R.: Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation. *Nat. Protoc.* **2**, 2930–2944 (2007)
47. Laughlin, S.T., Baskin, J.M., Amacher, S.L., Bertozzi, C.R.: In vivo imaging of membrane-associated glycans in developing zebrafish. *Science* **320**, 664–667 (2008)
48. Boyce, M., Carrico, I.S., Ganguli, A.S., Yu, S.H., Hangauer, M.J., Hubbard, S.C., Kohler, J.J., Bertozzi, C.R.: Metabolic cross-talk allows labeling of O-linked beta-*N*-acetylglucosamine-modified proteins via the *N*-acetylgalactosamine salvage pathway. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3141–3146 (2011)
49. Sletten, E.M., Nakamura, H., Jewett, J.C., Bertozzi, C.R.: Difluorobenzocyclooctyne: synthesis, reactivity, and stabilization by beta-cyclodextrin. *J. Am. Chem. Soc.* **132**, 11799–11805 (2010)
50. Mbua, N.E., Guo, J., Wolfert, M.A., Steet, R., Boons, G.J.: Strain-promoted alkyne-azide cycloadditions (SPAAC) reveal new features of glycoconjugate biosynthesis. *ChemBioChem* **12**, 1911–1920 (2011)
51. Kaji, H., Saito, H., Yamauchi, Y., Shinkawa, T., Taoka, M., Hirabayashi, J., Kasai, K., Takahashi, N., Isobe, T.: Lectin affinity capture, isotope-coded tagging, and mass spectrometry to identify *N*-linked glycoproteins. *Nat. Biotechnol.* **21**, 667–672 (2003)
52. Zhang, H., Li, X.J., Martin, D.B., Aebersold, R.: Identification and quantification of *N*-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **21**, 660–666 (2003)
53. Agarwal, A.K., Gude, R.P., Kalraiya, R.D.: Regulation of melanoma metastasis to lungs by cell surface lysosome associated membrane protein-1 (Lamp1) via galectin-3. *Biochem. Biophys. Res. Commun.* **449**, 332–337 (2014)
54. Kall, L., Storey, J.D., MacCoss, M.J., Noble, W.S.: Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J. Proteome Res.* **7**, 29–34 (2008)
55. Kim, W., Bennett, E.J., Huttlin, E.L., Guo, A., Li, J., Possemato, A., Sowa, M.E., Rad, R., Rush, J., Comb, M.J., Harper, J.W., Gygi, S.P.: Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011)
56. Chen, W.X., Smeekens, J.M., Wu, R.H.: A universal chemical enrichment method for mapping the yeast *N*-glycoproteome by MS. *Mol. Cell. Proteomics* **13**, 1563–1572 (2014)
57. Wang, Q., Chan, T.R., Hilgraf, R., Fokin, V.V., Sharpless, K.B., Finn, M.G.: Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J. Am. Chem. Soc.* **125**, 3192–3193 (2003)
58. Dube, D.H., Prescher, J.A., Quang, C.N., Bertozzi, C.R.: Probing mucin-type O-linked glycosylation in living animals. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4819–4824 (2006)
59. Hu, Y., Fortini, M.E.: Different cofactor activities in gamma-secretase assembly: evidence for a nicastrin-Aph-1 subcomplex. *J. Cell Biol.* **161**, 685–690 (2003)
60. Huang, D.W., Sherman, B.T., Lempicki, R.A.: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009)
61. Lodish, H.: *Molecular Cell Biology*, 4th ed. W. H. Freeman, New York (2008)
62. Hediger, M.A., Clemençon, B., Burrier, R.E., Bruford, E.A.: The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol. Asp. Med.* **34**, 95–107 (2013)
63. Mastroberardino, L., Spindler, B., Pfeiffer, R., Skelly, P.J., Loffing, J., Shoemaker, C.B., Verrey, F.: Amino-acid transport by heterodimers of 4F2hc/CD98 and members of a permease family. *Nature* **395**, 288–291 (1998)
64. Varki, A.: Biological roles of oligosaccharides—all of the theories are correct. *Glycobiology* **3**, 97–130 (1993)
65. Zola, H., Swart, B., Banham, A., Barry, S., Beare, A., Bensussan, A., Boumsell, L., Buckley, C.D., Buhning, H.J., Clark, G., Engel, P., Fox, D., Jin, B.Q., Macardle, P.J., Malavasi, F., Mason, D., Stockinger, H., Yang, X.F.: CD molecules 2006—human cell differentiation molecules. *J. Immunol. Methods* **319**, 1–5 (2007)
66. Barczyk, M., Carracedo, S., Gullberg, D.: Integrins. *Cell Tissue Res.* **339**, 269–280 (2010)